A THEORY OF SIGNALS

ROBERT E. WERNIKOFF

TECHNICAL REPORT 331

JANUARY 31, 1958

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory of the Department of Electrical Engineering and the Department of Physics.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the U. S. Army (Signal Corps), the U. S. Navy (Office of Naval Research), and the U. S. Air Force (Office of Scientific Research, Air Research and Development Command), under Signal Corps Contract DA36-039-sc-64637, Department of the Army Task 3-99-06-108 and Project 3-99-00-100.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 331 January 31, 1958

A THEORY OF SIGNALS

Robert E. Wernikoff

This report was reproduced by photo-offset printing from the thesis copy which was submitted to the Department of Electrical Engineering, M.I.T., January, 1958, in partial fulfillment of the requirements for the degree of Doctor of Science.

Abstract

An experiment is presented in which an attempt is made to arrive at a mathematical description of physical signals that embodies, more realistically than the usual functional representation, our limitations in performing measurements. The object is to achieve a closer relation between the structure of the mathematical description and the finite-resolution properties of the detector that characterizes any real measurement process.

An algebra of signals is obtained, appropriate to the model in which essentially frequency-limited signals interact with linear, time-invariant systems, and observations are made by means of a linear, finite-resolution oscilloscope. The properties of this algebra are studied, and a metric that indicates which operators give physically indistinguishable outputs is defined. The algebra is used to study problems in uniform and nonuniform sampling, the discrimination of two events from one in noisy, radar-like systems, and the conditions under which a signal is indistinguishable from its short-time average. A general procedure for linear, least-peak-error prediction is obtained. In the limit as the detector resolution becomes perfect, the present model is shown to tend smoothly to the usual functional model.

ACKNOWLEDGEMENT

These notes always fall short of the mark: there seems to be no way of describing, within one short paragraph, the variety and extent of the benefits derived from associating for four years with a large group of generous and stimulating people, or the warmth with which the obligation is recognized. Still, within the limitations of a form that makes a genuine feeling seem dull and dutiful, it gives me great pleasure to acknowledge my indebtedness to Professor Y.-W. Lee for his friendly and careful guidance during my stay in the Graduate School, and for supervising this thesis; and to Professor C. E. Shannon and Professor Peter Elias for encouraging me, and for their helpful suggestions and criticisms. I am more grateful than I can express to Dr. M. V. Cerrillo for the generous and inspiring introduction to research that he gave me. I am indebted to the Research Laboratory of Electronics for extending to me its support and the use of its facilities; and to many members of its staff for numerous informative and stimulating discussions.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 THE PROBLEM.

In studying a communications problem or the interaction of a signal with a system analytically, the usual approach is to start by representing signals as functions, make what appear to be intuitively reasonable assumptions about the analytic properties of these functions, and proceed from there. We do this so often and so naturally that we tend to forget that there is nothing self-evident about thinking of signals as functions, and even less about what, in effect, we actually do, which is to identify signals and functions.

While it is true that any process occurring in time can be thought of as defining, in some way, a function of time, this function must forever remain unknown to us, because when we observe the process through instruments, and, ultimately, our own senses – and these are the only ways that we can ever be aware of the process – what we perceive is, of course, not the process itself, but the process-as-seen-through-our-instruments. If the instruments were ideal, there would be no difference. But our instruments are not ideal: they have finite resolution, finite noise, finite inertia, and so forth. And the final display, which might take the form of an oscillographic recording or a pencilled graph, is a finite collection of dots of finite size observed by the imperfect mechanism of the human eye.

The relation between the functions that we talk about on paper and the physical phenomena observed in the laboratory is a somewhat arbitrary one. If it simplifies our work, we can, if we want to, study functions instead of signals, and say that we shall think of functions as models of the physical phenomena. But there is nothing inevitable about this choice. It is completely arbitrary, there exist no theoretical justifications for it, and the only criteria for doing it that we can appeal to are: it is simple; it works. All this is obvious and well-known, and therefore sometimes forgotten. The functional modelling of signals is so successful so often that it is sometimes overlooked that it is just a model. And far from being an obvious model, it is, if one thinks about it, if anything rather surprising that the functional model is so successful, because in many ways there is very little resemblance between a signal and a function.

Let us discuss this a bit further. We shall, here and in the remainder of this paper, restrict the use of the word "signal" to the thing that actually exists in the laboratory, voltages on wires, currents issuing from plugs, and so forth. If we wish to study a signal, we must do it through instruments (real ones, with real limitations), and ultimately through our senses. Let us now imagine that we are going to study a signal, the object of this brief study being the comparison of the properties of a signal with those of a mathematical function. Some of the most fundamental properties of functions are their analytical properties. Let us start our study with them.

In mathematics, when we say that a function f(t) is continuous at the point t_o, we mean that, given any positive number ϵ, however small, there exists a positive number δ with the property that $|f(t_o) - f(t)| < \epsilon$ whenever $|t_o - t| < \delta$. The function is continuous on an interval if it is continuous at every point of that interval.

What do we mean when we ask whether or not a signal is continuous? If we think of signals as identified with functions, then presumably we mean that the same test should be applied to signals, and, of course, we must do this through our instruments. We run into difficulties immediately: first of all, it is impossible to make measurements at every point of an interval of time, as the definition requires; next, even if we limit our question to asking about continuity at a point, it is usually impossible to identify one specific time t_o exactly. But even ignoring these problems, it is clear that we cannot implement the definition because, sooner or later, we shall reach a value of ϵ or δ that will be below our limit of resolution. Since the definition requires specifically that we perform the test for every value of ϵ however small (otherwise, how can we be sure that the function does not have discontinuities of magnitude smaller than our limit of resolution?), we see that we cannot verify the continuity of a signal.

Clearly, the same thing would happen if we considered differentiability, or, more generally, if we considered any property whose definition depends on a limit process. Questions about such properties all have in common the fact that, because of our and our

instruments' limitations, they are operationally meaningless (see, e.g., Bridgeman (1)). That is, we cannot prescribe a set of experiments that will decide them.

Let us try a different tack. A mathematical function is defined (see, e.g., Hobson (2)) as follows:

> If, to each point of the domain of the independent variable x, there be made to correspond in any manner a definite number, such that all such numbers form a new aggregate which can be regarded as the domain... of a new variable y, this variable y is said to be a (single-valued) function of x. ...the conception of the functional relation contains nothing more than the notion of determinate correspondence

Now, is a signal a function? As seen through our instruments and senses, clearly, no, because we cannot establish the required determinate correspondence if we cannot isolate instants in time and if, for example, our voltage-measuring devices have finite resolution (say, a meter needle of finite width).

The point of this perhaps unnecessarily long discussion was to show that there is nothing inevitable about thinking of signals as functions. The functional representation is just one possible model, and in some ways not a very close one at that. Thus refreshed, we can proceed to consider other possibilities.

What is a signal? Obviously, there is no general answer to this. If we persevere in our operational viewpoint, a signal is completely unknown to us until we observe it, and what it is when we observe it depends very much on what we observe it with. How should we construct a model to talk about signals? Again, there is no general answer, but it is possible that something in the way of

simplicity or efficiency might be gained from making the model resemble the signal as we observe it. (For example, the most economical model of a signal-as-observed-on-a-linear-oscilloscope is probably different from that of a signal-as-observed-on-a-zero-crossing-detector.) From this point of view, the functional model is the model of signals-as-observed-on-infinite-resolution-ideal-instruments.

The functional model serves us well, but we know that in most of our work our equipment has finite resolution. The model should therefore be of signals-as-observed-with-finite-resolution. Can such models be constructed? And can greater simplicity result from using such a model? There is no unqualified answer to this question, because the simplicity of a model depends to a great extent on the available analytical technique. As it happens, the most extensively developed one is the calculus, which is the technique appropriate to the functional model. (It is largely to this somewhat fortuitous circumstance that the functional model owes its success, rather than to any intrinsic appropriateness of the functional representation of signals.) But there are other branches of mathematics that are fairly well developed, enough perhaps to be able to work with them with reasonable ease. Perhaps one could be found that will allow us to construct, and work easily with, a model of signals-as-observed-with-finite-resolution. It is clear that there are always two elements that must somehow be combined: properties desirable on physical grounds, and mathematical possibilities.

Attempting to construct such a model is, perhaps, a somewhat academic project, but it seems to be an experiment worth trying (once, at any rate) because, after all, in most physical situations we are limited to observation with finite resolution.

1.2 THE THESIS.

This dissertation is a report on just such an experiment. We construct a model of signals, and study their interactions with linear, time-invariant systems, under the supposition that the ultimate observation instrument is a linear oscilloscope with finite trace width. We do not, in this model, attempt to include all possible signals — only those which, like phonograph music or FCC-controlled broadcast signals, suggest the idea of band-limitation or, more generally, of frequency-limitation. As described in Section 2.1, the model of a signal is taken to be a class of functions rather than a single function. It then follows, as is shown in Sections 2.2 and 2.3, that the linear systems with which the signals interact can be represented as elements of a unitary linear vector space, on which is also defined a commutative multiplication. This gives us an algebra whose properties are then studied, with examples, for the remainder of Chapter 2.

In the limit, as the resolution of our oscilloscope becomes better and better, our results should approach smoothly those of the functional model. This is shown to be the case in Chapter 3, where we also present various examples of the

application of our model to the study of signals and other questions. Among these results is the derivation of a minimal requirement for the discrimination of individual events in noisy radar-like systems; and the derivation of the optimum (in the sense of least peak error) predicting filter. The latter is a special case of the derivation of the procedure for least-peak-error interpolation and extrapolation.

The derivation of this optimum filter gives us at least one example of a problem that can be solved (quite easily, too) with our model, and that cannot (or, at least, apparently never has been) solved with the functional model, because here is one case where the usually compliant calculus becomes intractable. And since there is a smooth transition from our model to the functional one, the solution within our model suggests the solution within the functional one. This hints at (but no more than hints, since one single example is hardly a compelling argument in favor of anything) an additional benefit to be derived from this model, or, more generally, to be derived from having available a multiplicity of models that are consistent in some appropriate limit. This benefit, of course, is the availability of alternative lines of attack on problems.

This dissertation does not by any means exhaust the results and applications inherent in even this model, let alone those in the many useful variants of this model that are mentioned here or that will suggest themselves to the reader. But it does perhaps give a hint of the possibilities for the construction of useful, simple models that are designed purposely to be cogent, economical

representations of physical situations. It is our thesis that perhaps modelmaking, considered as a problem in its own right, can lead to simpler and therefore more successful analyses of some of our unsolved problems.

1.3 HISTORICAL SKETCH.

This problem has virtually no history, at least not in the technical literature. Its philosophical basis, extensive both in time and quantity, is splendidly presented by such authors as, for example, Mach (3), Born (4), Bridgeman (2), and, more recently, Brillouin (5), (6). The starting point for this work was a paper by Cerrillo (7), which first suggested to the author the tangible benefits potentially derivable from the explicit inclusion of a finite error in the formulation of a problem. The physical sources of finite resolution are discussed in Gabor (8), but he does not study the modelling problem further. A very large amount of work has of course been done on approximate techniques for the study of signal/system interaction and the synthesis of systems, but these are not models, they are just computational aids. An exhaustive bibliography of these papers (up to 1954) is given in Winkler (9). More pertinent to the present work is a group of papers on signal theory (31) in which some of these general questions are discussed. Recently, some of the ideas in this thesis were applied by Jacobs (10).

CHAPTER II

AN ALGEBRA OF SIGNALS

In this chapter we present and develop a formalism for the analysis of signals and their interactions with linear systems specifically suited to the case when the signals are viewed with a linear, finite-resolution measurement device. For simplicity and brevity, the presentation will be axiomatic rather than heuristic. Examples illustrating the physical significance of the mathematical concepts will be interspersed at appropriate points in the discussion.

2.1 INITIAL ASSUMPTIONS: THE MEASUREMENT PROCESS

The thought experiment appropriate to the analysis is the following:

We consider two linear, time-invariant networks, excited from a common source, and observe the outputs on an oscilloscope, as in Fig. 2.1.1.

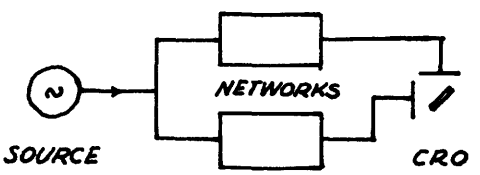

Fig. 2.1.1.

The measurement process.

Concerning the experiment, we make the following three assumptions:

1. The oscilloscope is linear, with a trace of finite length and width.

2. The probe (i.e., the excitation) is any member of the class $\underline{S}_{\psi}$ of functions that are L_2 solutions of the equation

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\tau)\psi(t - \tau)\, d\tau \qquad (2.1.1)$$

where ψ is any real, L_2 function with an idempotent spectrum.*

The probe functions f(t) are thought of as representative of bandlimited or, more generally, frequency-limited signals such as, for example, phonograph or broadcast music. This does not mean, for example, that we think of some specific $f_o(t) \in \underline{S}_{\psi}$ as <u>being</u> a particular performance of Beethoven's Ninth Symphony, since this would be logically and semantically absurd. It does not even mean that we suppose that any sort of correspondence can be established between individual members of the set of performed symphonies and individual members of a set $\underline{S}_{\psi}$. All that it means is that we shall further assume that

3. A network or experiment that is simultaneously successful with <u>all</u> the functions in any one $\underline{S}_{\psi}$ class will also be successful with, say, phonograph music represented by that class.

---

* The spectrum $\Psi(\omega)$ of $\psi(t)$ is idempotent if it assumes only the values 0 or 1, so that it is a solution of the equation $\Psi(\omega) = [\Psi(\omega)]^n$ for any positive integer n. In particular, this implies that $\Psi(\omega)$ is real. Since the $\psi(t)$ we are interested in are also real, the corresponding $\Psi(\omega)$ are even, and therefore so are the $\psi(t)$ also. Examples of acceptable spectra are shown in Fig. 2.1.2, at the end of this section.

The last assumption states the possibility of supposing a relationship to exist between ensembles of functions (the classes $\underline{S}_\psi$) and sets of physical signals (e. g., phonograph music) without making any definite claims about the possibility of representing a specific signal by a specific function. Accordingly, it specifies that any mathematical result, to be physically useful or even interpretable, must be established for a whole ensemble $\underline{S}_\psi$ of functions rather than for just one member function. This assumption seems to be intuitively reasonable and in accord with the point of view of modern communication theory.

Besides establishing the ground rules for the game, the function of the third assumption is to provide a bridge between the mathematical model to be developed here and a useful set of physical situations. As usual, there exists no paper method for establishing such a relationship. The very desire to attempt to do so is logically meaningless, since it corresponds to trying to prove propositions concerning the equivalence or correspondence of pairs of objects of which half are mathematical and the other half physical. Such a bridge between model and physical situation can, apparently, only be intuitive, and is subject to partial confirmation or complete rejection by experiment.

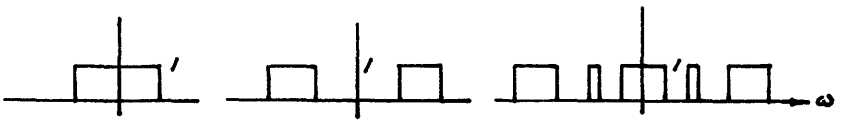

Fig. 2.1.2.

Acceptable spectra.

2.2 SINGULAR NETWORKS AND THEIR ELEMENTARY PROPERTIES

A time-invariant linear network whose impulse response consists of a finite number of impulses of finite area distributed over a finite time interval shall be called a singular network. The impulse response of a typical singular network might be

$$h(t) = \sum_{n=1}^{N} a_n \delta(t - t_n) \tag{2.2.1}$$

Such a network shall be represented by the operator

$$\Omega = \sum_{n=1}^{N} a_n E^{-t_n} \tag{2.2.2}$$

where the shift operator E^{-t_n} is defined by

$$E^{-t_n} f(t) = f(t - t_n) \tag{2.2.3}$$

so that, if f(t) is applied to a network characterized by Ω, the output will be

$$f_o(t) = \Omega f(t) = \sum_{n=1}^{N} a_n f(t - t_n) \tag{2.2.4}$$

It is clear that the same input f(t), when convolved with the impulse response of (2.2.1), results in the same output $f_o(t)$, so that the representations (2.2.1) and (2.2.2) are indeed equivalent.

Clearly, many other representations are possible, and some ($\mathcal{Z}$-transforms, time-series) are extensively discussed in the literature (for references, see e.g., (18), (19), (20), (20)). The numerical methods are rather rigid and cumbersome and therefore somewhat sterile except for actual computation. The transform approaches are

rather sophisticated and therefore tend to cloud the issues without any compensating advantages. Formally, of course, they are no different from the operators used here. We shall employ the operator representation (2.2.2); it is simple, direct, and, as later developments will show, very suggestive.

We consider now the interaction of probe functions and singular networks. From (2.1.1), a probe function is any L_2 solution of the equation

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\tau) \psi(t - \tau)\, d\tau \tag{2.2.5}$$

where ψ, which is descriptive of the particular class of signals under consideration, is considered fixed in any one discussion. If the probe function f(t) is used as input to a network Ω, the output can be represented by

$$f_o(t) = \Omega f(t) = \sum_n a_n f(t - t_n) = \sum_n a_n \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\tau) \psi(t - t_n - \tau)\, d\tau$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\tau) \sum_n a_n \psi(t - t_n - \tau)\, d\tau = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\tau) \left[\Omega \psi(t - \tau)\right] d\tau \tag{2.2.6}$$

Since, in (2.2.5), the change of variable $u = t - \tau$ is legitimate, and yields

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t - u) \psi(u)\, du$$

the reasoning leading to (2.2.6) would produce, in this case,

$$f_o(t) = \Omega f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\sum_n a_n f(t - t_n - u)\right] \psi(u)\, du = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\Omega f(t - u)\right] \psi(u)\, du \tag{2.2.7}$$

If now $\mathcal{A}$ is time invariant, so that the a_n are constants, we can change back to $t = u + \tau$, obtaining

$$f_o(t) = \mathcal{A}\, f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\mathcal{A}\, f(\tau)] \psi(t - \tau)\, d\tau \qquad (2.2.8)$$

Eq. (2.2.8) expresses the fact that if f(t) is a solution of (2.2.5) and $\mathcal{A}$ is time-invariant, then the output $\mathcal{A}\, f(t)$ is also a solution of (2.2.5). This establishes the obvious result that, for time-invariant networks, $\mathcal{A}\, f(t)$ is a probe function if f(t) is one, and that both belong to the same class $\underline{S}_\psi$. It is therefore meaningful to define a succession of operations and therefore, as will be done later, a multiplication for the $\mathcal{A}$'s. The conclusion about class preservation is not true when the coefficients a_n in (2.2.7) are functions of time. It is for this reason that time-varying systems will not be considered here.

Before continuing, it will be convenient to present some elementary identities which, while somewhat dull in themselves, will be useful later on.

Let the symbol $\mathcal{A}^*$ denote the operator obtained from

$$\mathcal{A} = \sum_{n=1}^{N} a_n E^{-t_n} \qquad (2.2.9)$$

by replacing $-t_n$, in the exponent of the shift operator, with $+t_n$. That is, if $\mathcal{A}$ is given by (2.2.9), then $\mathcal{A}^*$ is defined by

$$\mathcal{A}^* = \sum_{n=1}^{N} a_n E^{+t_n} \qquad (2.2.10)$$

so that $\mathcal{A}^*$ is just $\mathcal{A}$ folded over in time. Clearly, $(\mathcal{A}^*)^* = \mathcal{A}$. (There is an obvious analogy between the star operation and complex conjugation,

the reason being the parallelism between operator notation and the Fourier transform of the impulse response represented by the operator.) In terms of this notation we have

$$1.\quad \int_{-\infty}^{\infty} f(t)\left[\Omega\psi(t)\right] dt = \int_{-\infty}^{\infty} \left[\Omega^{*} f(t)\right] \psi(t)\, dt \qquad (2.2.11)$$

Proof: Starting from the left side,

$$\int f(t)\left[\Omega\psi(t)\right] dt = \int f(t) \sum_{n} \alpha_n \psi(t - t_n)\, dt = \sum_{n} \alpha_n \int f(t)\psi(t - t_n)\, dt$$

In the n^{th} term of the sum, the change of variable $u = t - t_n$ then yields

$$= \sum_{n} \alpha_n \int f(u + t_n)\psi(u)\, du = \int \sum_{n} \alpha_n f(u + t_n)\psi(u)\, du$$

$$= \int \left[\Omega^{*} f(t)\right]\psi(t)\, dt$$

Before presenting the next identity, we must first specify how to interpret the successive application of operators. If $\Omega_1 = \sum_{n} \alpha_n E^{-t_n}$ and $\Omega_2 = \sum_{m} \beta_m E^{-t_m}$, then

$$\Omega_2\left[\Omega_1 f(t)\right] = \Omega_2\left[\sum_{n} \alpha_n f(t - t_n)\right] = \sum_{n} \alpha_n \Omega_2 f(t - t_n)$$

$$= \sum_{n} \alpha_n \sum_{m} \beta_m f(t - t_m - t_n) = \sum_{n} \sum_{m} \alpha_n \beta_m f(t - t_m - t_n)$$

$$= \sum_{n} \sum_{m} \alpha_n \beta_m E^{-t_n - t_m} f(t) \qquad (2.2.12)$$

$$= \sum_{m} \beta_m \sum_{n} \alpha_n f(t - t_n - t_m) = \sum_{m} \beta_m \Omega_1 f(t - t_m)$$

$$= \Omega_1 \sum_{m} \beta_m f(t - t_m) = \Omega_1\left[\Omega_2 f(t)\right] \qquad (2.2.13)$$

These results show that

$$\Omega_2[\Omega_1 f(t)] = (\Omega_2\Omega_1) f(t) = (\Omega_1\Omega_2) f(t) = \Omega_1[\Omega_2 f(t)]$$

so that the successive application of two operators can be considered as a "product" of the operators, with multiplication defined as for polynomials (Cauchy product) as shown in (2.2.12).

