THE UNIVERSITY OF MICHIGAN

INDUSTRY PROGRAM OF THE COLLEGE OF ENGINEERING

SIGNAL EXTRACTION OF SYMMETRIC FUNCTIONS FROM NOISE

Richard S. Norman

December, 1967

IP-795

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Signals, Noise, and Prior Knowledge

In the collection and processing of data about any real system, the desired information is always perturbed or disrupted by irregularities or uncontrolled variation in the data. If some prior knowledge is available concerning the expected properties of the information sought, it may be possible to analyze the collected data to determine more from it than is obvious by direct inspection. In short, if you know what to look for, you will have an easier time finding it.

The data considered in this thesis are obtained by photographing symmetrical structures. The prior knowledge concerning the data is the knowledge of the symmetry of the object photographed. It will be helpful to keep in mind the idea that the data is a photograph, even though the notion of the observed data will be generalized considerably. In particular, the data are given over a region of space (the area of the photograph) and the prior knowledge is a knowledge of certain relations between the images at different spots on the photograph.

The development of information theory and the analysis of information handling systems has led to very powerful techniques for processing the data that arise in communications. Very general data handling problems can also be formulated within the framework of information theory, independent of the context of the application or the source of the data. Described in communications terms, the data processing problem becomes one of detecting and analyzing a "signal" in the

presence of obscuring "noise."

To make clear the connection between analyzing the photograph and processing communications data, consider the following paradigm for the observation of data (Fig. 1.1). The system under study produces (or emits or outputs) information, the signal, which is transmitted to the observer through a channel. The signal is that aspect of the system in which we are currently interested. Simultaneously the system may be producing extraneous information which is included as the noise source to the channel. In addition, other sources may contribute to the channel; these are all lumped together as noise. The result is the channel output, the observed data.

Of primary importance to us is the prior knowledge we may have concerning the system under study, the noise, and the channel. This prior knowledge may have been obtained previously through the channel, through a different channel, or may even be a hypothesis concerning the system. In the language of information theory, the presence of this prior knowledge means that the signal is redundant. This redundancy allows the recovery of part of the signal from the noise.

The type of prior knowledge usually considered in communications theory is the knowledge of the probability distributions of the signal parameters. In information theory, the signal may consist of discrete symbols and the prior knowledge is given in terms of joint probability distributions or groups or words of these symbols. In signal theory, the signal is considered a function of time and the prior knowledge is knowledge of certain correlation or power spectral density functions. In signal detection theory, the signal is of a
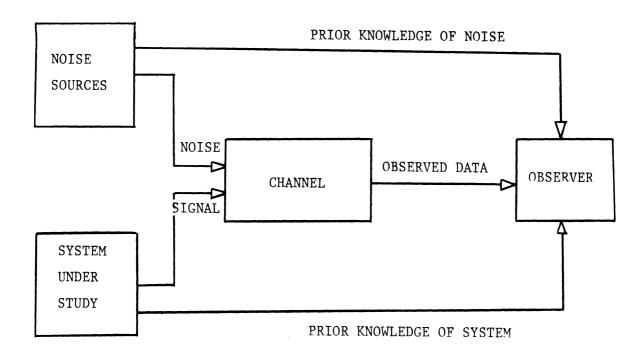
Figure 1.1.   Paradigm for Data-collection.

known form but with an unknown, random parameter such as its time of occurrence.

In the problem considered here, neither the statistical parameters nor the form of the signal is known a priori. Instead the signal is assumed to have a kind of internal structure, a symmetry, so that certain relations are known between one part of the signal and another.

There are two significant differences between the signals I will consider and those usually considered in communications theory. First, the prior knowledge concerning the signals in this thesis is given by geometric structure relations between different parts of the signal, rather than by assumptions on the statistical parameters of the signal. And second, the signals are defined in terms of space, rather than in terms of time. It is not meaningful to speak of one observation as occurring before another -- of data that have been observed in the past and data that have not yet been collected.

As a consequence of the non-temporal nature of the problem, two considerations of great importance in communications theory do not arise -- prediction and realizability. The prediction problem is the estimation of a future value of the signal on the basis of knowledge of past values of the data. The realizability criterion expresses the metaphysical idea that the processing cannot involve data not yet collected. Since the data are all obtained simultaneously, there is no prediction problem, and all the data may be used in the estimation of any value.

The idea of processing this type of data grew from the problem of analyzing electron micrographs of virus particles. In order to

provide motivation for the concepts to be developed, the structure of
viruses is briefly reviewed. The form of the present problem is derived
by considering the problems I encountered in trying to process micro-
graphs of viruses.