We now obtain

$$2. \quad \int_{-\infty}^{\infty} [\Omega_1 f(t)][\Omega_2 \psi(t)]\, dt = 2\pi \Omega_2^* \Omega_1 f(0) \qquad (2.2.14)$$

Proof: From (2.2.8) we have

$$\Omega_1 f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\Omega_1 f(\tau)] \psi(t - \tau)\, d\tau$$

and from (2.2.6)

$$\Omega_2^*[\Omega_1 f(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\Omega_1 f(\tau)][\Omega_2^* \psi(t - \tau)]\, d\tau$$

At t = 0,

$$\Omega_2^* \Omega_1 f(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\Omega_1 f(\tau)][\Omega_2^* \psi(-\tau)]\, d\tau$$

In the second factor of the integrand, because of the evenness of $\psi(\tau)$,

$$\Omega_2^* \psi(-\tau) = \sum_n a_n \psi(-t_m - \tau) = \sum_n a_n \psi(t_n + \tau) = \Omega_2^{**} \psi(\tau) = \Omega_2 \psi(\tau).$$

Making this substitution establishes (2.2.14).

$$3. \quad \int_{-\infty}^{\infty} [\Omega f(t)]^2\, dt = \Omega^* \Omega R_f(0) \qquad (2.2.15)$$

where $R_f(\tau)$ is the autocorrelation function of f(t) defined by

$$R_f(\tau) = \int_{-\infty}^{\infty} f(t)\, f(t + \tau)\, dt$$

The proof is obtained by straightforward substitution, in (2.2.15), of Ω expressed in terms of shifts.

4. $$\int_{-\infty}^{\infty} [\Omega\psi(t)]^2 \, dt = 2\pi \Omega^* \Omega \psi(0) \tag{2.2.16}$$

Proof: This result is obtained from (2.2.14) by setting $f(t) = \psi(t)$ and $\Omega_1 = \Omega_2 = \Omega$, or from (2.2.15) by noting that, because of the idempotence of the spectrum of $\psi(t)$,

$$R_\psi(\tau) = 2\pi \psi(\tau) \tag{2.2.17}$$

whence (2.2.16) follows directly.

Before closing this section, we notice that, with the operators Ω and the probe functions as defined, it is possible to place an upper bound on the output of any network when the input is a probe function. In fact, if the input to Ω is

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\tau)\psi(t - \tau) \, d\tau$$

then, as usual, the output is

$$\Omega f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\tau)[\Omega\psi(t - \tau)] \, d\tau$$

and use of the Schwartz inequality in this last integral yields

$$|\Omega f(t)| \leq \left(\frac{1}{2\pi} \int f^2 \, dt\right)^{1/2} \left(\frac{1}{2\pi} \int (\Omega\psi)^2 \, dt\right)^{1/2} = \left(\frac{1}{2\pi} \int f^2 \, dt\right)^{1/2} \sqrt{\Omega^* \Omega \psi(0)} \tag{2.2.18}$$

where the last relation follows from (2.2.16). Furthermore, the bound (2.2.18) is the best obtainable, since if

$$f(t) = \Omega^* \psi(t) \tag{2.2.19}$$

then equality is actually reached, and

$$\left|\mathcal{A} f(t)\right|_{max} = \left|\mathcal{A}^{*}\mathcal{A}\psi(t)\right|_{max} = \mathcal{A}^{*}\mathcal{A}\psi(0) \qquad (2.2.20)$$

The reason for our interest in bounding the output from $\mathcal{A}$ is that in performing the thought experiment of section 2.1 we must use an oscilloscope whose trace length is finite, which implies that we must be able to guarantee a priori that the signal applied to the oscilloscope never exceeds some preassigned maximum amplitude. It is precisely because of the possibility of arriving at a relation like (2.2.18) that we chose as probes function-classes that can be given an integral characterization like (2.2.1). An important feature of this way of characterizing probe functions is that the input f(t) appears on both sides of the equation, which permits the bound (as in (2.2.18)) to be expressed in terms of a product of two numbers, of which one is characteristic of the input alone (in our case essentially the square root of the input energy) and the other is characteristic of the system alone. This form of expression will be very useful in succeeding sections.

2.3 PROBE FUNCTIONS AND OPERATORS AS ELEMENTS OF VECTOR SPACES

In this section we shall study the operators $\mathcal{A}$ as elements of a vector space. While this formalization does not add any new content to our work, it will clarify and streamline it because it will allow us to apply our geometric intuition to the study of the problem.

First, and more for the sake of orderliness than for its intrinsic interest, we notice that the set $\underline{S}_{\psi}$ of probe function, together with addition and multiplication by a (real) scalar as usually defined for

functions, constitute a linear vector space. (For the axioms of a linear vector space, see, for example, Halmos (11), von Neumann (12), Kolmogorov and Fomin (13).) This vector space, which we shall denote with the same symbol $\underline{S}_{\psi}$ as we use for the underlying set of functions, is the domain on which our operators Ω are defined, and herein lies its main interest.

For our purposes, the only interesting feature of the space $\underline{S}_{\psi}$ is that it is possible to define on it a norm (i.e., a distance function, a metric). That is, with every vector $f \in \underline{S}_{\psi}$, we can associate a number $\|f\|$, given by

$$\|f\| = \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} f^2 \, dt \right)^{1/2} \tag{2.3.1}$$

with the properties (axioms of the norm):

$$\begin{aligned} &\text{a) } \|f\| \geq 0, \text{ and } \|f\| = 0 \text{ if and only if } f = 0 \\ &\text{b) } \|af\| = |a| \cdot \|f\| \\ &\text{c) } \|f + g\| \leq \|f\| + \|g\| \end{aligned} \tag{2.3.2}$$

Proof: Property (a) follows immediately from the non-negativeness of the integrand in (2.3.1). (Actually, this only establishes that $\|f\| = 0$ if and only if $f = 0$ almost everywhere; however, it will be shown later that, for any ψ, every $f \in \underline{S}_{\psi}$ is continuous (in fact, analytic entire) so that $f = 0$ almost everywhere implies $f = 0$.) Property (b) also follows by inspection from (2.3.1). To establish (c), notice that

$$\|f+g\|^2 = \frac{1}{2\pi} \int (f+g)^2 = \frac{1}{2\pi} \int f^2 + \frac{1}{2\pi} \int g^2 + \frac{2}{2\pi} \int fg = \|f\|^2 + \|g\|^2 + \frac{2}{2\pi} \int fg$$

But, by the Schwartz inequality,

$$\frac{1}{2\pi} \left| \int fg \right| \leq \left(\frac{1}{2\pi} \int f^2 \right)^{1/2} \left(\frac{1}{2\pi} \int g^2 \right)^{1/2} = \|f\| \cdot \|g\|$$

Therefore $\| f + g \|^2 \leq \| f \|^2 + \| g \|^2 + 2 \| f \| \cdot \| g \| = \left[\| f \| + \| g \| \right]^2$ and taking square roots of both sides establishes the desired result.

Now consider the set of all finite linear combinations of shift operators, i.e., the set of objects Ω of the form $\Omega = \sum_{n=1}^{N} a_n E^{t_n}$ (where t_n is any real number). It is clear from their definition that the Ω are a set of linear operators on the space $\underline{S}_\psi$ and that, because of the class-conservation property (2.2.8), $f \in \underline{S}_\psi$ implies $(\Omega f) \in \underline{S}_\psi$. If, in the usual way we define addition $(\Omega_1 + \Omega_2)$ by

$$(\Omega_1 + \Omega_2) f = \Omega_1 f + \Omega_2 f \qquad (f \in \underline{S}_\psi)$$

and multiplication by a (real) scalar $(a\Omega)$ by

$$(a\Omega) f = a(\Omega f) \qquad (f \in \underline{S}_\psi)$$

corresponding to the physical situations shown in Fig. 2.3.1, then we obtain the usual result that the set of operators Ω, together with the operations just defined, also constitute a vector space. This space, which is the principal object of study, will be denoted by $\underline{O}$.

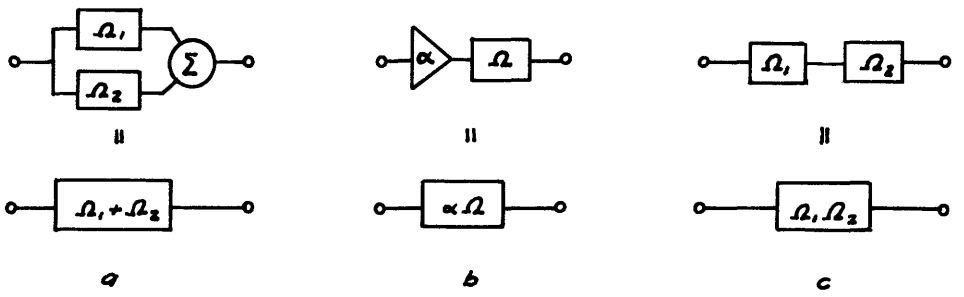

Fig. 2.3.1

Illustrating a) vector addition, b) multiplication of a vector Ω by a real scalar a, and c) multiplication of two operators (or vector product).

The reader will notice that the $\mathcal{A}$, as defined, have a dual nature, being operators on the space $\underline{S}_4$ and vectors in the space $\underline{O}$. This is convenient when we come to make analogies between our mathematical definitions and physical operations. The linear operations represented by Fig. 2.3.1a and b are "natural" operations in a linear vector space; on the other hand, there is no abstract concept of vector product, and this would leave us at a loss in attempting to obtain the mathematical equivalent of the cascading operation shown in Fig. 2.3.1c. Here it is "natural" to think of the $\mathcal{A}$ as operators because, while the sort of operation required to represent cascading is not natural within the structure of a linear vector space, it <u>is</u> a natural way of combining operators. For operators, "multiplication" corresponds simply to "successive application," and that is precisely what we mean by cascading.

In symbols, we define the <u>product</u> of $\mathcal{A}_1$ and $\mathcal{A}_2$ by

$$(\mathcal{A}_2\mathcal{A}_1)f = \mathcal{A}_2(\mathcal{A}_1 f) \qquad (f \in \underline{S}_4)$$

As was shown in sect. 2.2 (Eqs. 2.2.12, 2.2.13), when the $\mathcal{A}$ are represented as combinations of shifts, the product is formally the same as that for polynomials. It is also easy to verify that a product of $\mathcal{A}$'s corresponds to a convolution of the corresponding impulse responses. Notice, incidentally, that for time-invariant $\mathcal{A}$'s (the only kind that we shall consider in this paper) the product is commutative.

2.4 THE NORM OF $\underline{O}$

The most interesting thing about the space $\underline{O}$ is that it is possible

to define on it a metric that is quite meaningful physically. We shall first simply define it and then show, by studying its properties and by an example in the next section, what its physical meaning is.

With every vector $\mathcal{A} \in \underline{O}$ we associate a number $\|\mathcal{A}\|_\psi$ defined by

$$\|\mathcal{A}\|_\psi = \left[\frac{1}{2\pi\psi(0)} \int_{-\infty}^{\infty} [\mathcal{A}\psi(t)]^2 \, dt\right]^{1/2} = \sqrt{\frac{\mathcal{A}^+\mathcal{A}\psi(0)}{\psi(0)}} \qquad (2.4.1)$$

This number $\|\mathcal{A}\|_\psi$ is the <u>norm</u> of $\mathcal{A}$. (Note that the subscript ψ in $\|\ldots\|_\psi$ serves as a reminder of which probe ensemble is being studied, and hence which function ψ appears in the definition of the norm. The numerical values, though not the properties, of the norm are different for different ψ functions. However, since ψ is usually fixed in any one discussion, or else is arbitrary, the subscript will usually be omitted, and only be inserted when there is room for confusion.) To see that the name "norm" is deserved, we must show that $\|\mathcal{A}\|_\psi$ satisfies the following requirements:

a) $\|\mathcal{A}\|_\psi \geq 0$, and $\|\mathcal{A}\|_\psi = 0$ if and only if $\mathcal{A} = 0$

b) $\|a\mathcal{A}\|_\psi = |a| \cdot \|\mathcal{A}\|_\psi$

c) $\|\mathcal{A}_1 + \mathcal{A}_2\|_\psi \leq \|\mathcal{A}_1\|_\psi + \|\mathcal{A}_2\|_\psi$

Proof: If we note that $\mathcal{A}\psi(t)$ is just a function contained in $\underline{S}_\psi$, we can immediately use the results (2.3.2) to establish (b), (c) and the first part of (a). To see that $\|\mathcal{A}\|_\psi = 0$ if and only if $\mathcal{A} = 0$, we must prove that the functions ψ form a linearly independent set with respect to translations; that is, that for arbitrary distinct $t_1, \ldots, t_n$ there exists no set of coefficients $a_1, \ldots, a_n$ with values other than zero, and with the property that $\sum_{i=1}^{n} a_i \psi(t - t_i) \equiv 0$. This result is obtained in Appendix A.1.

The denominator in (2.4.1) was chosen so that $\|E^0\| = \|1\| = 1$.

Two features of this definition of the operator norm are interesting. First we shall present a result that justifies the use of this norm on the grounds of its physical meaningfulness, and then show that the norm induces a useful inner product.

We notice first that, in the notation of the present section, the inequality (2.2.18) becomes

$$|\Omega f(t)| \leq \|f\| \cdot \|\Omega\|_\psi \sqrt{\psi(0)} = \|f\| \cdot \|\Omega\|_\psi \cdot \|\psi\| \qquad (2.4.2)$$

If we consider all probes $f \in \underline{S}_\psi$ whose norm is one (i.e., all probes of the form $\frac{f(t)}{\|f\|}$ or, what is the same thing, all points on the unit sphere of $\underline{S}_\psi$) then we see that for all such inputs, the output of the network, $\Omega f(t)$, is small whenever $\|\Omega\|_\psi$ is small.

A sufficiently strong converse is also true: whenever the output $\Omega f(t)$ is small for <u>all</u> inputs on the unit sphere, then $\|\Omega\|_\psi$ is also small. This results from the fact that (2.4.2) was originally obtained by means of the Schwartz inequality, so that there always exists one specific function f for which (2.4.2) becomes an equality for at least one value of t. It has already been shown (2.2.19) that this function – clearly the one that gives the tallest output – turns out to be $f(t) = \Omega^* \psi(t)$. Therefore, if the output $\left| \frac{\Omega f(t)}{\|f\|} \right| < \epsilon$ for all non-zero inputs $f \in \underline{S}_\psi$, then, in particular, for the maximal input $\left| \Omega \left(\frac{\Omega^* \psi(t)}{\|\Omega^* \psi\|} \right) \right| < \epsilon$ for all values of t. Therefore, at t= 0,

$$\left| \Omega \frac{\Omega^* \psi(0)}{\|\Omega^* \psi\|} \right| = \sqrt{\psi(0)} \frac{\|\Omega\|^2}{\|\Omega\|} \quad \|\psi\| \cdot \|\Omega\| < \epsilon .$$

If now Ω is the difference between two specified operators Ω_1 and Ω_2, $\Omega = \Omega_1 - \Omega_2$, then the discussion indicates that (a) if the distance

$\|\mathcal{A}_1 - \mathcal{A}_2\|$ is small, then the difference in the outputs, $|\mathcal{A}_1 f(t) - \mathcal{A}_2 f(t)| = |(\mathcal{A}_1 - \mathcal{A}_2) f(t)|$, will also be small for all acceptable inputs; and (b) that if the difference in outputs is small for all acceptable inputs, then the distance $\|\mathcal{A}_1 - \mathcal{A}_2\|$ is small. Notice that the impulse responses corresponding to $\mathcal{A}_1$ and $\mathcal{A}_2$ can, at the same time, be wildly different.

Since the quantity $|\mathcal{A}_1 f(t) - \mathcal{A}_2 f(t)|$ is precisely what is measured by the oscilloscope in our thought experiment, the above result establishes a two-way correspondence between CRO-measurable differences and spheres in $\underline{O}$-space. That is, if a difference is too small to be measured on an oscilloscope with resolution ($\equiv$ trace-width/trace-length) ϵ, the corresponding operators must be contained in a sphere in $\underline{O}$ space of radius less than ϵ, and conversely. This result will become clearer when it is applied to a specific experiment in the next section; for the moment, we proceed to the study of another useful property of the norm, which is that it induces an inner product, or dot product.

The abstract idea of inner product (in our case real) is given in the following definition: The inner product is a real number $[\mathcal{A}_1, \mathcal{A}_2]$ associated with each pair of elements $\mathcal{A}_1$ and $\mathcal{A}_2$ of $\underline{O}$ such that:

$$
\begin{aligned}
&\text{a) } [\mathcal{A}_1, \mathcal{A}_2] = [\mathcal{A}_2, \mathcal{A}_1] \\
&\text{b) } [\mathcal{A}_1, \mathcal{A}_1] > 0 \text{ if } \mathcal{A}_1 \neq 0 \\
&\text{c) } [a\mathcal{A}_1, \mathcal{A}_2] = a[\mathcal{A}_1, \mathcal{A}_2] \text{ (where a is a real number)} \\
&\text{d) } [\mathcal{A}_1 + \mathcal{A}_2, \mathcal{A}_3] = [\mathcal{A}_1, \mathcal{A}_2] + [\mathcal{A}_2, \mathcal{A}_3]
\end{aligned}
\qquad (2.4.3)
$$

As is shown in von Neumann (12), Theorem 12.10, the necessary and sufficient condition that an inner product $[\mathcal{A}_1, \mathcal{A}_2]$ may be defined in such a way that $\sqrt{[\mathcal{A}_1, \mathcal{A}_1]} = \|\mathcal{A}_1\|$ is

$$\|\Omega_1 + \Omega_2\|^2 + \|\Omega_1 - \Omega_2\|^2 = 2\left(\|\Omega_1\|^2 + \|\Omega_2\|^2\right) \tag{2.4.4}$$

Furthermore, the appropriate inner product is unique.

Straightforward substitution of $\|\Omega\| = \sqrt{\frac{\Omega^*\Omega\Psi(0)}{\Psi(0)}}$ immediately establishes that the identity (2.4.4) is satisfied by our norm. This is the existence theorem. The inner product induced by a norm can be constructed from the norm, but it is easier simply to guess at it, the uniqueness of the correct answer being a sufficient guide in choosing the right guess. We therefore define

$$[\Omega_1, \Omega_2]_\Psi = \frac{1}{2\pi\Psi(0)} \int_{-\infty}^{\infty} [\Omega_1\Psi(t)][\Omega_2\Psi(t)]\, dt = \frac{\Omega_2^*\Omega_1\Psi(0)}{\Psi(0)} \tag{2.4.5}$$

the second equality following from the identity (2.2.14) when f(t) is replaced by $\Psi(t)$. It is immediately verified that the definition (2.4.5) has all the properties (2.4.3), the truth of (b) having already been verified in connection with the discussion of the norm, and all the rest being trivial. Therefore, $[\Omega_1, \Omega_2]$ is an inner product. Since it obviously yields the right norm, it is the unique inner product induced by the norm. The same remark as was made in connection with the definition (2.4.1) regarding the subscript Ψ in $[\ldots]_\Psi$ applies here also.

It is worth noticing, by the way, that the Schwartz inequality, which is true for any inner product, holds here also:

$$|[\Omega_1, \Omega_2]| \leq \|\Omega_1\| \cdot \|\Omega_2\| \tag{2.4.6}$$

The existence of an inner product gives us the possibility of defining orthogonality for the Ω's, and hence of constructing orthonormal bases in $\underline{O}$. More importantly, perhaps, it establishes that $\underline{O}$ is a unitary vector space and hence in many respects similar to the Euclidean space of ordinary experience, thus allowing us to use our geometric

intuition in suggesting relations and results. One of the important respects in which $\underline{O}$ differs from the Euclidean space of ordinary experience is that it is not finite-dimensional. These matters will be discussed in detail in section 2.6. Meanwhile, we stop to describe an experiment with which some of the properties of the norm are exhibited and interpreted.

2.5 AN EXPERIMENT OF CERRILLO

Consider the following experiment, performed by Dr. M. V. Cerrillo and K. Joannou. A set of M networks is prepared, whose impulse responses are the Cerrillo pure transmission kernels of orders $0, 1, \ldots, M$. These kernels are derived and discussed in Cerrillo (*7*). The kernel of order m is composed of $m+1$ contiguous, unidirectional pulses, of base-width 2λ chosen appropriately small; only the area, not the shape, of these pulses is defined, the k^{th} pulse ($k < m+1$) having an area

$$a_{k,m} = (-1)^k \frac{(m+1)!}{(k+1)!(m-k)!} = (-1)^k \binom{m+1}{k+1} \qquad (2.5.1)$$

where the last symbol represents the binomial coefficient. A few of these impulse responses are sketched in Fig. 2.5.1.

Suppose that, as was actually done in the laboratory, these networks are all excited by an arbitrary band-limited source (phonograph music was used in the experiment) and that the outputs of the networks are compared with the input and with each other by placing them, two at a time, on the vertical and horizontal plates of a laboratory oscilloscope. The arrangement is shown in Fig. 2.5.1. It turned out that,

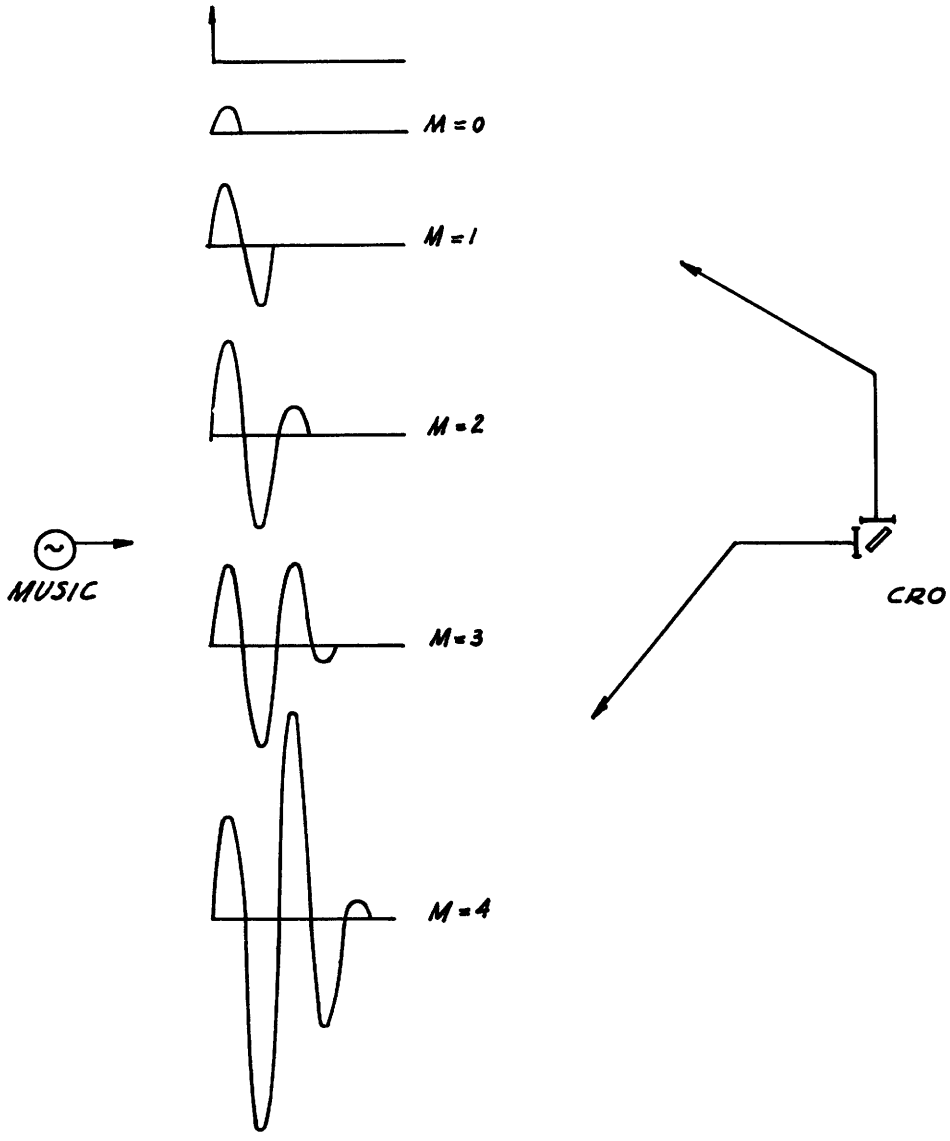

Fig. 2.5.1

Illustrating Cerrillo's experiment.

for all networks of order higher than a sufficiently large N, the network outputs could not be distinguished from the input or from each other. This result indicates, first, that Cerrillo kernels of sufficiently high order are actually capable of producing pure transmission, as far as a CRO with finite resolution can tell; and second, that as far as such an oscilloscope can tell, there is no difference, in the operation performed, between any two networks of order greater than N. (This number N can be as small as zero if the pulse width is small enough, or the oscilloscope crude enough. The relationships between pulse width, CRO resolution, bandwidth, and replaceability of pulses by impulses will be explored fully in Chapter III. For the moment, it must be assumed that the phrases "small enough," "crude enough," and so forth can be given definite meanings.)

To analyze the experiment, let us assume that (as will be shown in Chapter III) it is possible to replace the pulses in the Cerrillo kernels by impulses located at the centers of the corresponding pulses and with areas equal to those of the pulses. Then the vector corresponding to the transmission kernel of order m can be written

$$T_m = \sum_{k=0}^{m} a_{k,m} E^{-(2k+1)\lambda} = E^{-\lambda} \sum_{k=0}^{m} (-1)^k \binom{m+1}{k+1} E^{-2\lambda k} \qquad (2.5.2)$$

It can then be shown (see Appendix A.2) that the vectors T_m satisfy the difference equation

$$T_m - T_{m-1} = E^{-\lambda} \nabla_{2\lambda}^{m} \qquad (m = 1, 2, \ldots) \qquad (2.5.3)$$

with $T_o = E^{-\lambda}$. The operator $\nabla_{2\lambda}^{m}$ is the m^{th} backward difference operator, with sampling spacing 2λ, defined recursively starting from $\nabla_{2\lambda} f(t) = f(t) - f(t - 2\lambda)$; $\nabla^2 f(t) = \nabla(\nabla f(t)) = f(t) - 2f(t - 2\lambda) + f(t - 4\lambda)$; . . . ; $\nabla^n f(t) = \nabla(\nabla^{n-1} f(t))$. (See Hildebrand (14).) For consistency, ∇^0 is

defined to be the identity operator E^0, so that $\nabla^0 f(t) = E^0 f(t) = f(t)$.

Summing both sides of (2.5.3) with respect to m from 1 to n yields, since the left-hand side telescopes,

$$T_n = E^{-\lambda} \sum_{i=0}^{n} \nabla^i \tag{2.5.4}$$

a representation which is considerably more convenient than (2.5.2).

We need not go any farther than this to see, for example, that for n large enough (or λ small enough), the operation of two successive kernels can be made indistinguishable. From (2.4.2), we obtain

$$\left| T_n f(t) - T_{n-1} f(t) \right| = \left| (T_n - T_{n-1}) f(t) \right| \leq \| f \| \cdot \| T_n - T_{n-1} \| \cdot \| \psi \| \tag{2.5.5}$$

But, from (2.5.3)

$$\| T_n - T_{n-1} \| = \| E^{-\lambda} \nabla_{2\lambda}^n \| = \| \nabla_{2\lambda}^n \| \tag{2.5.6}$$

At this point, where we need definite estimates for the norms of operators, we must specialize the function $\psi(t)$ to that appropriate to the model of the actual experiment. We shall assume

$$\psi(t) = 2W \frac{\sin Wt}{Wt} \tag{2.5.7}$$

where W is the radian/sec bandwidth of the low-pass idempotent spectrum $\Phi(\omega)$. It can then be shown (see Appendix A.3) that

$$\| \nabla_{2\lambda}^n \| \leq \left(\frac{1}{2n+1}\right)^{1/2} (2\lambda W)^n \tag{2.5.8}$$

As long as $2\lambda < \frac{1}{W}$, it is therefore clearly possible to choose n and λ so as to make the bound in (2.5.5) smaller than the resolution of any preassigned oscilloscope.

As a matter of fact, we can show that, beyond some order n, any two networks (not just successive networks) are indistinguishable

if $2\lambda < \frac{1}{W}$. For suppose we wish to compare T_n and T_{n+p} (p is any positive integer). From (2.5.5) we see that we need only obtain a bound on the norm of the difference $T_{n+p} - T_n$. But, using (2.5.4)

$$T_{n+p} - T_n = E^{-\lambda} \sum_{i=n+1}^{n+p} \nabla^i$$

so that, using (2.5.8),

$$\| T_{n+p} - T_n \| = \| \sum_{i=n+1}^{n+p} \nabla^i \| \leq \sum_{i=n+1}^{n+p} \| \nabla^i \| \leq \sum_{i=n+1}^{n+p} (2\lambda W)^i$$

$$= (2\lambda W)^{n+1} \sum_{i=0}^{p-1} (2\lambda W)^i \leq \frac{(2\lambda W)^{n+1}}{1 - 2\lambda W} \quad \text{for all } p > 0. \qquad (2.5.9)$$

(It is certainly possible to obtain better bounds, but this one suffices for our present purpose. Note also that the condition $2\lambda < \frac{1}{W}$ is sufficient, but certainly not necessary.)

It is natural to wonder what it is that T_n approaches as $n \to \infty$. If our space were complete (see remarks at the beginning of section 2.6) then, since (2.5.9) shows that for $2\lambda < \frac{1}{W}$ the sequence $\{T_n\}$ is Cauchy-convergent, (2.5.9) would prove that a limit T_∞ exists. We shall, purely formally, determine what this limit T_∞ might be. (We have to proceed formally because, among other reasons, we do not even know yet how to interpret an infinite series of operators.)

Notice first that

$$(1 - \nabla) \sum_{i=0}^{\infty} \nabla^i = 1 \qquad (2.5.10)$$

because, multiplying out,

$$\sum_0^\infty \nabla^i - \nabla \sum_0^\infty \nabla^i = \sum_0^\infty \nabla^i - \sum_1^\infty \nabla^i = \nabla^0 = 1 .$$

From (2.5.10), it is therefore also true that

$$(1-\nabla)\, E^{-\lambda} \sum_{0}^{\infty} \nabla^i = E^{-\lambda} \tag{2.5.11}$$

But, from (2.5.4),

$$T_\infty = E^{-\lambda} \sum_{0}^{\infty} \nabla^i$$

so that (2.5.11) becomes

$$(1-\nabla)\, T_\infty = E^{-\lambda}$$

or, expanding

$$(1 - (1 - E^{-\lambda}))\, T_\infty = E^{-\lambda}\, T_\infty = E^{-\lambda}$$

Multiplying both sides by E^{λ} thus yields $T_\infty = 1$. That is, in the infinite limit the Cerrillo kernels produce pure transmission.

We can use this result to show that, beyond a sufficiently high order N, the outputs of the Cerrillo networks, besides being indistinguishable from each other, are indistinguishable from the input. Let us use T_∞ to represent pure transmission. Then

$$\| 1 - T_n \| = \| T_\infty - T_n \| = \| \sum_{0}^{\infty} \nabla^i - \sum_{0}^{n} \nabla^i \| = \| \sum_{n+1}^{\infty} \nabla^i \|$$

$$= \| \nabla^{n+1}\, E^{-\lambda} \sum_{0}^{\infty} \nabla^i \| = \| \nabla^{n+1}\, T_\infty \| = \| \nabla^{n+1} \|$$

Using (2.5.8), we see that for $2\lambda < \frac{1}{W}$ our statement is true.

To us, the interesting feature of this experiment is that, although the impulse responses of the networks of Fig. 2.5.1 differ enormously, for $m \geq n$ the networks are indistinguishable in their mode of operation

on music. That this is possible is not surprising, since the result obtains only for a restricted, although large, class of sources, and the oscilloscope has finite resolution. In fact, in the frequency domain, the responses of the networks are approximately equal at low-pass and differ appreciably only at frequencies higher than those of the excitation. But it raises the question: How can we describe or explicitly exhibit in the time domain the fact that, in spite of their very different impulse responses, these networks are so similar in their mode of operation that their outputs are indistinguishable ?

In terms of the metric established by our norm, the property that is common to the networks of Fig. 2.5.1 (of order higher than some n) is that they are all contained within a hypersphere in $\underline{O}$-space of radius ϵ smaller than the resolution of the oscilloscope that is used for measuring.

2.6 THE GEOMETRY OF $\underline{O}$: ORTHOGONAL BASES

In this section we shall study the geometry of our vector space $\underline{O}$. After some introductory observations concerning the structure of $\underline{O}$, we shall use the definition of inner product given in section 2.4 to obtain a meaning for orthogonality, and use this concept to construct an orthonormal set. We then prove that this set is complete with respect to our vectors Ω.

Recall that $\underline{O}$ was defined as the set of all vectors Ω that are finite linear combinations of shift operators. Despite the superficial impression of finite dimensionality caused by the wording of the definition, $\underline{O}$ is not a finite-dimensional space. The reason for this is that

we have not restricted in any way the exponents of the shift operators, so that any arbitrary shifts, not only, for example, integral multiples of some basic shift, are allowed. The fact that arbitrary shifts entail infinite dimensionality will become clearer when we study orthonormal bases in $\underline{O}$, and discover that for such a base to be complete it must have an infinite number of members. In any case, there is no satisfactory way of making a space like $\underline{O}$ finite dimensional because, as has been shown, the cascading of systems corresponds to a vector (or operator) multiplication in which the dimensionality of the product vector is, in general, one less than the sum of the dimensions of the component vectors. As a result, no matter how we fix the finite dimension, the attempt to consider the vectors under study as members of a finite-dimensional space would lead to vectors (corresponding to perfectly acceptable physical systems) that are outside the space.

The second remark is purely technical; it concerns the fact that, as defined, the space $\underline{O}$ is not complete, in the sense that a Cauchy-convergent, or fundamental, sequence in $\underline{O}$ does not necessarily have a limit in $\underline{O}$. (Convergence is defined with the metric determined by the norm.) This difficulty results from the fact that a Cauchy-convergent sequence of vectors composed of a finite number of shifts (and therefore members of $\underline{O}$) might have as limit a vector composed of an infinite number of shifts (and therefore not a member of $\underline{O}$). Such a sequence would then have no limit in $\underline{O}$. This problem is not important until we study orthonormal bases and, in any case, it is immediately obviated by the theorem (see, e.g., Kolmogorov and Fomin (13), p. 40) that every metric space has a completion, and all of its completions are

isometric. Thus we may as well consider our space $\underline{O}$ to be embedded in an appropriate completion, and ignore the whole question. However, in all questions concerning representation or completeness (of orthonormal sets) we shall remember that we are only interested in the members of $\underline{O}$, and not in those of its completion that are not in $\underline{O}$, and will therefore limit ourselves to establishing our results for members of $\underline{O}$ only.

We begin by recalling the definition (2.4.5) of the inner product

$$[\mathcal{A}_1, \mathcal{A}_2]_\psi = [\mathcal{A}_2, \mathcal{A}_1]_\psi = \frac{1}{2\pi\psi(0)} \int_{-\infty}^{\infty} [\mathcal{A}_1\psi(t)][\mathcal{A}_2\psi(t)]\, dt$$

$$= \frac{1}{\psi(0)} \mathcal{A}_2^* \mathcal{A}_1 \psi(0) = \frac{1}{\psi(0)} \mathcal{A}_1^* \mathcal{A}_2 \psi(0) \qquad (2.6.1)$$

It will be instructive to determine the meaning of orthogonality, that is, the physical situation that corresponds to two $\mathcal{A}$-vectors being orthogonal in the sense of definition (2.6.1). Notice that the very definition of orthogonality depends on the function ψ that is characteristic of a particular problem. Because various cases have to be distinguished in discussing orthogonality and completeness with arbitrary ψ-functions, the presentation will be clearer if we restrict ourselves to the case for which

$$\psi(t) = 2W \frac{\sin Wt}{Wt} \qquad (2.6.2)$$

For brevity, we shall denote this specific ψ-function by ϕ. For practical problems, this is perhaps the most important case. Substantially the same procedure as used on ϕ will establish our results for other ψ-functions.

Consider a pair of shift vectors E^{t_1} and E^{t_2}. Under what conditions

will these vectors be orthogonal ? The inner product will be zero when

$$\left[E^{t_1}, E^{t_2}\right] = \frac{1}{\phi(0)} E^{t_1^*} E^{t_2} \phi(0) = \frac{1}{\phi(0)} E^{(t_2 - t_1)} \phi(0) = \frac{1}{\phi(0)} \phi(t_2 - t_1) = 0 \quad (2.6.3)$$

The zeros of ϕ occur at the points $(t_2 - t_1) = \frac{n\pi}{W}$ $(n = \pm 1, \pm 2, \ldots)$. If we choose one of the vectors to be E^o, the set of vectors $E^{\frac{k\pi}{W}}$ $(k = \pm 1, \pm 2, \ldots)$ is therefore orthogonal to it. We shall now prove that the set $E^{\frac{k\pi}{W}}$ $(k = 0, \pm 1, \pm 2, \ldots)$ is a complete orthogonal basis in $\underline{O}$.

Let E^{-t_o} be an arbitrary shift vector, and let us see to what extent we can approximate it by a linear combination of N orthogonal vectors $E^{-\frac{k\pi}{W}}$, $k = 1, 2, \ldots, N$. That is, we wish to approximate E^{-t_o} with a vector $\mathcal{A} = \sum_{k=1}^{N} a_k E^{-\frac{k\pi}{W}}$, and we shall choose the coefficients a_k by requiring $\mathcal{A}$ to be the projection of E^{-t_o} in the linear manifold spanned by the vectors $\left\{E^{-\frac{k\pi}{W}}\right\}$, $k = 1, \ldots, N$. Accordingly, we have first that the components of E^{-t_o} along the vectors $\left\{E^{-\frac{i\pi}{W}}\right\}$ are

$$\left[E^{-t_o}, E^{-\frac{i\pi}{W}}\right] = \frac{1}{\phi(0)} E^{(t_o - \frac{i\pi}{W})} \phi(0) = \frac{1}{\phi(0)} \phi(t_o - \frac{i\pi}{W}) \quad (2.6.4)$$

Next we see that the corresponding components of $\mathcal{A}$ are

$$\left[\mathcal{A}, E^{-\frac{i\pi}{W}}\right] = \left[\sum_j a_j E^{-\frac{j\pi}{W}}, E^{-\frac{i\pi}{W}}\right] = \sum_j a_j \left[E^{-\frac{j\pi}{W}}, E^{-\frac{i\pi}{W}}\right]$$

$$= a_i \left\| E^{-\frac{i\pi}{W}} \right\|^2 \qquad \text{if} \quad i \in (1, 2, \ldots, N) \quad (2.6.5)$$

$$= 0 \qquad \text{if} \quad i \notin (1, 2, \ldots, N)$$

We shall achieve our object if we set

$$\left\| E^{-\frac{i\pi}{W}} \right\|^2 a_i = \frac{1}{\phi(0)} \phi(t_o - \frac{i\pi}{W}) \qquad \text{for} \quad i = 1, 2, \ldots, N$$

or, noting that $\left\| E^{-\frac{i\pi}{W}} \right\|^2 = \frac{1}{\phi(0)} E^{\frac{i\pi}{W}} E^{-\frac{i\pi}{W}} \phi(0) = 1$, if we set

$$a_i = \frac{\phi(t_o - \frac{i\pi}{W})}{\phi(0)} \tag{2.6.6}$$

These coefficients determine the vector Ω that we wish to use to approximate E^{-t_o}. If now the two networks Ω and E^{-t_o} are excited by a common input $f \in \underline{S}_\phi$, what is the peak difference in their outputs ? As usual, the peak difference is determined by the norm $\left\| E^{-t_o} - \Omega \right\|$, which we now proceed to evaluate.

We can use our geometric language to advantage here. Ω is the projection of E^{-t_o} onto the manifold defined by the set $\left\{ E^{-\frac{\pi}{W}}, \ldots, E^{-\frac{N\pi}{W}} \right\}$. Therefore, $E^{-t_o} - \Omega$ must be a vector perpendicular to Ω (this can be verified by taking the inner product of these two vectors). But (as is easily proved) if two vectors u and v are perpendicular, $\| u + v \|^2 = \| u \|^2 + \| v \|^2$. Therefore,

$$\left\| E^{-t_o} - \Omega + (\Omega) \right\|^2 = \left\| E^{-t_o} - \Omega \right\|^2 + \| \Omega \|^2 \equiv \left\| E^{-t_o} \right\|^2 \tag{2.6.7}$$

so that

$$\left\| E^{-t_o} - \Omega \right\|^2 = \left\| E^{-t_o} \right\|^2 - \| \Omega \|^2 \tag{2.6.8}$$

The norms on the right are given by:

$$\left\| E^{-t_o} \right\|^2 = \frac{1}{\phi(0)} E^{t_o} E^{-t_o} \phi(0) = 1$$

$$\| \Omega \|^2 = \frac{1}{\phi(0)} \left(\sum_k a_k E^{\frac{k\pi}{W}} \sum_j a_j E^{-\frac{j\pi}{W}} \right) \phi(0) = \frac{1}{\phi(0)} \sum_{k,j} a_k a_j \delta_{kj} \phi(0)$$

$$= \sum_k a_k^2 \tag{2.6.9}$$

Therefore,

$$\| E^{-t_o} - \Omega \|^2 = 1 - \sum_{k=1}^{N} a_k^2 \qquad (2.6.10)$$

Let us denote the index set of the linear manifold containing Ω by Γ. Using (2.6.6) the sum in (2.6.9) can be written more explicitly as

$$\sum_{k \in \Gamma} \left[\frac{\sin W(t_o - \frac{k\pi}{W})}{W(t_o - \frac{k\pi}{W})} \right]^2 \qquad (2.6.11)$$

The known identity

$$\sum_{k=-\infty}^{\infty} \left[\frac{\sin W(t_o - \frac{k\pi}{W})}{W(t_o - \frac{k\pi}{W})} \right]^2 \equiv 1 \qquad \text{for all } t_o$$

then shows that, as the index set Γ in (2.6.11) stretches to include all integers, $\| E^{-t_o} - \Omega \|^2 \to 0$. In the limit it must then be true that $E^{-t_o} = \Omega$ (properties of the norm, axiom (a), section 2.4). Note that in this case it must also be true that $\| E^{-t_o} \|^2 = \| \Omega \|^2 = \sum_{-\infty}^{\infty} a_k^2$ which is Parseval's Theorem for our orthogonal set.

We have shown that any shift operator can be represented without error with our orthogonal base $\left\{ E^{\frac{k\pi}{W}} \right\}$ $(k = 0, \pm 1, \pm 2, \ldots)$. Since any $\Omega \in \underline{O}$ is a finite linear combination of shifts, any $\Omega \in \underline{O}$ can be represented also. This establishes the completeness of $\left\{ E^{\frac{k\pi}{W}} \right\}$ with respect to our vector space $\underline{O}$.

Notice that the completeness of our orthogonal base is equivalent to the Shannon Sampling Theorem (Shannon (15) and (16)). To see the equivalence, suppose we wish to represent E^{t_o} with $\left\{ E^{\frac{k\pi}{W}} \right\}$. In the manifold determined by the finite index set Γ we obtain, by proceeding as before

$$\Omega = \sum_{k \in \Gamma} \frac{\phi(t_o - \frac{k\pi}{W})}{\phi(0)} E^{\frac{k\pi}{W}}$$

Then, for any $f \in \underline{S}_{\phi}$,

$$\left| E^{t_o} f(t) - \left(\sum_{k \in \Gamma} \frac{\phi(t_o - \frac{k\pi}{W})}{\phi(0)} E^{-\frac{k\pi}{W}} \right) f(t) \right| = \left| (E^{t_o} - \mathcal{A}) f(t) \right| \leq \| f \| \cdot \| E^{t_o} - \mathcal{A} \| \cdot \| \varphi \|$$

Therefore,

$$\left| f(t + t_o) - \sum_{k \in \Gamma} \frac{\phi(t_o - \frac{k\pi}{W})}{\phi(0)} f(t + \frac{k\pi}{W}) \right| \leq \| f \| \cdot \| E^{t_o} - \mathcal{A} \| \cdot \| \varphi \|$$

As Γ is stretched to cover all integers, $\| E^{t_o} - \mathcal{A} \| \to 0$ (Parseval's theorem). Therefore, in the limit,

$$f(t + t_o) = \sum_{k=-\infty}^{\infty} f(t + \frac{k\pi}{W}) \frac{\phi(t_o - \frac{k\pi}{W})}{\phi(0)}$$

At $t = 0$,

$$f(t_o) = \sum_{k=-\infty}^{\infty} f(\frac{k\pi}{W}) \frac{\phi(t_o - \frac{k\pi}{W})}{\phi(0)} \qquad (2.6.12)$$

Equation (2.6.12), with the meaning (2.6.2) substituted for ϕ, has precisely the form of Shannon's theorem. The content is the same too: (2.6.12) says that the value of any $f \in \underline{S}_{\phi}$ at any arbitrary point t_o can be determined exactly if the values of f at the points $\frac{k\pi}{W}$ $(k = 0, \pm 1, \ldots)$ are known.

Other orthogonal systems can be constructed, and other $\underline{S}_{\mathcal{A}}$ classes used in their study. However, without a specific problem in mind there is not much point in going into these further details. In the next section we study briefly a question of a different kind: what can be done with basis vectors that correspond to time-limited or nonuniform sampling.

2.7 THE GEOMETRY OF $\underline{O}$: NON-ORTHOGONAL BASES

It will be easier to visualize the following discussion if we think in terms of the impulse responses associated with our vectors. For example, if a unit impulse is applied at t= 0 to a network represented by the vector E^{-t_o}, then the corresponding impulse response consists of a unit impulse at $t = t_o$. An analogous result would of course be obtained for a vector consisting of more than one shift. Let us consider the impulse response which would correspond to a vector having unit components along all the orthogonal axes $\left\{E^{\frac{k\pi}{W}}\right\}$ if such a vector existed. This impulse response would consist of a train of unit impulses spaced regularly (with spacing $\frac{\pi}{W}$) along the time axis on both sides of the origin, and with a unit impulse at the origin corresponding to the component E^{o}. This situation is sketched in Fig. 2.7.1.

Suppose that, because of considerations of physical realizability or practical limitations, we only have available for representation purposes the portion of the time axis shown darkened in Fig. 2.7.1.

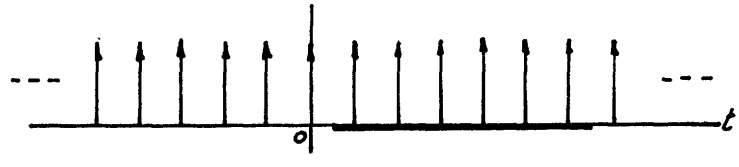

Fig. 2.7.1

The orthogonal base $\left\{E^{\frac{k\pi}{W}}\right\}$ as impulse response.

If we wish to continue restricting ourselves to the use of orthogonal vectors for representation purposes, and if at the same time we accept our physical limitations as constraints, then, from the discussion of the

last section (still restricting ourselves to probes in $\underline{S}_\phi$) we have the following situation:

Suppose we wish to represent E^{-t_o} and, in accordance with the above discussion, limit ourselves to use of the set $\{E^{-\frac{\pi}{W}}, \ldots, E^{-\frac{N\pi}{W}}\}$ for representation purposes. From (2.6.10), the error will be determined by

$$\|E^{-t_o} - \underline{\Omega}\|^2 = 1 - \sum_{k=1}^{N} a_k^2 \tag{2.7.1}$$

in which

$$\sum_{k=1}^{N} a_k^2 = \sum_{k=1}^{N} \left[\frac{\sin W(\frac{k\pi}{W} - t_o)}{W(\frac{k\pi}{W} - t_o)}\right]^2 = 1 \quad \text{if} \quad t_o = j\frac{\pi}{W},\ j = 1, 2, \ldots, N \tag{2.7.2}$$

$$= 0 \quad \text{if} \quad t_o = j\frac{\pi}{W},\ j = 0, -1, -2, \ldots \quad N+1, N+2, \ldots \tag{2.7.3}$$

and $\sum a_k^2$ is greater than zero for any other t_o (the first assertion is obvious, since then none of the terms of the sum is zero, and the second follows from Parseval's theorem). What does this mean ? If t_o has one of the values in (2.7.2) then E^{-t_o} belongs to our set and of course we can represent it without error; if t_o has one of the values in (2.7.3) then it is orthogonal to our set, and we can't represent it at all; and if t_o has any other value, then, although E^{-t_o} is not orthogonal to our set ($\sum a_k^2 > 0$), it is linearly independent of our set ($\sum a_k^2 < 1$, whence error > 0). This means, for example, that shift vectors whose impulses are located between those of members of our set, being linearly independent of our set, must have components elsewhere. That is, they must have components in the part of $\underline{O}$ that is outside the manifold spanned by our orthogonal vectors, and therefore, with those components,

must be bringing in information from outside our manifold, as sketched in Fig. 2.7.2. Notice that when a vector E^{-t_o} as shown in Fig. 2.7.2

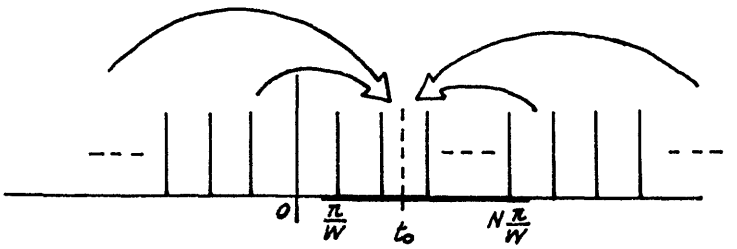

Fig. 2.7.2

The effect of dense sampling.

is added to our original set $E^{-\frac{\pi}{W}}, \ldots, E^{-\frac{N\pi}{W}}$, there are no outside pure shift vectors perpendicular to <u>all</u> the vectors of our new set, so that the error in representing vectors at the positions (2.7.3) is immediately reduced.

This demonstrates the obvious fact that by increasing the density of sampling, we can reduce the approximation error. As the density is increased indefinitely, there are two possibilities: either the error approaches a constant greater than zero, so that an irreducible error exists, or the error tends to zero. The second possibility can in fact be realized. This follows from the two observations:

1. Any function belonging to a ψ-class is an analytic entire function of t, when t is considered a complex variable (see Appendix A.4 for proof; for the special case $\psi = \phi$, this is also noted in Whittaker (17)).
2. The Wierstrass theorem asserts that knowledge of the samples of an analytic function over any countable set containing a limit point specifies the function uniquely over

the region of analyticity. Since any bounded countable set contains at least one limit point (Bolzano-Wierstrass theorem), this shows that time-limited, dense sampling can, in the limit, specify any probe function uniquely for all values of t.

This implies that we can represent an arbitrary shift vector with arbitrary accuracy using vectors corresponding to a time-limited impulse response. It therefore also implies that we can represent any $\mathcal{A} \in \underline{O}$, with arbitrary accuracy, so that it is possible to obtain vector sets, with time-limited impulse responses, that are complete in $\underline{O}$. This result has, however, little practical value since, as will be shown in Chapter III, when the measurement process has finite resolution (caused either by instrument limitations or noise) there exists a finite density level such that no new information is made perceptible by increasing the density beyond that level.

The actual mechanics of expanding an arbitrary vector in terms of a non-orthogonal (but linearly independent) set of vectors $E_1, \ldots, E_N$ is simple in principle, but involved in practice. The error expression can also be written down immediately, but no general answer has been obtained to the most interesting practical question, which is: How fast does the error decrease with increasing density ?

To represent E^{t_o} in terms of the set $E_1, \ldots, E_N$, we determine coefficients a_i such that $\sum_{i=1}^{N} a_i E_i$ equals the projection of E^{t_o} in the manifold determined by $\{E_1, \ldots, E_N\}$. These coefficients are clearly specified by the set of linear simultaneous equations

$$[E^{t_o}, E_j] = \sum_{i=1}^{N} a_i [E_i, E_j] \qquad j = 1, \ldots, N \tag{2.7.4}$$

To obtain the a's it is therefore necessary to invert the $N \times N$ matrix whose elements are the inner products $[E_i, E_j] = \frac{\varphi(t_i - t_j)}{\varphi(0)}$. Since it is still true that the error will be determined by

$$\| E^{t_o} - \sum_i a_i E_i \|^2 = \| E^{t_o} \|^2 - \| \sum_i a_i E_i \|^2 \tag{2.7.5}$$

we must determine the second term in (2.7.5):

$$\| \sum_i a_i E_i \|^2 = \frac{1}{\varphi(0)} \sum_i a_i E_i^* \sum_j a_j E_j \varphi(0) = \sum_i \sum_j a_i a_j [E_i, E_j]$$

$$= \underline{a} \left[[E_i, E_j] \right] a \tag{2.7.6}$$

These results will be used in section 3.6.

CHAPTER III

DENSITY THEOREMS AND APPLICATIONS

We have two objectives in this chapter: first, to study the relation between the present model and the functional model, especially in the limit as they approach each other; and second, to illustrate the use of the present model and its algebra in the solution of some physical problems. The first goal is reached in obtaining a theorem of, roughly speaking, the type: singular systems are dense in the space of all systems. This result suggests that, since finite resolution prevents us from distinguishing systems that are closer than some determinable finite distance, we can, whenever it is easier to do so, use the present model to solve problems. Scattered throughout the chapter, but mainly in the last section, we present some examples of such questions and the simplicity with which their answers are obtained. One of these answers, the one concerned with least-peak-error interpolation and extrapolation, is a rather useful result in itself.

3.1 PREPARATORY RESULTS: THE EFFECT OF SMALL SHIFTS

In this section we present some lemmas which are not particularly meaningful in themselves, but are useful in the sequel.

We wish to determine the effect, at the output of a singular network, resulting from shifting the impulse response of the network by a small amount λ, as in Fig. 3.1.1. We note, to start, that shifting an impulse response by an amount λ (further into the past) corresponds to multiplying the associated vector

$$\Omega = \sum_n a_n E^{-t_n} \tag{3.1.1}$$

by $E^{-\lambda}$, thus obtaining

$$E^{-\lambda}\Omega = \Omega E^{-\lambda} = \sum_n a_n E^{-(t_n+\lambda)} \tag{3.1.2}$$

Let us denote $E^{-\lambda}\Omega$ by Ω_1. Our problem then is to place a bound on $|\Omega f(t) - \Omega_1 f(t)| = |(\Omega - \Omega_1) f(t)|$. We shall, in this section and the

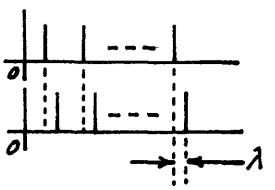

Fig. 3.1.1

Impulse response shifted an amount λ.

next, always present two such bounds: one which is useful for the numerical determination of the relative error, but is somewhat conservative, and another which is more complicated, but has the advantage of being a "best possible" bound, in the sense that we shall, in every case, exhibit an $f \in \underline{S}_\psi$ that makes the inequality an equality for at least one value of t. Both forms will be useful in the work to follow.

FIRST BOUND. Starting from $|(\Omega - \Omega_1) f(t)|$, we obtain, as usual,

$$|(\Omega - \Omega_1) f(t)| \leq \|f\| \cdot \|\psi\| \cdot \|\Omega - \Omega_1\| \tag{3.1.3}$$

so that our problem is to determine the distance $\|\Omega - \Omega_1\|$ from the original to the new (shifted) vector Ω_1. In terms of the inner product, the distance is given by

$$\|\Omega - \Omega_1\|^2 = [\Omega - \Omega_1, \Omega - \Omega_1] = [\Omega, \Omega] + [\Omega_1, \Omega_1] - [\Omega, \Omega_1] - [\Omega_1, \Omega] \tag{3.1.4}$$

Since the norm is unaffected by a translation,

$$[\Omega_1, \Omega_1] = \|\Omega_1\|^2 = \|\Omega\|^2 = [\Omega, \Omega]$$

and, because of the symmetry of the inner product, the last two terms in (3.1.4) are equal, so that

$$\|\Omega - \Omega_1\|^2 = 2(\|\Omega\|^2 - [\Omega, \Omega_1]) . \tag{3.1.5}$$

From the definition (2.4.8) of the inner product, we see that

$$[\Omega, \Omega_1] = \frac{1}{2\pi\psi(0)} \int_{-\infty}^{\infty} [\Omega\psi(t)][\Omega\psi(t-\lambda)]\, dt = \frac{1}{2\pi\psi(0)} R_{\Omega\psi}(\lambda) \tag{3.1.6}$$

where $R_{\Omega\psi}(\lambda)$ is the autocorrelation function of $\Omega\psi$ evaluated at λ. Since $\lambda = 0$ corresponds to the special case $\Omega = \Omega_1$, we have also

$$[\Omega, \Omega] = \|\Omega\|^2 = \frac{1}{2\pi\psi(0)} R_{\Omega\psi}(0) .$$

Using these results in (3.1.5), we have

$$\|\Omega - \Omega_1\|^2 = \frac{2}{2\pi\psi(0)} \left[R_{\Omega\psi}(0) - R_{\Omega\psi}(\lambda)\right] \tag{3.1.7}$$

To obtain an explicit bound in terms of λ, we must now specialize the function ψ. If we choose

$$\psi(t) = 2W \frac{\sin Wt}{Wt} ,$$

which corresponds to the low-pass idempotent spectrum of radian bandwidth W, (i.e., spectrum nonzero only on (-W, W)) then it has been shown (see Wernikoff (22), p. 38) that

$$R_{\Omega\psi}(0) - R_{\Omega\psi}(\lambda) \leq R_{\Omega\psi}(0)\, (W\lambda)^2 \tag{3.1.8}$$

Substituting in (3.1.7), we obtain

$$\|\Omega - \Omega_1\|^2 \leq \frac{2}{2\pi\psi(0)} R_{\Omega\psi}(0)\, (W\lambda)^2 = 2\|\Omega\|^2 (W\lambda)^2$$

Substituting back in (3.1.3) we therefore find that, for the chosen $\psi(t)$, the difference in output from an operator Ω and the same operator shifted by an amount λ is bounded by

$$|(\Omega - \Omega_1) f(t)| \leq \|f\| \cdot \|\psi\| \cdot \|\Omega\| \cdot \sqrt{2}\, W |\lambda| \tag{3.1.9}$$

This is the desired first result. It is conservative for any ψ function whose highest cutoff frequency does not exceed W. The first three factors in (3.1.9) define, essentially, the maximum output from Ω. We can normalize with respect to them, and find that the relative error due to a shift λ is bounded by $\sqrt{2}\, W\, |\lambda|$.

SECOND BOUND. We can obtain a sharper bound (in fact, the best possible bound) from the following calculation:

$$|(\Omega - \Omega_1) f(t)| = |\Omega(1 - E^{-\lambda}) f(t)| = |\Omega \nabla_\lambda f(t)|$$

$$= |\nabla_\lambda (\Omega f(t))| \leq \|\Omega f\| \cdot \|\psi\| \cdot \|\nabla_\lambda\| \tag{3.1.10}$$

In (3.1.10), the last step follows from considering $\Omega f(t)$ as a probe, used as input to ∇_λ. The last factor can be determined exactly, from the definition of the norm (2.4.1):

$$\|\nabla_\lambda\|^2 = \frac{\nabla_\lambda^* \nabla_\lambda \psi(0)}{\psi(0)} = \frac{(1 - E^{-\lambda})^* (1 - E^{-\lambda}) \psi(0)}{\psi(0)} = \frac{(1 - E^{\lambda})(1 - E^{-\lambda}) \psi(0)}{\psi(0)}$$

$$= \frac{(2 - E^{\lambda} - E^{-\lambda}) \psi(0)}{\psi(0)} = \frac{2}{\psi(0)} [\psi(0) - \psi(\lambda)] \tag{3.1.11}$$

Substituting the result in (3.1.10), we obtain

$$|(\Omega - \Omega_1) f(t)| \leq \|\Omega f\| \cdot \|\psi\| \cdot \left[\frac{2}{\psi(0)} (\psi(0) - \psi(\lambda)) \right]^{1/2}$$

Therefore

$$|(\Omega - \Omega_1) f(t)| \leq \|\Omega f\| \cdot \|\psi\| \left[2 \left(1 - \frac{\psi(\lambda)}{\psi(0)} \right) \right]^{1/2} \tag{3.1.12}$$

This is the desired result. It is the best general result in the sense that it is possible to find an operator Ω and a probe f(t) that make the inequality (3.1.12) an equation for at least one value of t. In fact, it is easily verified from (3.1.10) that (3.1.12) is an equality when

$$\Omega = \nabla_\lambda E^\gamma \quad \text{and} \quad f(t) = \psi(t+\delta) \tag{3.1.13}$$

where γ and δ are arbitrary. To see this, recall that (3.1.10) was originally derived from the Schwartz inequality, which reaches equality when the component integrands become equal; this is the condition brought about by (3.1.13).

Let us evaluate the bound (3.1.12), under the maximizing condition (3.1.13), to compare it with our first bound under the same conditions. Substituting (3.1.13) in (3.1.9) yields, for the right-hand side,

$$\|\psi\|^2 \, \|\nabla_\lambda^*\| \sqrt{2}\, W|\lambda| \tag{3.1.14}$$

On the other hand, (3.1.12) yields, noticing that $\|\Omega f\|^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty}[\nabla_\lambda \psi]^2\, dt$ $= \psi(0)\,\|\nabla_\lambda\|^2$, and that $\|\psi\| = \sqrt{\psi(0)}$,

$$\|\psi\|^2 \, \|\nabla_\lambda\| \cdot \left[2\left(1 - \frac{\psi(\lambda)}{\psi(0)}\right)\right]^{1/2} \tag{3.1.15}$$

Equation (3.1.9) was determined for $\psi(t) = 2W\,\frac{\sin Wt}{Wt}$, so we must determine (3.1.15) for the same case. Noticing that

$$1 - \frac{\psi(\lambda)}{\psi(0)} = \frac{(W\lambda)^2}{3!} - \frac{(W\lambda)^4}{5!} + \ldots - \ldots$$

we obtain

$$\left[2\left(1 - \frac{\psi(\lambda)}{\psi(0)}\right)\right]^{1/2} = \left[\frac{(W\lambda)^2}{3} - \frac{(W\lambda)^4}{5.4.3} + \ldots\right]^{1/2} = \sqrt{\frac{(W\lambda)^2}{3}}\left[1 - \frac{(W\lambda)^2}{5.4} + \frac{(W\lambda)^4}{6.5.4} - \ldots\right]^{1/2} \tag{3.1.16}$$

If $|W\lambda| < 1$ (this is what we mean by the phrase "λ is a small shift") then the terms of the series in the second factor decrease monotonically in absolute value and, from the elementary theory of alternating series (see, e.g., Knopp (32), p. 250) we obtain

$$\sqrt{\frac{(W\lambda)^2}{3}}\left[1 - \frac{(W\lambda)^2}{5.4}\right]^{1/2} \leq \left[2\left(1 - \frac{\psi(\lambda)}{\psi(0)}\right)\right]^{1/2} \leq \sqrt{\frac{(W\lambda)^2}{3}} \tag{3.1.17}$$

This shows that if we estimate the middle term with $\frac{W|\lambda|}{\sqrt{3}}$, we shall be off by less than 3 per cent for $W\lambda < 1$. Substituting in (3.1.15), we can compare with (3.1.14) and find that our first bound is a factor of $\sqrt{6} = 2.4$ larger than the best bound. Thus we have gained some idea of the quality of our bound.

3.2 APPROXIMATION THEOREMS

So far we have always restricted our analysis to singular networks; in this section we shall begin to attempt to decide whether this is only a special case, or whether, as in fact turns out to be the case, the study of singular networks is as general as that of more arbitrary linear systems. Intuitively, we feel that, using instruments of limited resolution on specified classes of signals, it must be impossible, beyond a certain point, to decide in the laboratory whether the black box under study should be modelled by a singular or a functional impulse response. A striking physical example of this situation is provided in the study of ionospheric multipath transmission problems. There, with a transmitter of essentially limited bandwidth and power, and in the inescapable presence of noise, it ultimately becomes impossible to decide whether the transmission medium, considered as a linear system, consists

of a finite number of closely spaced, discrete paths, or of a continuum of paths.

Our ultimate problem, more specifically, is to determine how the two models, singular and smooth responses, are related; and to do this in terms of the spacings, the probe-class parameters, and the resolution appropriate to a given situation. We shall give our present results in the form of bounds on the maximum difference in the outputs from the singular network and its corresponding smooth network. As before, two types of bound will be presented: the first very simple analytically, capable of yielding a relative error, but conservative; the second less simple, but sharper. These derivations will be followed by examples and physical interpretations.

FIRST BOUND. We shall proceed as follows: Using the results of the last section, which specify what happens to the output when a singular response is shifted slightly, we shall determine what happens with a combination of such shifts, as in Fig. 3.2.1, and then use this technique to build up pulses out of impulses. Thus we shall obtain

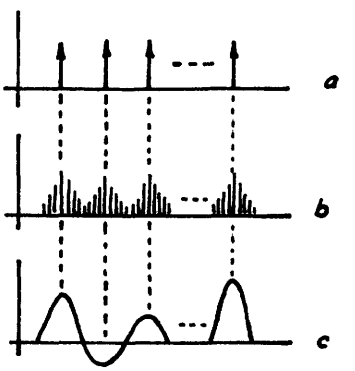

Fig. 3.2.1

Derivation of pulses from impulses.

smooth networks, or simple-function approximations to smooth networks, from impulses.

We start with an operator $\mathcal{A}$. Consider, instead, the operator $\mathcal{A}_i = a_i \mathcal{A}$ (a_i, a real number) and shift $\mathcal{A}_i$ by an amount λ_i. Then, from (3.1.9), we have *

$$\left| (\mathcal{A}_i - E^{-\lambda_i \mathcal{A}_i}) f(t) \right| \leq \|f\| \cdot \|\psi\| \cdot \|\mathcal{A}_i\| \sqrt{2}\, |\lambda_i| W \tag{3.2.0}$$

and since $\|\mathcal{A}_i\| = |a_i| \cdot \|\mathcal{A}\|$,

$$\left| (\mathcal{A}_i - E^{-\lambda_i \mathcal{A}_i}) f(t) \right| \leq \|f\| \cdot \|\psi\| \cdot \|\mathcal{A}\| \cdot \sqrt{2}\, |a_i \lambda_i| W .$$

Suppose that this is done several times (i.e., for a sequence of values of i) and that the results are added:

$$\sum_i \left| a_i \mathcal{A} f(t) - a_i \mathcal{A} f(t - \lambda_i) \right| \leq \sqrt{2} \|f\| \cdot \|\psi\| \cdot \|\mathcal{A}\| \cdot W \sum_i |\lambda_i a_i| \tag{3.2.1}$$

Since it is always true that

$$\left| \sum_i c_i \right| < \sum_i |c_i| , \tag{3.2.2}$$

equation (3.2.1) implies that

$$\left| \sum_i a_i \mathcal{A} f(t) - \sum_i a_i \mathcal{A} f(t - \lambda_i) \right| \leq \sqrt{2} \|f\| \cdot \|\psi\| \cdot \|\mathcal{A}\| \cdot W \sum_i |\lambda_i a_i| . \tag{3.2.3}$$

Now suppose that our purpose is to convert each impulse of the original response into the pulse $p(\lambda)$ (origin of λ at each impulse). Let $a_i = p(\lambda_i) \Delta\lambda$, so that an impulse shifted to the position λ_i has its area multiplied by a number a_i which is proportional to the desired pulse height at that point. Then (3.2.3) becomes

$$\left| \mathcal{A} f(t) \left[\sum_i p(\lambda_i) \Delta\lambda \right] - \mathcal{A} \left[\sum_i p(\lambda_i)\, f(t - \lambda_i) \Delta\lambda \right] \right| \leq \sqrt{2} \|f\| \cdot \|\psi\| \cdot \|\mathcal{A}\| \cdot W \cdot \sum_i |\lambda_i\, p(\lambda_i)| \Delta\lambda$$

---

* Notice that by using (3.1.9) we have already restricted ourselves to the case $\psi(t) = 2W \frac{\sin Wt}{Wt}$.

In the limit, as the number of shifts increases and $\Delta\lambda \to 0$, the sums tend to Riemann integrals; and, since it is easily shown that the inequality is preserved in the course of the limit process, we obtain

$$\left| \left[\int_{\Lambda} p(\lambda)\, d\lambda\right] \Omega f(t) - \Omega \left[\int_{\Lambda} p(\lambda)\, f(t-\lambda) d\lambda\right] \right| \leq \sqrt{2}\, \|f\| \cdot \|\Psi\| \cdot \|\Omega\| \cdot W \cdot \int_{\Lambda} |\lambda\, p(\lambda)|\, d\lambda \,. \tag{3.2.4}$$

In (3.2.4), Λ is the base or support of the pulse $p(\lambda)$. This is the desired bound.

To interpret (3.2.4), let us first normalize our pulses $p(\lambda)$ so that $\int_{\Lambda} p(\lambda)\, d\lambda = 1$. Then the first term on the left-hand side is just $\Omega f(t)$, which is the output from the singular network Ω. The second term,

$$\Omega \int_{\Lambda} p(\lambda)\, f(t-\lambda)\, d\lambda \,, \tag{3.2.5}$$

is just the output of the smooth network that is obtained from Ω by replacing all its impulses with pulses in such a way that corresponding impulses and pulses have the same area. To see this, let $\Omega = \sum_{k=1}^{N} a_k E^{-t_k}$; then the corresponding smooth response, denoted by $h(t)$, can be written

$$h(t) = \sum_{k=i}^{N} a_k\, p(t - t_k) = \left[\sum_{k} a_k\, E^{-t_k}\right] p(t) = \Omega p(t) \tag{3.2.6}$$

The response from $h(t)$ to an input $f(t)$ is given by

$$\int h(\lambda)\, f(t-\lambda)\, d\lambda = \int \Omega p(\lambda)\, f(t-\lambda)\, d\lambda = \Omega \int p(\lambda)\, f(t-\lambda)\, d\lambda$$

which is just the expression (3.2.5) since $p(\lambda) \equiv 0$ outside the interval Λ.

Before continuing to the derivation of the second type of bound, we pause briefly to examine some of the content of the present result.

Notice first that if $p(\lambda)$ is a unidirectional pulse (i.e., a pulse that does not change sign), if it is normalized, and if it extends from

$-\lambda_o$ to λ_o, then in (3.2.4) the last factor becomes

$$\int_{\mathcal{A}} |\lambda\, p(\lambda)|d\lambda = \int_{-\lambda_o}^{\lambda_o} |\lambda\, p(\lambda)|\, d\lambda < \lambda_o \int_{-\lambda_o}^{\lambda_o} |\, p(\lambda)|d\lambda = \lambda_o \left| \int_{-\lambda_o}^{\lambda_o} p(\lambda)d\lambda \right| = \lambda_o \tag{3.2.7}$$

so that for this case, (3.2.4) becomes

$$\left| \mathcal{A}f(t) - \mathcal{A}\int_{-\lambda_o}^{\lambda_o} p(\lambda)\, f(t-\lambda)\, d\lambda \right| < \sqrt{2}\, \|f\| \cdot \|\mathcal{A}\| \cdot \|\mathcal{A}\|\ W\lambda_o \tag{3.2.8}$$

so that the bound on the error is independent of pulse shape.

We remark, parenthetically, that since unidirectional pulses are precisely what Cerrillo calls "windows," (3.2.6) shows that it is possible to associate a singular response with every window function, with a very clear estimate of the error incurred by this procedure (or, with a conservative estimate of how small the probe bandwidth W must be in order to make the right-hand side of (3.2.6) less than the prescribed tolerable error). Furthermore, since a singular response is just a numerical operator, the procedure shows how to use the results of numerical analysis to obtain the Cerrillo kernel appropriate to a given operation directly. (This topic is more fully discussed in Wernikoff (23), p. 44.) More generally, since the pulse $p(\lambda)$ may be chosen rectangular, (3.2.4) or (3.2.6) show the relationship between singular responses and simple-function, or staircase, responses. Since it is known that the set of simple functions is uniformly dense in the set of time-limited continuous functions, these equations show that it is possible to associate a singular response with any continuous (time-limited) response, and that this can be done with arbitrarily small error at the output, for all members of a specified ensemble of inputs. This last possibility

will be discussed more fully later.

Returning now to a consideration of the first bound (3.2.4), we notice that there are two reasons for suspecting that it is rather conservative: 1) the result that ultimately determined the bound entered the derivation as early as (3.2.0), before anything at all could be said about the pulse we wanted to construct, so that no use could be made of the pulse information; and 2) the use of (3.2.2), a crude inequality and, in our context, a somewhat unnatural one because there is no reasonable set of circumstances that make the inequality an equality.

So much for the simple bound. Now we proceed to obtain its corresponding "best" bound.

SECOND BOUND.

To avoid the difficulties in the first derivation, we shall attempt to go from (a) to (b) in Fig. 3.2.1 directly, without considering, as before, the error due to individual displacements. That is (still with reference to Fig. 3.2.1), we shall consider each aggregate of impulses in (b) associated with an impulse in (a) as an operator in its own right, and denote it by

$$\mathcal{A}_p = \sum_{i=1}^{N} p(\lambda_i)\, \Delta\lambda\, E^{-\lambda_i} . \tag{3.2.9}$$

We shall then suppose that each impulse of the original operator $\mathcal{A}$ has been replaced by $\mathcal{A}_p$ multiplied by the corresponding area. It is clear that the new operator $\mathcal{A}_1$ depicted in (b) is given by

$$\begin{aligned} \mathcal{A}_1 = \mathcal{A}\mathcal{A}_p &= \sum_n a_n E^{-t_n} \sum_i p(\lambda_i)\, \Delta\lambda\, E^{-\lambda_i} \\ &= \sum_n a_n \sum_i p(\lambda_i)\, \Delta\lambda\, E^{-(t_n+\lambda_i)} \end{aligned} \tag{3.2.10}$$

We are, as usual, interested in the difference

$$\left|(\mathcal{A} - \mathcal{A}_1)\, f(t)\right| = \left|(\mathcal{A} - \mathcal{A}\mathcal{A}_p)\, f(t)\right| = \left|\mathcal{A}(1 - \mathcal{A}_p)\, f(t)\right|$$

$$\leq \|\mathcal{A} f\| \cdot \|\psi\| \cdot \|1 - \mathcal{A}_p\| \tag{3.2.11}$$

Our problem is to determine the last term in the product, that is, to evaluate

$$\|1 - \mathcal{A}_p\|^2 = \frac{1}{2\pi\psi(0)} \int_{-\infty}^{\infty} \left[(1 - \mathcal{A}_p)\psi(t)\right]^2 dt \tag{3.2.12}$$

If we apply our usual identity (2.4.1), we shall obtain a determination in the time domain that turns out to be somewhat too unwieldy for use. On the other hand, it is possible to obtain a very succinct frequency-domain expression. Using the convention

$$f(t) = \int_{-\infty}^{\infty} F(\omega)\, e^{j\omega t}\, d\omega\,; \qquad F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)\, e^{-j\omega t}\, dt \tag{3.2.13}$$

the Parseval theorem for Fourier transforms is

$$\int_{-\infty}^{\infty} f^2(t)\, dt = 2\pi \int_{-\infty}^{\infty} |F(\omega)|^2\, d\omega \tag{3.2.14}$$

Making the identification

$$f(t) = (1 - \mathcal{A}_p)\psi(t) = \psi(t) - \sum_i p(\lambda_i)\, \Delta\lambda\, \psi(t - \lambda_i)$$

we obtain

$$F(\omega) = \Psi(\omega)\left[1 - \sum_i p(\lambda_i)\, \Delta\lambda\, e^{-j\omega\lambda_i}\right]$$

whence, using Parseval's theorem in (3.2.12), we have

$$\|1 - \mathcal{A}_p\|^2 = \frac{1}{2\pi\psi(0)}\, 2\pi \int_{-\infty}^{\infty} \Psi(\omega) \left|1 - \sum_i p(\lambda_i)\, e^{-j\omega\lambda_i}\, \Delta\lambda\right|^2 d\omega \tag{3.2.15}$$

Substituting back in (3.2.11), and remembering that $\psi(0) = \|\psi\|^2$, we obtain

$$\left| \mathcal{A}f(t) - \mathcal{A}\sum_i p(\lambda_i) f(t-\lambda_i) \Delta\lambda \right| \leq \|\mathcal{A}f\| \left[\int_{-\infty}^{\infty} \Psi(\omega) \left| 1 - \sum_i p(\lambda_i) e^{-j\omega\lambda_i} \Delta\lambda \right|^2 d\omega \right]^{1/2} \tag{3.2.16}$$

The integral inside the bracket in (3.2.16) is infinite only in appearance; actually, it can only extend over a set of finite total length because $\Psi(\omega)$ is both idempotent and L_2 by assumption. There is no difficulty now in performing the limit process in which

$$\sum_i p(\lambda_i) f(t-\lambda_i) \Delta\lambda \longrightarrow \int_{\Lambda} p(\lambda) f(t-\lambda) d\lambda$$

$$\sum_i p(\lambda_i) e^{-j\omega\lambda_i} \Delta\lambda \longrightarrow \int_{\Lambda} p(\lambda) e^{-j\omega\lambda} d\lambda \equiv P(\omega) \tag{3.2.17}$$

$P(\omega)$ is 2π times the Fourier transform of $p(\lambda)$, since $p(\lambda) \equiv 0$ outside of Λ. Since, in (3.2.16), $\Psi(\omega)$ is nothing more than the characteristic function of the frequency-set on which our probe functions have energy, we can omit it and just indicate with the symbol $\int_X$ the limited range of integration. With these changes, our results, substituted into (3.2.16), yield

$$\left| \mathcal{A}f(t) - \mathcal{A}\int_{\Lambda} p(\lambda) f(t-\lambda) d\lambda \right| \leq \|\mathcal{A}f\| \cdot \sqrt{\int_X |1 - P(\omega)|^2 d\omega} \tag{3.2.18}$$

which is the desired result.

It is instructive to examine this expression carefully because, as will be proved presently, (3.2.18) is the best possible general result, in the sense that, given a desired arbitrary $p(\lambda)$, it is always possible to find an operator $\mathcal{A} \in \underline{O}$ and a probe $f \in \underline{S}_{\Psi}$ that make the inequality an equality for at least one value of t. Thus no sharper or smaller general upper bound is possible: relation (3.2.18) specifies the least upper bound.

We notice first that (3.2.18) has a very good chance of being an improvement on the previous expressions because it predicts that the error can be made identically zero by a very reasonable choice of pulse shape $p(\lambda)$. If, for example, in (3.2.18) we set $P(\omega) = 1$ for $\omega \in X$ and zero outside, so that $p(\lambda) = \psi(\lambda)$, we see that the error vanishes identically. This is reasonable because $p(\lambda)$ is the function that is being used to replace impulses in the impulse response of Ω, so that for probes $f \in \underline{S}_{\psi}$, $\psi(\lambda)$ is the impulse as they see it. Therefore, there should be no difference in the outputs from the singular network and the smooth network obtained in this way. More generally, of course, $p(t)$ will yield zero error if it is any L_2 function such that $p(t) = \psi(t) + g(t)$, where the Fourier transform $G(\omega)$ of $g(t)$ vanishes identically for $\omega \in X$ and is completely arbitrary (except for being L_2) on the rest of the real line.

It is perhaps worth emphasizing that no requirements (except integrability) have ever been placed on the pulses $p(\lambda)$, or on the spacing of the impulses in the response of Ω. Both are completely arbitrary, and pulses associated with various impulses may overlap. In the figures the pulses were drawn narrow and the impulses were spaced uniformly only for graphical simplicity.

To prove that (3.2.18) is in fact the <u>least</u> upper bound, we shall show that the operator E^{γ} (γ arbitrary) and the probe function

$$f_o(t) = \psi(t) - \int_{\Lambda} p(\lambda)\, \psi(t+\lambda)\, d\lambda \qquad (3.2.19)$$

make (3.2.18) an equality when $t = 0$. (It is clear that $E^{\gamma} \in \underline{O}$. To see that $f_o \in \underline{S}_{\psi}$, we notice that it is the sum of two terms of which the

first is in $\underline{S}_\psi$, and the second is also, as can be seen most easily by noticing that, whatever be $p(\lambda)$, the Fourier transform of the second term contains $\Psi(\omega)$ as a factor, and is therefore frequency-limited in the same way that $\psi(t)$ is.)

We begin by evaluating the left-hand side of (3.2.18):

$$\left| f(t') - \Omega \int_\Lambda p(\lambda) f(t'-\lambda)\, d\lambda \right| = \left| f_o(t) - \int_\Lambda p(\lambda) f_o(t-\lambda)\, d\lambda \right| \qquad (t = t' + \gamma)$$

$$= \left| \psi(t) - \int_\Lambda p(\lambda)\psi(t+\lambda)\, d\lambda - \int_\Lambda p(\lambda)\left[\psi(t-\lambda) - \int_\Lambda p(\xi)\psi(t-\lambda+\xi)\, d\xi\right] d\lambda \right|$$

$$= \left| \psi(t) - \int_\Lambda p(\lambda)\psi(t+\lambda)\, d\lambda - \int_\Lambda p(\lambda)\psi(t-\lambda)\, d\lambda + \int_\Lambda d\lambda\, p(\lambda) \int_\Lambda d\xi\, p(\xi)\psi(t-\lambda+\xi) \right| \qquad (3.2.20)$$

Let us express the various terms of (3.2.20) in terms of their Fourier transforms:

$$\psi(t) = \int_{-\infty}^{\infty} \Psi(\omega)\, e^{j\omega t}\, d\omega = \int_X e^{j\omega t}\, d\omega \qquad (3.2.21a)$$

$$\int_\Lambda p(\lambda)\psi(t+\lambda)\, d\lambda = \int_{-\infty}^{\infty} P^*(\omega)\Psi(\omega)\, e^{j\omega t}\, d\omega = \int_X P^*(\omega)\, e^{j\omega t}\, d\omega \qquad (3.2.21b)$$

$$\int_\Lambda p(\lambda)\psi(t-\lambda)\, d\lambda = \int_{-\infty}^{\infty} P(\omega)\Psi(\omega)\, e^{j\omega t}\, d\omega = \int_X P(\omega)\, e^{j\omega t}\, d\omega \qquad (3.2.21c)$$

$$\int_\Lambda p(\lambda)\, d\lambda \int_\Lambda p(\xi)\psi(t-\lambda+\xi)\, d\xi = \int_{-\infty}^{\infty} P(\omega)\, P^*(\omega)\Psi(\omega)\, e^{j\omega t}\, d\omega = \int_X P(\omega)P^*(\omega)e^{j\omega t}\, d\omega \qquad (3.2.21d)$$

(the asterisk means complex conjugate here). Substituting these relations in (3.2.20), evaluating them at $t = 0$, and collecting terms, we find that (3.2.20) equals

$$\int_X |1 - P(\omega)|^2\, d\omega \qquad (3.2.22)$$

This is the left-hand side of (3.2.18). To determine the right side, we need only evaluate $\| f_o \|$. But, by definition,

$$\| f_o \|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} f_o^2(t)\, dt$$

so that, using Parseval's theorem (3.2.14),

$$\| f_o \|^2 = \int_{-\infty}^{\infty} |F_o(\omega)|^2\, d\omega$$

where $F_o(\omega)$ is the Fourier transform of $f_o(t)$ which, from (3.2.19) and (3.2.21b), is given by

$$F_o(\omega) = \Psi(\omega) - P^*(\omega)\Psi(\omega)$$

Therefore,

$$\| f_o \|^2 = \int_{-\infty}^{\infty} \Psi(\omega)\, |1 - P^*(\omega)|^2\, d\omega = \int_X |1 - P(\omega)|^2\, d\omega \qquad (3.2.23)$$

Substituting $\| f_o \|$ into the right-hand side of (3.2.18) we obtain again (3.2.23), so that (3.2.18) is an equality. This proves our assertion.

It is worth noting that there is nothing even remotely artificial about the circumstances described in this last proof, in which a pulse shape $p(\lambda)$ is specified a priori, and the bound is shown to be attained for $\Lambda = E^Y$ and $f(t) = f_o(t)$. This is, for example, precisely the situation corresponding to the physically interesting question: Under what conditions is a signal indistinguishable from its short-time average ? That is, under what conditions will there be no loss of information in a signal if the instrument with which it is observed yields (because of its limited bandwidth, resolution, and so forth) the short-time average of the signal rather than the signal itself ?

This question will be studied in detail in the next section, principally for the opportunity it gives us to insert some numbers in our results and to compare them with intuitive thinking. The remainder of this work will also be concerned with examining some of the implications of the bounds (3.1.12) and (3.2.18), with special regard to the fact that they are least upper bounds. In section 3.4 we consider some examples, couched in practical language, of what the bounds tell us about the relation between spectrum parameters, amplitude resolution, and time resolution. In section 3.5 we discuss what to the author seems the most thought-provoking consequence of these bounds, the density of simple systems in the space of all (linear, time-invariant) systems. Section 3.6 presents two simple but interesting uses of the density idea; the second application, that concerned with linear least-peak-error interpolation and extrapolation, is rather interesting and suggestive in itself.

3.3 COMPARISON OF PURE TRANSMISSION AND SHORT-TIME INTEGRATION

We wish to determine conditions under which a signal is indistinguishable from its short-time average. In terms of the $\underline{S}_{\psi}$ class pertinent to the signal, the signal will be indistinguishable from its short-time average if this is true for all probe functions belonging to $\underline{S}_{\psi}$. To reduce this problem to our standard form, we think of a probe $f \in \underline{S}_{\psi}$ being applied to two networks, of which the first has an impulse response consisting of just one unit impulse (pure transmission), and the second has a rectangular pulse of unit area and base $2\lambda_o$ centered

at the impulse position. We compare the outputs of these two networks on an oscilloscope with finite trace-length and -width, as in Fig. 3.3.1.

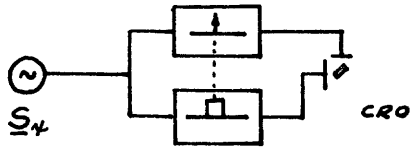

Fig. 3.3.1

Comparison of exact viewing and short-time averaging.

We can immediately apply the results of (3.2.18) et seq., with $\Omega = E^o = 1$ and the prescribed pulse shape $p(\lambda)$ given by

$$p(\lambda) = \begin{cases} \frac{1}{2\lambda_o} & \text{for} \quad -\lambda_o < \lambda < \lambda_o \\ 0 & \text{outside} \end{cases} \tag{3.3.1}$$

From (3.2.18), the difference in the outputs of the two networks is bounded by

$$\left| \Omega f(t) - \Omega \int_{\Lambda} p(\lambda)\, f(t-\lambda)\, d\lambda \right| = \left| f(t) - \frac{1}{2\lambda_o} \int_{t-\lambda_o}^{t+\lambda_o} f(\lambda)\, d\lambda \right|$$

$$\leq \| f \| \cdot \sqrt{\int_X |1 - P(\omega)|^2 \, d\omega} \tag{3.3.2}$$

Since $P(\omega)$ is 2π times the Fourier transform of $p(\lambda)$, it is given by $P(\omega) = \frac{\sin \omega\lambda_o}{\omega\lambda_o}$. To obtain a numerical evaluation of the integral in (3.3.2) we must first settle on some definite set X, i.e., choose some specific class S_ψ. In other words, we must choose a function $\psi(t)$. We shall use

$$\phi(t) = 2W \frac{\sin Wt}{Wt} \tag{3.3.3}$$

a function whose spectrum is unity in $-W \leqslant \omega \leqslant W$ and zero outside, so that X is the interval $-W \leqslant \omega \leqslant W$. The rest of the problem is the evaluation of the integral in (3.3.2), that is, just arithmetic.

NUMERICAL INTERLUDE. Consider the integral

$$I = \frac{1}{2W}\int_{-W}^{W}\left(1 - \frac{\sin \omega\lambda_o}{\omega\lambda_o}\right)^2 d\omega = \frac{1}{\lambda_o W}\int_0^{W\lambda_o}\left(1 - \frac{\sin x}{x}\right)^2 dx$$

$$= \frac{1}{\lambda_o W}\left[\int_0^{W\lambda_o} dx - 2\int_0^{W\lambda_o}\frac{\sin x}{x}\,dx + \int_0^{W\lambda_o}\left(\frac{\sin x}{x}\right)^2 dx\right] \qquad (3.3.4)$$

where we introduced the change of variable $\omega\lambda_o = x$. The last two integrals cannot be determined in closed form; however, the second one is extensively tabulated (see, e.g., (24)), and the third one can be reduced to the second by integration by parts:

$$\int_0^x (\sin^2 t)\left(\frac{dt}{t^2}\right) = -\frac{\sin^2 x}{x} + \int_0^{2x}\frac{\sin t}{t}\,dt$$

Using the notation $\underline{\mathrm{Si}}(x)$ for the sine-integral, $\underline{\mathrm{Si}}(x) = \int_0^x \frac{\sin t}{t}\,dt$, (3.3.4) becomes

$$I = 1 - \frac{2}{W\lambda_o}\underline{\mathrm{Si}}(W\lambda_o) + \frac{1}{W\lambda_o}\underline{\mathrm{Si}}(2W\lambda_o) - \left(\frac{\sin W\lambda_o}{W\lambda_o}\right)^2 \qquad (3.3.5)$$

Formula (3.3.5) is deceptively simple; actually its use is more unpleasant than seems at first sight, because the answer comes out as the difference of two numbers both of which are a factor of 10^3 larger than their difference, so that many digits have to be carried along to make the answer meaningful. Furthermore, an expression such as (3.3.5) is difficult to interpret and understand. Therefore, a very accurate approximate evaluation of (3.3.5) was obtained, using a series

expansion of the integrand. It is shown in Appendix A.5 that, for $W\lambda_o \leqslant 1$ (the region of most interest),

$$I \leqslant \frac{(W\lambda_o)^4}{5(3!)^2} = \frac{(W\lambda_o)^4}{5(36)} \qquad (W\lambda_o \leqslant 1) \tag{3.3.6}$$

This bound on I is only imperceptibly different from I itself (for example, see the numerical comparisons following (3.3.10)).

We can now use these numbers to answer our original question. Substituting back in (3.3.2) we obtain (remembering that $\sqrt{2W} = \sqrt{\phi(0)} = \|\phi\|$)

$$\left| f(t) - \frac{1}{2\lambda_o} \int_{t-\lambda_o}^{t+\lambda_o} f(\lambda)\, d\lambda \right| \leqslant \|f\| \sqrt{2WI} = \|f\| \cdot \|\phi\| \sqrt{I} \leqslant \|f\| \cdot \|\phi\| \frac{(W\lambda_o)^2}{6\sqrt{5}} \tag{3.3.7}$$

the last step being valid only for $W\lambda_o \leqslant 1$. Eq. (3.3.7) gives the absolute error, and we know from (3.2.19) that there exists a probe function $f_o(t) \in \underline{S}_\phi$ that actually reaches this maximum absolute error (except for the negligible difference between $\sqrt{I}$ and its approximate value), this function being given by

$$f_o(t) = \phi(t) - \frac{1}{2\lambda_o} \int_{-\lambda_o}^{\lambda_o} \phi(t+\lambda)\, d\lambda \tag{3.3.8}$$

According to our discussion, the relative error is just $\sqrt{I}$ $\left[\approx \frac{(W\lambda_o)^2}{6\sqrt{5}} \text{ for } W\lambda_o \leqslant 1\right]$. Suppose that our tolerable relative error is ϵ_o (that is, on our CRO with finite trace thickness, we cannot distinguish network outputs whose relative absolute difference is less than ϵ_o). Then, from our Initial Assumptions (sect. 2.1), a signal and its short-time average are indistinguishable if the relative error $\sqrt{I} \leqslant \epsilon_o$ for all $f \in \underline{S}_\phi$. However, since one probe function $f_o(t)$ actually

achieves the bound, the signal and its short-time average will be indistinguishable if and only if

$$\sqrt{I} < \epsilon_o \tag{3.3.9}$$

For our approximate evaluation, valid for $W\lambda_o \leq 1$, the corresponding condition is

$$\frac{(W\lambda_o)^2}{6\sqrt{5}} < \epsilon_o \tag{3.3.10}$$

For example, using (3.3.10)

if $\lambda_o = \frac{1}{2W}$, the tolerable relative error must satisfy $\epsilon_o \geq 1.86\ \%$

if $\lambda_o = \frac{1}{W}$, the tolerable relative error must satisfy $\epsilon_o \geq 7.45\ \%$

in order for the signal and its short-time average over an interval of length $2\lambda_o$ to be indistinguishable. By the way, the corresponding numbers obtained directly from (3.3.5) using 9-place tables are 1.85 % and 7.1 %, which shows how good the approximation is.

An interesting special case is the one in which the integration time is chosen to be the Nyquist spacing appropriate to functions bandlimited to the radian band $(-W, W)$. This spacing is $(2\lambda_o) = \frac{\pi}{W}$, so that here $\lambda_o = \frac{\pi}{2} \cdot \frac{1}{W}$. A rough plausibility argument, based on the relative constancy or smoothness of bandlimited functions between Shannon sample points, might have suggested the Nyquist spacing as an integration interval for which the signal and its short-time average would not differ appreciably. Substituting $\lambda_o W = \frac{\pi}{2}$ in (3.3.5) (we cannot use the approximate determination since now $\lambda_o W > 1$), we find that the relative error under these conditions is 16.8 %, which shows that the plausibility reasoning is rougher than might have been anticipated.

Of course, it is possible that the requirements that we have derived from analysis may be somewhat more stringent than those we know we can get away with in practice. This is necessarily a vague statement, but if its import is correct, it probably means that our thought experiment is not as close a model of a real situation as it might be. For example, it is possible that in the laboratory two networks are accepted as equivalent even if every now and then, but infrequently, the tolerable error level is exceeded, or if it is exceeded only for certain inputs that are considered unimportant or unrepresentative. The key words are "infrequent," "unimportant," "unrepresentative," and thinking along these lines, at Prof. P. Elias's suggestion, brings up the question of fidelity criteria [see Shannon (15), Kolmogorov (25)]. Our requirement that an absolute difference be uniformly small is a fidelity criterion, possibly an excessively stringent or unrealistic one. For example, $\underline{S}_{\psi}$ contains a probe f_o that turns our inequalities into equalities. Is this function f_o "unusual" or "exceptional" in some sense ? Are there many functions for which equality is almost reached ? Is f_o unusual enough so that for a large proportion of functions in $\underline{S}_{\psi}$ ours is not the best bound, and a better one could be found ? These questions would be meaningful only if some value or density function were defined on $\underline{S}_{\psi}$, something like a probability or content measure. It is clear that if such a measure were available, its use could only cause our bounds to decrease or remain constant – they certainly could not increase – and our results would perhaps model physical experience more closely.

However, there is no general criterion for choosing a criterion,

fidelity or other. It would have to be suggested by a specific problem.

3.4 RESOLUTION OF EVENTS IN NUCLEAR COUNTERS AND RADAR

We shall consider here two quasi-practical situations in which the only important information carried by a signal concerns the occurrence or nonoccurrence of events, and the number of events taking place. The real object of the discussion is to show how our results can be used, under certain conditions, to relate time-resolution, frequency parameters, and amplitude resolution or signal/noise ratio.

In the case of the counter, we suppose that the events being measured give rise to impulses which are applied to our amplifier (the counter). The problem is to determine under what conditions two events appearing close together will be resolved as two individual events and not be counted as one. We assume that the linear counter-amplifier has some specified, frequency-limited transfer function; and that the amplifier output is observed on a thick-traced oscilloscope, or, equivalently, that the amplifier generates noise, so that its noise/signal ratio at the output is greater than zero. Fig. 3.4.1 summarizes the situation in this problem.

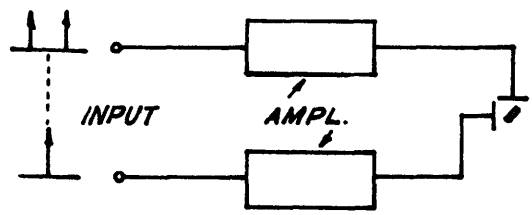

Fig. 3.4.1

The problem of distinguishing two closely-spaced events from one event.

In the case of radar, we might be interested in knowing whether a target under observation consists of just one object, or of several separate objects. The impulse response of the system "transmitter-to-target distribution-to-receiver" is of course just that of a singular network (in one dimension). When the radar transceiver can be considered a linear, frequency-limited system with a finite/signal noise ratio, the problem of deciding when two targets will be seen as one is again summarized by Fig. 3.4.1.

Our analysis in preceding sections was done on the basis of observation performed with a CRO whose trace has finite thickness. It is intuitively clear, however, that in their final effect on measurement accuracy, a thick trace and a small amount of additive noise (independent of signal) are similar. To the extent that they are, we can consider our "finite-resolution CRO"-situation to be like the "ideal CRO with some noise present"-situation. In the second case, the relative error ϵ_o = trace width/trace length would be obtained as a noise/signal ratio. Just which noise statistics and which signal statistics are pertinent to forming this ratio is somewhat arbitrary. They cannot be defined once and for all, because they depend on the detailed physical properties of a specific system, details such as persistence and luminescence qualities of CRO phosphors, integration properties of the human eye or other ultimate observer, and so forth. The pertinent signal parameters might be, for example, peak signal, r.m.s. signal, etc.; and the appropriate noise statistic might be the noise r.m.s. value, or the amplitude interval in which it spends, say, 90% of the time, or, in fact, any statistic that conveys the idea of a strip of uncertainty. To make our results applicable, the guide in defining the noise/signal ratio appropriate

to a specific problem is to make the definition lead to the same model as the oscilloscope with finite trace thickness.

To proceed with the problem, we notice first that the diagram of Fig. 3.4.1 can immediately be converted to our standard form just by interchanging the roles of source and excitation, as in Fig. 3.4.2.

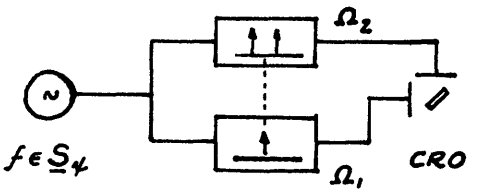

Fig. 3.4.2

Standard form of the resolution problem.

The probe function $f \in \underline{S}_\psi$ in Fig. 3.4.2 is just the impulse response of the amplifiers of Fig. 3.4.1. Our problem consists simply in determining a good bound on $\left| (\Omega_1 - \Omega_2) f(t) \right|$ for the specific given f(t). However, since our machinery is capable of working simultaneously with the whole class $\underline{S}_\psi$, we may as well follow the more general procedure, since then we shall be solving the problem for all possible frequency-limited amplifiers simultaneously. In particular, we recall that there exists a specific probe $f_o \in \underline{S}_\psi$ that actually achieves the least upper bound on $\left|(\Omega_1 - \Omega_2) f(t)\right|$. In our past formulation, this probe represented the least favorable case since it produces the maximum possible difference in the outputs from Ω_1 and Ω_2. For the same reason, in the present problem this probe represents the best possible case, since it emphasizes as much as possible the difference between two events and one. Therefore, the probe that actually attains the least upper bound

defines the optimum amplifier response. We may as well determine resolution for this optimum case, since that will yield us a number meaningful for the whole class of amplifiers rather than for just one specific one.

We shall examine one special case, the one for which $\mathcal{A}_1 = E^0$ and $\mathcal{A}_2 = \frac{1}{2}(E^\lambda + E^{-\lambda})$, and merely point out how to proceed in an arbitrary situation. This is the symmetric case shown in Fig. 3.4.3. The difference, $\mathcal{A}_1 - \mathcal{A}_2$, is given by

$$\mathcal{A}_1 - \mathcal{A}_2 = \frac{1}{2}(E^\lambda - 2 + E^{-\lambda}) = \frac{E^\lambda}{2}(1 - 2E^{-\lambda} + E^{-2\lambda}) = \frac{E^\lambda}{2}\nabla_\lambda^2 \tag{3.4.1}$$

whence

$$\left|(\mathcal{A}_1 - \mathcal{A}_2) f(t)\right| = \left|\left(\tfrac{1}{2} E^\lambda \nabla_\lambda^2\right) f(t)\right| \leq \|f\| \cdot \|\psi\| \cdot \tfrac{1}{2}\|\nabla_\lambda^2\| \tag{3.4.2}$$

As usual, (3.4.2) becomes an equality (at t = 0) for

$$f_o(t) = \frac{1}{\psi(0)}\left(\tfrac{1}{2} E^\lambda \nabla_\lambda^2\right)^* \psi(t) \tag{3.4.3}$$

and, by definition of the operator norm, the value reached is

$$\|\tfrac{1}{2} E^\lambda \nabla_\lambda^2\|^2 = \tfrac{1}{4}\|\nabla_\lambda^2\|^2 \tag{3.4.4}$$

This is the maximum difference-output attainable. It is obtained using the (unrealistic) optimum amplifier (which, incidentally, is nothing more

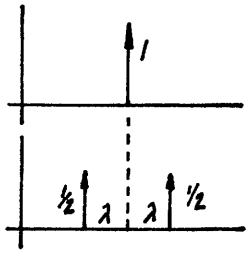

Fig. 3.4.3

Discrimination of two events from one.

than the filter matched to the difference-input).

Precisely the same result is obtained if the more general (3.2.18) is employed, using $p(\lambda) = \frac{1}{2}\left[\delta(\lambda - \lambda_o) + \delta(\lambda + \lambda_o)\right]$, where $\delta(\lambda)$ is the unit impulse at $\lambda = 0$. In this second way we can also treat the general asymmetric case or, in fact, any desired distribution of input impulses.

In (3.4.4) we have the maximum difference-output. At the same time, the actual output from, say, the first of the systems in Fig. 3.4.3 is just $f_o(t)$. It can be shown (see Appendix A.6) that the maximum value of the output occurs at $t = 0$, and is given by

$$|f_o(0)| = 1 - \frac{\psi(\lambda)}{\psi(0)} \tag{3.4.5}$$

At that point, therefore, the relative error is given by

$$\frac{\frac{1}{4}\|\nabla_\lambda^2\|_\psi^2}{1 - \frac{\psi(\lambda)}{\psi(0)}} = \frac{\psi(0)\,\|\nabla_\lambda^2\|_\psi^2}{4\left[\psi(0) - \psi(\lambda)\right]} \tag{3.4.6}$$

In order for two events to be distinguishable from one, this relative error must exceed the uncertainty introduced by finite CRO resolution or noise. If we let ρ_o stand for either the CRO resolution parameter ϵ_o or for the noise/signal ratio, and if we recall that this whole problem was studied on an optimum basis, we find that the <u>minimal</u> condition for discrimination is that

$$\frac{\psi(0)\,\|\nabla_\lambda^2\|_\psi^2}{4\left[\psi(0) - \psi(\lambda)\right]} > \rho_o \tag{3.4.7}$$

Using the definitions of the norm and the difference operator, it can be shown, with some algebraic manipulation, that

$$\|\nabla_\lambda^2\|^2 = \frac{2}{\psi(0)}\left[3\psi(0) - 4\psi(\lambda) + \psi(2\lambda)\right]$$

so that (3.4.7) can be written more explicitly,

$$\frac{3\psi(0) - 4\psi(\lambda) + \psi(2\lambda)}{2(\psi(0) - \psi(\lambda))} > \rho_o \qquad (3.4.8)$$

To make relation (3.4.8) more easily visualizable, it is plotted in Fig. 3.4.4 for the special case $\psi(t) = \phi_W(t) = 2W \frac{\sin Wt}{Wt}$, with $x = \lambda W$ used as abscissa. The minimal requirement for discrimination, for a given displacent-bandwidth product, is then that ρ_o be smaller than the value of the curve at that point. An interesting feature of the curve is that it is not monotonic, as might perhaps have been expected. The same behavior is exhibited in Fig. 3.4.5, which shows the corresponding result obtained in studying the discrimination problem when, instead of an optimum amplifier, a flat band-limited amplifier is used. The analysis for this second case, which can also be done without using our methods and therefore serves as a check, is as follows:

From (3.4.1), the difference-output is given by $\frac{1}{2} E^{\lambda} \nabla_{\lambda}^2 \phi_W(t)$. Because of the symmetry of the operator, the proof given before (Appendix A.6) shows that the peak difference-output occurs at $t = 0$ and is given by $\phi_W(0) - \phi_W(\lambda)$. The output from the amplifier is $\phi_W(t)$, whose peak value, $\phi_W(0)$, is also reached at $t = 0$. The relative error (the number that must be greater than the uncertainty for successful discrimination) is then

$$1 - \frac{\phi_W(\lambda)}{\phi_W(0)} = 1 - \frac{\sin x}{x}$$

where $x = W\lambda$. This is the curve shown in Fig. 3.4.5.

The most significant difference between the optimum case and the

flat case is in the asymptotic value of the attainable relative error, which is 3/2 in the former and 1 in the latter. This is an important difference, but it is not as immediately useful as it might seem because the optimum situation requires an amplifier matched to the displacement to be detected, and in many practical cases this is unknown a priori.

Actually, real amplifiers are not frequency-limited and for them the difference-output is greater than that predicted by (3.4.2). As a result, (3.4.7), instead of being a minimal requirement, is a conservative one. Just how conservative, our present analysis does not allow us to estimate exactly.

Our problem (the resolution problem) has an important analogue in optics,* with a rather extensive history both in pre-World War II classical optics, and in the more recent statistical studies in which noise is taken into account. In this context, the non-monotonicity of our discrimination curves (Figs. 3.4.4 and 3.4.5) corresponds to the phenomenon, well-known in optical circles, of "spurious resolution." References to the optical work on the resolution problem are grouped in the Bibliography under Number 30.

3.5 DENSITY THEOREMS

In this section we shall examine the relation that singular networks, considered as a class, bear to the class of all networks. [When we say all networks, we shall always mean, in this section, linear, time-invariant networks whose impulse responses are L_2.] Specifically, we wish to show that the study of singular networks is not the study of a very special case, but rather that it is, in a sense, just as general as

*I am indebted to Prof. Peter Elias for pointing this out.

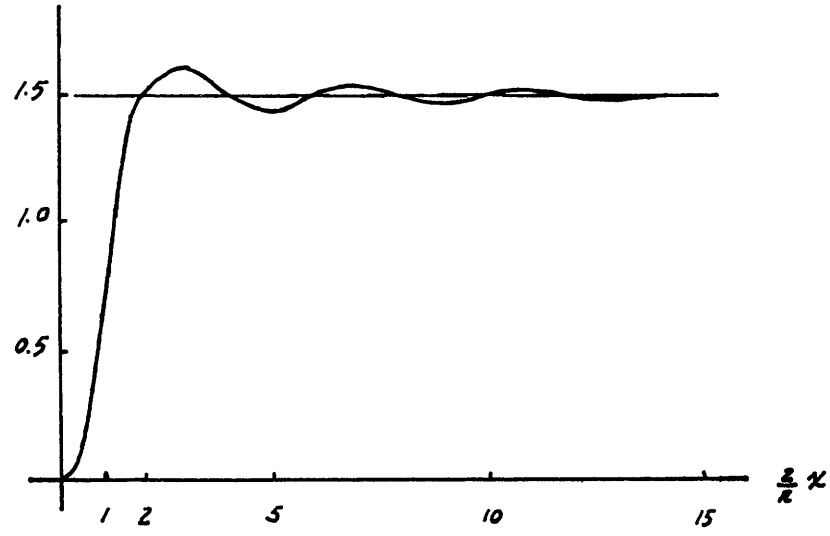

Fig. 3.4.4

Plot of the ratio (maximum difference-output)/(maximum output) versus the displacement-bandwidth product $x = \lambda W$, when the optimum amplifier is used.

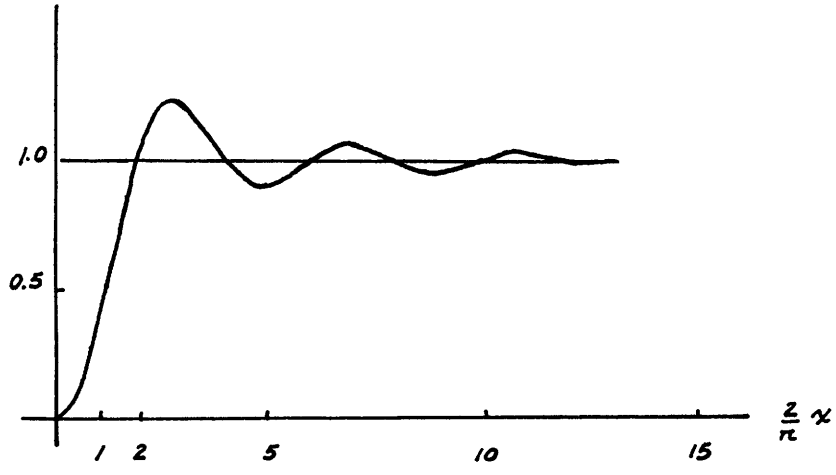

Fig. 3.4.5

The same as Fig. 3.4.4, for a flat amplifier.

studying the set of all possible impulse responses. The reason is that the set of singular networks is, in a sense, dense in the space of all possible networks. The special sense in which this is true involves the nature of our measurement process – in particular, the finiteness of its resolution – and the fact that our probes belong to restricted classes of functions.

The statement that we shall establish is the following:

> Given any arbitrary (linear, time-invariant, L_2) network N, it is always possible to find a singular network $\mathcal{Q}$ with the property that, for all probes of any arbitrary (but fixed) class $\underline{S}_{\psi}$, and for any preassigned (finite-resolution) oscilloscope, the outputs from N and $\mathcal{Q}$ will be indistinguishable. (3.5.1)

Before proceeding to the proof of (3.5.1), two remarks:

1. It will be recalled from the discussion following Eq. (3.2.18) that it is also true that with any singular network can be associated a smooth network; and that, with respect to the probes of any arbitrary but fixed $\underline{S}_{\psi}$, the error incurred by their interchange can be made zero. In fact, as was shown rigorously by (3.2.18), this occurs when the impulses in the singular response are replaced with $\psi(t)$-functions because, to probes in $\underline{S}_{\psi}$, the $\psi(t)$-functions <u>are</u>, so to speak, impulses.

2. The statement (3.5.1) promises slightly more than our proof will justify. The proof holds for classes $\underline{S}_{\psi}$ that are completely arbitrary <u>except</u> that there must exist a finite number W beyond which $\Psi(\omega)$ is definitely zero. To see that there are idempotent, L_2 spectra for which the required W does not exist, consider the spectrum $\Psi_0(\omega)$ defined by:

$\Psi_o(\omega) = 1$ for $\omega \in \bigcup_{n=0}^{\infty} [n, n + \frac{1}{2^n}]$ (and, of course, symmetrically for negative frequencies)

$\Psi_o(\omega) = 0$ elsewhere

$\Psi_o(\omega)$ is idempotent. To see that it is L_2, note that

$$\int_{-\infty}^{\infty} |\Psi(\omega)|^2 \, d\omega = 2 \int_0^{\infty} \Psi(\omega) \, d\omega = 2\left[\sum_{n=0}^{\infty} \frac{1}{2^n}\right] = 2(2) = 4 < \infty$$

Obviously, there exists no finite W beyond which $\Psi(\omega)$ is zero. Spectra of this type are not negligible or especially unimportant, but we shall ignore them in proving (3.5.1). [Actually, to include them it is only necessary to carry along in the proof an extra contribution to the error, due to the arbitrarily small energy which is neglected if a suitably large, finite number W is associated with such a spectrum.]

Proof of (3.5.1):

Given an arbitrary probe class $\underline{S}_{\psi}$, let W be a positive number with the property that $\Psi(\omega) = 0$ for $|\omega| \geq W$. Define the symbol $\phi_W(t)$ to mean

$$\phi_W(t) = 2W \frac{\sin Wt}{Wt} \tag{3.5.2}$$

Then $\phi_W(t)$ is the time-function corresponding to the spectrum $\Phi_W(\omega) = 1$ for $|\omega| < W$, $\Phi_W(\omega) = 0$ for $|\omega| \geq W$. Clearly, for any ψ-function that meets the stated requirements on W, $\underline{S}_{\psi} \subset \underline{S}_{\phi_W}$, so that proving the theorem for the probe class $\underline{S}_{\phi_W}$ also proves it for any other $\underline{S}_{\psi}$ for which W has the required property.

Given an arbitrary L_2 impulse response h(t), consider instead the bandlimited response $\tilde{h}(t)$ given by

$$\tilde{h}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(\tau)\, \phi_W(t-\tau)\, d\tau \tag{3.5.3}$$

It is easily verified (consider the Fourier transform of (3.5.3)) that for probes $f \in \underline{S}_{\phi_W}$ (and, a fortiori, for probes $f \in \underline{S}_{\psi} \subset \underline{S}_{\phi_W}$) the difference between the outputs from h(t) and $\tilde{h}(t)$ is zero. That is, for the specified probes, nothing is changed if h is replaced with $\tilde{h}$, and vice versa. Our problem is therefore reduced to associating a singular network with $\tilde{h}(t)$.

But this is easy. By the Shannon sampling theorem, $\tilde{h}(t)$, being bandlimited, can be expressed in the form

$$\tilde{h}(t) \sim \sum_{n=-\infty}^{\infty} \tilde{h}\left(\frac{n\pi}{W}\right) \frac{\phi_W\left(\frac{n\pi}{W} - t\right)}{\phi_W(0)} \tag{3.5.4}$$

Consider the N^{th} partial sum, $\sigma_N(t)$, of the series in (3.5.4):

$$\sigma_N(t) = \sum_{n=-N}^{N} \tilde{h}\left(\frac{n\pi}{W}\right) \frac{\phi_W\left(\frac{n\pi}{W} - t\right)}{\phi_W(0)} \tag{3.5.5}$$

From the discussion following (3.2.18) (and from the first remark preceding this proof) the singular network

$$\Omega = \sum_{n=-N}^{N} \tilde{h}\left(\frac{n\pi}{W}\right) E^{\frac{n\pi}{W}} \tag{3.5.6}$$

can be substituted for the network with response $\sigma_N(t)$ <u>with zero error</u> at the output, for any input $f(t) \in \underline{S}_{\phi_W}$. Since there is zero error in going from the smooth network with response $\sigma_N(t)$ to the singular network Ω, all that we have to show now is that the difference between the outputs from $\tilde{h}(t)$ and $\sigma_N(t)$, for inputs $f(t) \in \underline{S}_{\phi_W}$, can be made uniformly arbitrarily small. This, of course, can be proved immediately.

Let

$$g_1(t) = \int_{-\infty}^{\infty} \tilde{h}(\lambda)\, f(t-\lambda)\, d\lambda \tag{3.5.7a}$$

$$g_2(t) = \int_{-\infty}^{\infty} \sigma_N(t)\, f(t-\lambda)\, d\lambda \tag{3.5.7b}$$

be the outputs from $\tilde{h}$ and σ_N, when excited by $f \in \underline{S}_{\phi_W}$. Then

$$\left| g_1(t) - g_2(t) \right| = \left| \int_{-\infty}^{\infty} \left[\tilde{h}(\lambda) - \sigma_N(\lambda) \right] f(t-\lambda)\, d\lambda \right|$$

$$\leq \left(\int_{-\infty}^{\infty} \left[\tilde{h}(\lambda) - \sigma_N(\lambda) \right]^2 d\lambda \right)^{1/2} \left(\int_{-\infty}^{\infty} f^2\, d\lambda \right)^{1/2} \tag{3.5.8}$$

Our assumptions guarantee the existence and finiteness of both integrals. Furthermore, since the orthonormal set in the Shannon sampling theorem is complete, Parseval's theorem holds, which implies that

$$\lim_{N \to \infty} \int_{-\infty}^{\infty} \left[\tilde{h}(\lambda) - \sigma_N(\lambda) \right]^2 d\lambda = 0$$

This, of course, implies that given an arbitrary ϵ^2, there exists a sufficiently large (but finite) N so that

$$\int_{-\infty}^{\infty} \left[\tilde{h}(\lambda) - \sigma_N(\lambda) \right]^2 d\lambda < \epsilon^2$$

Substituting in (3.5.8) we obtain the result:

$$\left| g_1(t) - g_2(t) \right| < \epsilon K_f \tag{3.5.9}$$

We have written K_f for the finite second factor in (3.5.8). Thus we can make the difference between the outputs from $\tilde{h}(t)$ and $\sigma_N(t)$ arbitrarily small – in particular, smaller than the resolution of any preassigned oscilloscope.

Recall now, 1) that there is no difference between the outputs

from σ_N and Ω, so that we could just as well have written, in (3.5.7b), $g_2(t) = \Omega f(t)$; and, 2) that there is no difference in the outputs from $\tilde{h}(t)$ and $h(t)$, so that we could have written, in (3.5.7a), $g_1(t) = \int_{-\infty}^{\infty} h(\lambda)\, f(t-\lambda)\, d\lambda$. Substituting these meanings for g_1 and g_2 in (3.5.9), we obtain, finally,

$$\left| \Omega f(t) - \int_{-\infty}^{\infty} h(\lambda)\, f(t-\lambda)\, d\lambda \right| < \epsilon K_f \qquad (f \in \underline{S}_{\phi_W}) \tag{3.5.10}$$

Here h(t) was the given, arbitrary impulse response. For sufficiently large (but finite) N, the singular network Ω given by (3.5.6) has the required property. This establishes our theorem.

(We remark, by the way, that (3.5.6) is, from a practical point of view, a grossly inefficient way of arriving at a suitable singular network. Our present purpose was to prove the existence of a suitable Ω, not to give a practical algorithm for its construction.)

Our proof depended heavily on the special properties of bandlimitation. It is therefore worthwhile to point out briefly that, as Dr. M. V. Cerrillo has remarked,* bandlimitation really has no fundamental bearing on the problem. Actually there are several different bases for arriving at conclusions of the type represented by our density theorem. To give one example, it is easy to show (see Wernikoff (23), p. 44, (26), p. 70) that:

> If A(t) is the step response of any physically realizable, finite-memory network (with memory-time T), and if A(t) is continuous; if $A_n(t)$ is a simple-function approximation to A(t) (and therefore the step-response of a singular network); if f(t) is an input function and g(t) and $g_n(t)$ the corresponding outputs from A and A_n respectively; then

* Personal communication.

$$\left| g(t) - g_n(t) \right| \leq \epsilon_n \int_{t-T}^{t} \left| \frac{df(\xi)}{d\xi} \right| d\xi \qquad (3.5.11)$$

In (3.5.11), ϵ_n is the uniform bound on the approximation of A by A_n: $|A(t) - A_n(t)| < \epsilon_n$. Several different sets of assumptions are possible, some of which are mentioned in the reference. For example, (3.5.11) provides a uniform bound if f(t) has a uniformly bounded first derivative, or, what is less restrictive, if f(t) is of uniformly bounded variation over any interval of length T. The assumptions in these theorems are sufficient, but they are not necessary, as is easily shown by examples. For the statement (3.5.1), just consider an h(t) that is the sum of an impulse and an L_2 response $h_1(t)$. Clearly the sum h(t) is not L_2. On the other hand, the conclusion of the theorem still holds, because if we decompose h(t) into its two components, the impulse represents a singular network to start with, so that it contributes no error; and the L_2 component, $h_1(t)$, can be treated as before.

There is nothing surprising about theorems like (3.5.1). They are not widely known or used, but this seems to be true mainly because the usual approach to approximating network operation is very ambitious Besides not recognizing explicitly the acceptability of a finite error, it tries to perform the approximation for all possible inputs (including impulses) simultaneously, so that the problem becomes the very difficult one of point-by-point approximation of a function (the impulse response) rather than just the imitation of the mode of operation of the network. It is only when the process is required to work for all inputs and zero error that imitating the operation of a network and approximating its impulse response are the same thing. If finite error is admitted, and if it is recognized that in many important situations it

is only required to have the approximation hold for a restricted class of inputs (not including the impulse), then there is no longer a unique impulse response to characterize the network. In fact, there is now an infinite family of impulse responses associated with each network, most of them wildly different in appearance. (See section 2.5, especially Fig. 2.5.1) The only criterion of approximation available now must be in terms of the outputs of the networks resulting from all excitations belonging to the required class. In this case, an approximation is deemed successful if the outputs of the networks are "indistinguishable" in some sense. It is therefore possible, under these conditions, to imitate the network operation closely and, at the same time, have the impulse responses bear no visible relation to each other.

The real importance of (3.5.1) for our discussion is that it implies that, under the restrictions of our thought experiment, it is just as general to obtain results for singular networks as it is to obtain them for arbitrary networks. Therefore, if it should happen that certain types of problems are more easily studied with singular than with ordinary networks, this provides the justification for taking the easy way. This is really the point of the thesis. Our probe classes and singular networks provide a working model of signal/system interaction that is different from, but, in its special way, as complete as, the ordinary model. Since arbitrarily close to every network there is a singular network, and every singular network can be replaced with a smooth network with zero error, we see that, in effect, the set of singular networks is a subset of the set of all networks, a subset that

is dense in the set of all networks.* If a problem is simple in the space of singular networks, then we have a set of simple systems from which all other systems can be reached. This is especially useful for problems that can be solved in the subset but not in the set of all networks. An example of this type is given in the next section, where we determine the optimum least-peak-error predictor using singular networks and their algebra. This problem cannot be (or, at least, apparently never has been) solved using ordinary methods.

3.6 MATCHED FILTERS AND LINEAR LEAST-PEAK-ERROR INTERPOLATION AND EXTRAPOLATION

Some examples of the uses of the algebra of singular networks have already been given. In this section, we conclude this dissertation with two more examples. In the first one, on matched filters, a known result is re-derived, mainly to exhibit the simplicity of the algebra, but partly also to provide a partial check on the essential correctness of our work. In the second one, on linear least-peak-error interpolation and extrapolation, a new and rather interesting result is obtained. The second problem has the additional virtue that it does not seem to be amenable to solution by the usual methods, yet is quite easily solved here.

A. MATCHED FILTERS

We shall prove the well-known theorem** concerning the optimality of matched filters in detecting the presence of a signal when the

---

* A set A is dense in a set B if every point of B either belongs to A or is a limit point of A. That is, all points of B can be reached starting from points in A. For example, the set of rational numbers is dense in the real line.

** For references, see (27) and (28).

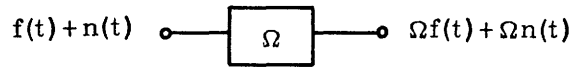

Fig. 3.6.1

A detection problem.

latter is submerged in noise.

With reference to Fig. 3.6.1, we assume the probe function $f(t) \in \underline{S}_{\psi}$ to be the "signal" whose presence we wish to detect, and $n(t)$ to be noise, which we assume to have a white spectrum in the bands where $\psi(t)$ has its energy, and no energy outside of those bands. (This last assumption seems natural, since it does not make sense, in a filtration problem, to allow the noise to have energy in bands where the signal does not have any.)

We are interested in maximizing the ratio

$$\rho = \frac{\text{peak signal}}{\text{r. m. s. noise}} = \frac{\text{peak}[\Omega f(t)]}{\text{r. m. s. noise}} \tag{3.6.1}$$

Let us first determine the root-mean-square output noise:

$$\text{mean square}[\Omega n(t)] = \overline{\Omega n(t) \Omega n(t)}^t = \sum_i \overline{a_i n(t-t_i) \sum_j a_j n(t-t_j)}^t$$

$$= \sum_{i,j} a_i a_j \overline{n(t) n(t+t_i-t_j)}^t = \sum_{i,j} a_i a_j R_n(t_i - t_j)$$

$$= \Omega \Omega^* R_n(0) \tag{3.6.2}$$

where $R_n(\lambda)$ is the autocorrelation function of the noise, $R_n(\lambda) = \overline{n(t)\, n(t+\lambda)}^t$. For noise that is white in the bands of $\psi(t)$, and with a $\frac{\text{noise-power}}{\text{radian/sec}} = N_o$,

$$R_n(\lambda) = N_o \psi(\lambda) \tag{3.6.3}$$

so that the mean square noise in (3.6.2) is given by

$$\Omega\Omega^* R_n(0) = N_o \Omega^*\Omega\psi(0) = N_o \psi(0) \|\Omega\|^2 \tag{3.6.4}$$

Therefore, the r.m.s. output noise is

$$\sqrt{N_o \psi(0)}\, \|\Omega\| \tag{3.6.5}$$

Our problem, therefore, is to maximize

$$\rho = \frac{\text{peak}\, \Omega f(t)}{\sqrt{N_o \psi(0)}\|\Omega\|} = \frac{1}{\sqrt{N_o \psi(0)}} \text{peak}\left[\frac{\Omega}{\|\Omega\|} f(t) \right] \tag{3.6.6}$$

Now, if Ω is given, then, from (2.2.21), the peak is reached when $f(t) = \Omega^*\psi(t)$. By the same proof (the Schwartz inequality) if f(t) is given, in the form* $\Omega_1\psi(t)$, then the peak is reached for $\Omega = \Omega_1$. In either case, the numerator in (3.6.6) becomes $\Omega^*\Omega\psi(0)$, so that

$$\rho = \frac{\psi(0)\|\Omega\|^2}{\sqrt{N_o \psi(0)}\|\Omega\|} = \frac{\sqrt{\psi(0)}\, \|\Omega\|}{\sqrt{N_o}} \tag{3.6.7}$$

To write this answer in a more familiar form, recall that

$$\sqrt{\psi(0)}\, \|\Omega\| = \sqrt{\psi(0)} \sqrt{\frac{1}{2\pi\psi(0)} \int_{-\infty}^{\infty} [\Omega\psi(t)]^2\, dt} = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} f^2(t)\, dt}$$

so that

$$\rho = \frac{\sqrt{\int_{-\infty}^{\infty} f^2(t)\, dt}}{\sqrt{2\pi N_o}}$$

which is the well-known result (the 2π in the denominator comes from

---

* There is a trifling bit of cheating here. Actually, Ω_1 could very well consist of an infinite series of terms, and if it did, it would no longer belong to $\underline{O}$ and our algebra would not apply to it. However, either by omitting the tail of the series or closing one eye, it is clearly possible to carry on.

the fact that our noise power density is defined per radian/sec rather than the usual cycles/sec).

Recalling that Ω^* is just Ω folded over in time, we see that our "signal" and network bear the expected relation to each other. Also, the final result depends only on signal energy and noise power density, as it should.

B. LEAST-PEAK-ERROR INTERPOLATION AND EXTRAPOLATION

Suppose that we are interested in finding a network Ω with the property that, for any probe $f(t) \in \underline{S}_\mathcal{E}$, the output will be the best estimate of $f(t \pm a)$, where by best we mean that the peak difference between the desired output, $f(t \pm a)$, and the actual output, $\Omega f(t)$, is a minimum. If we can obtain such a network Ω, and if its specification depends only on the class $\underline{S}_\mathcal{E}$ and not on the properties of a particular $f \in \underline{S}_\mathcal{E}$, then that network will also be the optimum operator for any signal corresponding to the class $\underline{S}_\mathcal{E}$.

It turns out that this problem can be solved very easily. With

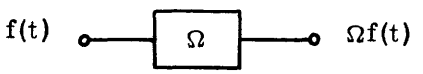

Fig. 3.6.2

The desired output is $f(t \pm a)$.

reference to Fig. 3.6.2, the instantaneous difference e(t) between actual and desired output is

$$e(t) = \Omega f(t) - f(t \pm a) = (\Omega - E^{\pm a}) f(t) \qquad (3.6.8)$$

From Chapter II we recall that

$$\left|e(t)\right|_{max} = \left| (\Omega - E^{\pm a})\, f(t)\right|_{max} \leq \|f\| \cdot \|\psi\| \cdot \|\Omega - E^{\pm a}\| \qquad (3.6.9)$$

It will be recalled that the bound in (3.6.9) is actually attained for some particular $f_o(t) \in \underline{S}_\psi$. Since equality <u>is</u> reached for some input, we shall solve our problem if we minimize the bound on the maximum error.

Suppose, for example, that we want a predictor E^a $(a > 0)$, and wish the optimum Ω to be physically realizable. This means that Ω must be of the form $\sum_n \beta_n E^{-t_n}$ $(t_n > 0)$. Clearly,* $\|\Omega - E^a\|$ is a minimum when Ω is the projection of E^a into the realizable subspace of $\underline{O}$; or, if (as is the case in some practical problems) a set $\{E^{-t_n}\}$ is prescribed, then $\|\ldots\|$ is a minimum when Ω is the projection of E^a into the manifold spanned by $\{E^{-t_n}\}$. Since we already know how to obtain the coefficients for these projections (see section 2.7), the problem is solved.

Before considering this problem a little further, it is interesting to notice that our simple discussion has actually yielded us a surprisingly general result:

> The linear extrapolation (interpolation) derived from projection is the best possible (in the sense of least peak error) for functions belonging to any ψ-class. For such functions, any other linear extrapolation (interpolation) technique known in numerical analysis can at best only be as good as projection. (3.6.10)

---

* For example, in 3-dimensional space, picture a vector v sticking out of a plane, but not perpendicular to it. Consider the projection v_p of v on the plane, and any other vector u in the plane. From the Pythagorean theorem, the distance from u to v, d(u, v), is given by $d^2(u, v) = d^2(u, v_p) + d^2(v_p, v)$. Since all terms are positive, and the last one is independent of u, the distance d(u, v) will be a minimum when $d(u, v_p) = 0$, which implies that $u = v_p$. This argument is the substance of the rigorous proof of our statement, and its validity is independent of the dimensionality involved.

Returning to the original problem, we recall that two situations were considered: one in which the only requirement was that the optimum network Ω be physically realizable, and the other in which a set of sample points $\{t_n\}$ were prescribed, so that Ω had to be a linear combination of the set $\{E^{-t_n}\}$.

Let us discuss the second possibility first. It might correspond to the physically important situation in which a scanning device yields arbitrarily spaced samples of a process. This might happen either because of randomly imperfect operation of the device, or because of its intrinsic nature (for example, a device that samples only when a certain threshold is exceeded). Suppose that the device has been operating for some time, so that a certain amount of data has been collected, the data consisting of sample values and corresponding sample times. The problem is; How should this information be used to estimate with least peak error the values of the process between sample times (interpolation) or the values of the process before or after the sampling started or stopped (extrapolation) ? For probes $f \in \underline{S}_\psi$, or for signals corresponding to some ψ-class, our result gives the best possible linear procedure.

Another situation in which a set $\{E^{-t_n}\}$ might be prescribed is if a tapped delay line is available, and it is desired to make a predictor (or increase the effective length of the delay line) using available equipment. In that case, the tap positions on the line determine the set $\{t_n\}$.

If there are no restrictions except realizability on Ω, the problem is no different from the previous one until the question of efficiency is

raised – and then it becomes prohibitively difficult. That is, if someone asks, for example, how should ten taps be placed on a line of fixed length so as to obtain the smallest possible minimum error, no answer is available. (There are, of course, a few common-sense rules: if the problem is to represent E^{α}, don't choose the $\{E^{-t_n}\}$ orthogonal to it; if $E^{-\beta}$ is contained within the available length of delay line, choose β as a tap point.) But the general question of the efficient representation of vectors by sets of other vectors is one that mathematicians have avoided most meticulously.

APPENDICES

A.1 LINEAR INDEPENDENCE OF $\{\psi(t - t_n)\}$.

We wish to show that, given an arbitrary finite set of numbers $\{t_n\}$, there exists no corresponding set of constants $\{a_n\}$ (not all of them zero) with the property that

$$f(t) = \sum_n a_n \psi(t - t_n) \equiv 0 \qquad (A.1.1)$$

for all values of t.

Consider the Fourier transform $F(\omega)$ of $f(t)$. Clearly, (A.1.1) requires that

$$F(\omega) = \Psi(\omega) \sum_n a_n e^{-j\omega t_n} = 0 \quad \text{almost everywhere} \qquad (A.1.2)$$

But then the known linear independence of any finite set of functions $\{e^{-j\omega t_n}\}$ over any interval (in particular, over the non-zero bands of Ψ) implies that (A.1.2) can be satisfied only if $a_n = 0$ for all n.

A.2. CERRILLO KERNELS.

We wish to show that the Cerrillo pure transmission kernels T_m, defined by

$$T_m = E^{-\lambda} \sum_{k=0}^{m} (-1)^k \binom{m+1}{k+1} E^{-(2\lambda)k} \qquad (A.2.1)$$

satisfy

$$T_m - T_{m-1} = E^{-\lambda} \nabla_{2\lambda}^m \qquad (A.2.2)$$

Recalling that, by definition, $\nabla_\mu = (1 - E^{-\mu})$; $\nabla_\mu^2 = \nabla_\mu(\nabla_\mu) = (1 - E^{-\mu})^2$; ... and so forth, we have

$$\nabla_\mu^m = (1 - E^{-\mu})^m = \sum_{k=0}^{m} \binom{m}{k} (-E^{-\mu})^k \qquad (A.2.3)$$

On the other hand,

$$E^\lambda(T_m - T_{m-1}) = \sum_{k=0}^{m} \binom{m+1}{k+1} (-E^{-\mu})^k - \sum_{k=0}^{m-1} \binom{m}{k+1} (-E^{-\mu})^k \qquad (A.2.4)$$

Equality of (A.2.3) and (A.2.4) requires that corresponding coefficients be equal:

$$\begin{aligned} &\binom{m+1}{k+1} - \binom{m}{k+1} = \binom{m}{k} \quad \text{for} \quad k = 0, \ldots, m-1 \\ &\binom{m+1}{k+1} - 0 = \binom{m}{m} \qquad \text{for} \quad k = m \end{aligned} \qquad (A.2.5)$$

The second relation in (A.2.5) is trivial. To see that the first is satisfied, it is only necessary to insert the meaning of the binomial coefficients in terms of factorials.

A.3 BOUNDS.

We desire a bound on

$$\|\nabla^n\| = \left[\frac{\nabla^{n^*}\,\nabla^n\,\phi(0)}{\phi(0)}\right]^{1/2} \qquad \text{(A.3.1)}$$

Since

$$\nabla^{n^*}\,\nabla^n = [(1-E^{-\mu})^*\,(1-E^{-\mu})]^n = [(1-E^{\mu})(1-E^{-\mu})]^n$$

$$= (-E^{\mu})^n\,(1-E^{-\mu})^{2n} = (-1)^n\,E^{n\mu}\,\nabla^{2n} \qquad \text{(A.3.2)}$$

our problem reduces to finding a bound on

$$\frac{1}{\phi(0)}\,(-1)^n\,\nabla^{2n}\,\phi(n\mu) \qquad \text{(A.3.3)}$$

We notice that it is easy to obtain bounds on the derivatives of $\frac{\sin x}{x}$. Therefore, if we can obtain a relation between ∇ and $\frac{d}{dx}$, we shall have our answer.

If we consider a continuously differentiable function f, the mean value theorem for derivatives asserts that

$$\nabla_\mu f(x) = f(x) - f(x-\mu) = \mu f'(\xi), \text{ where } \xi\epsilon[x, x-\mu]$$

(we denote differentiation by primes). Therefore,

$$|\nabla_\mu f| \leq \mu |f'|_{\max} = \mu[\text{uniform bound on } f'] \qquad \text{(A.3.4)}$$

Similarly, since

$$\nabla^2 f(x) = \nabla[\nabla f(x)],$$

$$|\nabla^2 f| = |\nabla(\nabla f)| \leq \mu |(\nabla f)'|_{\max} = \mu |\nabla f'|_{\max} \leq \mu^2 |f''|_{\max}$$

so that

$$|\nabla^2 f| \leq \mu^2 |f''|_{\max} = \mu^2[\text{uniform bound on } f']$$

Repeating this procedure n times we find that

$$|\nabla^n f| \leq \mu^n |f^{(n)}|_{max} \qquad (A.3.5)$$

With this result, we can immediately place a bound on (A.3.3). Recalling that the spectrum of

$$\phi_W(t) = 2W \frac{\sin Wt}{Wt}$$

is unity for $|\omega| \leq W$ and zero outside, we find that

$$\phi^{(k)}(t) = \int_{-W}^{W} (j\omega)^k e^{j\omega t} d\omega$$

so that

$$|\phi^{(k)}| \leq \int_{-W}^{W} |\omega|^k d\omega = 2 \int_0^W \omega^k d\omega = \frac{2W^{k+1}}{k+1} \qquad (A.3.7)$$

Therefore, from (A.3.5) and (A.3.7),

$$\left|\frac{\nabla^{2n} \phi(n\mu)}{\phi(0)}\right| \leq \frac{\mu^{2n}}{2W} \cdot \frac{2W^{2n+1}}{2n+1} = \frac{(\mu W)^{2n}}{2n+1} \qquad (A.3.8)$$

so that, finally,

$$\|\nabla_\mu^n\| \leq \frac{|\mu W|^n}{\sqrt{2n+1}} \qquad (A.3.9)$$

This bound becomes increasingly better as μW decreases to zero.

A.4 ANALYTICITY OF $f \in \underline{S}_{\psi}$.

We wish to prove that any probe belonging to any class $\underline{S}_{\psi}$ is, if time is considered a complex variable, an analytic entire function of time. We shall only show this for probe classes $\underline{S}_{\psi}$ that have a defining spectrum $\Psi(\omega)$ for which a finite number W exists beyond which $\Psi(\omega) \equiv 0$. (To see that this leaves out some possible cases, see Remark 2, Section 3.5). For a different proof of a similar statement, see Whittaker (17) who shows that his cardinal functions, which correspond to our probes, are analytic entire.

Let W be a finite number with the property that $\Psi(\omega) \equiv 0$ for $|\omega| > W$. Then, if $f \in \underline{S}_{\psi}$, it is clearly also contained in $\underline{S}_{\phi_W}$, where ϕ_W is defined in (A.3.6). Then f(t) can be represented by

$$f(t) = \int_{-W}^{W} F(\omega)\, e^{j\omega t}\, d\omega = \int_{-W}^{W} d\omega \sum_{k=0}^{\infty} \frac{(j\omega t)^k}{k!} F(\omega) \qquad (A.4.1)$$

The series for the exponential is uniformly convergent for all ω. If the spectrum $F(\omega)$ of $f(t)$ is reasonably well behaved (for example, if it is bounded), multiplying the terms of the series by $F(\omega)$ does not affect the uniformity of convergence. Therefore, summation and integration may be interchanged (see, e.g., Whittaker and Watson (29), p. 78) yielding for $f(t)$ the series representation

$$f(t) = \sum_{0}^{\infty} \frac{(jt)^k}{k!} \int_{-W}^{W} \omega^k F(\omega)\, d\omega \qquad (A.4.2)$$

To determine the radius of convergence of this series, we apply the root test:

$$|a_n|^{1/n} = \left[\frac{|jt|^n}{n!}\int_{-W}^{W} \omega^n F(\omega)\, d\omega\right]^{1/n} = \frac{|t|}{(n!)^{1/n}}\left[\int_{-W}^{W} \omega^n F(\omega)\, d\omega\right]^{1/n} \qquad (A.4.3)$$

Using the Schwartz inequality on the integral, and recalling that f and therefore F are integrable square,

$$|a_n|^{1/n} \leq \frac{|t|\ W^{\frac{2n+1}{2n}}}{(n!)^{1/n}\,(2n+1)^{1/n}}\left[\int_{-W}^{W} |F|^2\, d\omega\right]^{\frac{1}{2n}} \qquad (A.4.4)$$

As $n \to \infty$, the bracket approaches unity, and $W^{2n+1/2n}$ approaches W. The important term is the factorial, and using Stirling's approximation we see that $|a_n|^{1/n}$ approaches zero faster than $\frac{|t|}{n}$ as $n \to \infty$. Since the radius of convergence of the series (A.4.2) is given by

$$R = \lim \inf |a_n|^{-\frac{1}{n}},$$

we therefore see that the series converges for all finite $|t|$. Since (A.4.2) is a power series in t, f(t) must be an analytic entire function of t.

A.5 APPROXIMATE DETERMINATION OF AN INTEGRAL.

We desire an approximate determination of

$$I = \frac{1}{\lambda_o W} \int_0^{\lambda_o W} (1 - \frac{\sin x}{x})^2 \, dx \qquad (A.5.1)$$

From the analysis preceding (3.1.16) we have that

$$1 - \frac{\sin x}{x} = \frac{x^2}{3!} - \frac{x^4}{5!} + \ldots - \ldots$$

and also that, for $|x| \leq 1$,

$$\frac{x^2}{3!} - \frac{x^4}{5!} \leq 1 - \frac{\sin x}{x} \leq \frac{x^2}{3!}$$

so that

$$\int_0^{\lambda_o W} \left(\frac{x^2}{3!} - \frac{x^4}{5!}\right)^2 dx \leq \int_0^{\lambda_o W} (1 - \frac{\sin x}{x})^2 \, dx \leq \int_0^{\lambda_o W} \left(\frac{x^2}{3!}\right)^2 dx$$

$$(|\lambda_o W| \leq 1) \qquad (A.5.2)$$

Integrating the last member of (A.5.2), we obtain

$$I \leq \frac{1}{\lambda_o W} \frac{|\lambda_o W|^5}{5(3!)^2} = \frac{(\lambda_o W)^4}{5(36)} \quad (|\lambda_o W| \leq 1) \qquad (A.5.2)$$

It is clear from (A.5.2) that this bound does not differ appreciably from I itself. A numerical confirmation of this is given in the discussion following (3.3.10).

A.6 MAXIMUM VALUE OF $f_o(t)$.

We wish to determine where

$$f_o(t) = \frac{1}{2\psi(0)} (E^{\lambda} \nabla_{\lambda}^2)^* \psi(t) \qquad (A.6.1)$$

has its maximum value, and what this maximum value is.

From the definition of the operators

$$f_o(t) = \frac{1}{2\psi(0)} E^{-\lambda} (1 - 2E^{-\lambda} + E^{-2\lambda})^* \psi(t)$$

$$= \frac{1}{2\psi(0)} (E^{-\lambda} - 2E^0 + E^{\lambda}) \psi(t) \qquad (A.6.2)$$

Expressing (A.6.2) as an inverse Fourier transform,

$$f_o(t) = \frac{1}{2\psi(0)} \int_{-\infty}^{\infty} \Psi(\omega) (e^{-j\omega\lambda} - 2 + e^{j\omega\lambda}) e^{j\omega t} d\omega$$

$$= - \frac{1}{2\psi(0)} \int_X |1 - e^{j\omega t}|^2 e^{j\omega t} d\omega$$

$$= - \frac{1}{2\psi(0)} \int_X |1 - e^{j\omega\lambda}|^2 \cos \omega t \, d\omega \qquad (A.6.3)$$

the last step following from considerations of evenness of the integrand. But

$$\left| \int_X |1 - e^{j\omega\lambda}|^2 \cos \omega t \, d\omega \right| \leq \int_X |1 - e^{j\omega\lambda}|^2 |\cos \omega t| \, d\omega$$

$$\leq \int_X |1 - e^{j\omega\lambda}|^2 \, d\omega \qquad (A.6.4)$$

and the bound (A.6.4) is actually reached for $t = 0$. This establishes both the location and value of the maximum of $f_o(t)$.

To obtain a simpler explicit determination, we return to

(A.6.2) where, because of the symmetry of $\psi(t)$ about $t = 0$,

$$f_o(0) = \frac{1}{2\psi(0)} (2\psi(\lambda) - 2\psi(0))$$

so that the maximum value of f_o is given by

$$|f_o(0)| = 1 - \frac{\psi(\lambda)}{\psi(0)} \quad . \qquad \text{(A.6.5)}$$

BIBLIOGRAPHY

1. P. W. Bridgeman, The Logic of Modern Physics (The Macmillan Co., New York, 1927).
2. E. W. Hobson, The Theory of Functions of a Real Variable, 3rd Ed., Vol. I (Harren Press, Washington, D.C., 1950).
3. E. Mach, The Science of Mechanics (Open Court Publishing Co., Chicago, 1902).
4. M. Born, Experiment and Theory in Physics (Dover Publications, New York, 1956).
5. L. Brillouin, Science and Information Theory (Academic Press, Inc., New York, 1956).
6. L. Brillouin, "Editorial", Information and Control, 1, No. 1, 1957.
7. M. V. Cerrillo, "Sobre la Sintesis de Sistemas Lineales para la Transmision sin Retraso, Retrasada, y Prediccion Lineal de Señales", Revista Mexicana de Fisica, Vol. 4, No. 2, 1955. (English translation available in Research Laboratory of Electronics, Quarterly Progress Reports, Oct. 15, 1956, Jan. 15, 1957, April 15, 1957).
8. D. Gabor, Lectures on Communication Theory, Technical Report No. 238, Research Laboratory of Electronics, M.I.T., April 1952.
9. S. Winkler, "The Approximation Problem in Network Synthesis", Trans. IRE, Vol. CT-1, No. 3, Sept. 1954.
10. I. M. Jacobs, "Time Domain Synthesis by Area Approximation" S.M. Thesis, M.I.T., June 1957.
11. P. R. Halmos, Finite Dimensional Vector Spaces (Princeton University Press, Princeton, N.J., 1955).
12. J. von Neumann, Functional Operators, Vol. II (Princeton University Press, Princeton, N.J., 1950).
13. A. N. Kolmogorov and S. V. Fomin, Elements of the Theory of Functions and Functional Analysis, Vol. I (Greylock Press, Rochester, N.Y., 1957).
14. F. B. Hildebrand, Introduction to Numerical Analysis (McGraw-Hill Book Co., Inc., New York, 1956).
15. C. E. Shannon and W. Weaver, The Mathematical Theory of Communication (The University of Illinois Press, Urbana, Ill., 1949).
16. C. E. Shannon, "Communications in the Presence of Noise", Proc. IRE, Vol. 37, No. 1, Jan. 1949.
17. E. T. Whittaker, "On the Functions of Interpolation Theory", Proc. Roy. Soc. Edinburgh, 1915.

18. N. W. Lewis, "Waveform Computation by Time Series Method", Proc. IEE, Vol. 99, Part III, Sept. 1952. See also:

W. E. Thomson, "A Theory of Time Series for Waveform-Transmission Systems", IEE Monograph No. 53, Oct. 1952.

19. W. Hurewicz, "Filters and Servo Systems with Pulsed Data", Chapter 5 in H. M. James, N. B. Nichols, R. S. Phillips, Theory of Servomechanisms, Radiation Laboratory Series, M.I.T., Vol. 25 (McGraw-Hill Book Co., Inc., New York, 1947).

20. W. K. Linvill, "Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation", Trans. AIEE, Vol. 70, Part II, 1951.

21. J. R. Ragazzini and L. A. Zadeh, "The Analysis of Sampled Data Systems", Trans. AIEE, Vol. 71, Part II, 1952.

22. R. E. Wernikoff, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., April 15, 1955.

23. R. E. Wernikoff, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., July 15, 1956.

24. Tables of Sine and Cosine Integrals, National Bureau of Standards, 1942.

25. A. N. Kolmogorov, "To the Shannon Theory of Information Transmission in the Continuous Signals Case", Transactions, 1956 Symposium on Information Theory, M.I.T., Sept. 1956.

26. R. E. Wernikoff, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1957.

27. P. M. Woodward, Probability and Information Theory with Applications to Radar (McGraw-Hill Book Co., Inc., New York 1953).

28. J. H. Van Vleck and D. Middleton, J. Appl. Phys., Vol. 17, p. 940, Nov. 1946.

29. E. T. Whittaker and G. N. Watson, Modern Analysis, 4th Ed. (University Press, Cambridge (England), 1952).

30. E. L. O'Neill, "Spatial Filtering in Optics", Trans. IRE, Vol. IT-2, No. 2, June 1956. This paper is a good introduction to the relation between signal analysis and optics, and includes a short bibliography. See also:

M. Francon, Le Microscope a Contraste de Phase et le Microscope Interferentiel, Ed. du Centre National de la Recherche Scientifique, Paris, 1954.

31. Signal Theory Issue, Trans. IRE, Vol. CT-3, No. 4, Dec. 1956.

32. K. Knopp, Theory and Applications of Infinite Series, 2nd. Ed. (Blackie and Son, Ldd., London, 1951).