Virus Structure

Viruses are believed to be constructed from a core of nucleic
acid, containing the genetic information, and a protective coat of
protein, the capsid. This nucleo-capsid is sometimes surrounded by a
membrane. (See Horne, 1963 and 1964; Horne and Wildy, 1963.) Of parti-
cular interest in the processing problem are the so-called simple viruses
which almost invariably appear as rods or spheres in the electron micro-
scope. These may have a diameter from 150 Å to 1300 Å.

Early electron microscopic observation of these virus particles
did not disclose any internal structure, although a growing body of
evidence indicated that the protein coat was composed of a number of sub-
elements. X-ray diffraction studies, though, did show the shell to have
internal structure, leading Crick and Watson (1956,1957) to suggest
that the shell is composed of identical units packed together in an
orderly manner. The spherical viruses would have their subunits arranged
in a pattern similar to one of the regular polyhedra; the rod viruses
in a helical pattern. Two observations supported this hypothesis. First,
there was not enough genetic material to code a large number of different
proteins. Second, a regular arrangement of identical units would be a
minimum energy configuration, aiding in self-construction. Further
X-ray diffraction studies showed the internal structure of the spherical

viruses to be that of the icosahedron, one of the regular polyhedra.

Soon after, direct evidence for icosahedral symmetry was observed in the electron microscope, when a technique of double shadowing on the largest "spherical" virus showed it to have angular corners arranged in a pattern consistent with an icosahedral configuration.

With the development of the "negative staining" technique in 1959, surface structure became visible (Fig. 1.2). Since then, many viruses have been shown to be icosahedral and all viruses examined show that the surface is composed of an orderly array of small subunits, or capsomeres. The capsomeres are about 50 Å or 60 Å in diameter and may have a central hole of 30 Å or 40 Å. The number of capsomeres making up the capsid of spherical viruses varies from 12 to 812 (see Almeida, 1963; Hosaka, 1965).

The bewildering diversity in number of capsomeres was resolved by Caspar and Klug (1962), who found a rational basis for categorizing the capsomere counts and arrangements. Only certain arrangements and numbers are possible, and can be enumerated in advance.

The best resolution so far possible in electron micrographs of viruses is on the order of 20 Å, so that capsomere structure cannot be determined by observation. Indeed it is frequently not possible even to determine the number of arrangement of the capsomeres in the capsid (Howatson, 1965; Klug and Finch, 1965).

The limit on resolution is not set by the electron microscope, which has a theoretical resolution on the order of an Ångstrom (Heidenreich, 1964). Rather it is due to random noise, which may be caused by shot noise in the electron beam, photographic losses, and
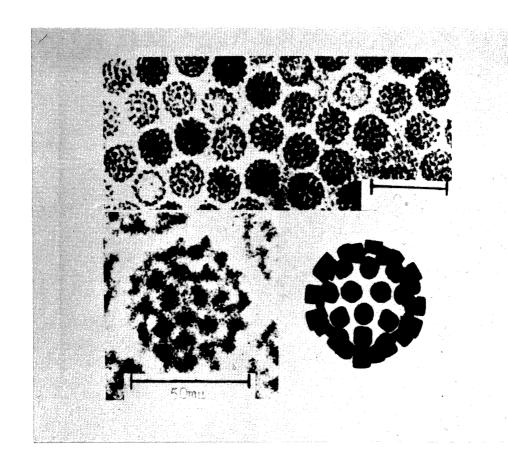
Figure 1.2. Electron Micrograph of Viruses. Upper, portion
of an electron micrograph field of human wart
viruses, negatively strained with phosphotungstic
acid. Scale: 0.1μ. Lower left, single virus
particle at high magnification. Scale: 50mμ.
Lower right, model in same orientation as par-
ticle. (Figure 3 of Williams, et. al., (1961),
reproduced with the kind permission of the
authors and Nature.)

random irregularities due to the staining technique and the supporting film. Valentine and Wrigley (1964) discuss the first two noise sources and some analysis has been made of the statistical properties of photographic granularity (Klein and Langner, 1963, and Thiry, 1963), but no data are yet available concerning the statistical properties of the staining and support film irregularities.

Processing Virus Micrographs

Since the object being examined has such a high degree of structure, the micrograph has a large amount of redundancy. It should thus be possible to recover some of the signal structure from the noise by suitable processing techniques. Two processing problems appear: signal classification to determine the arrangement or the number of capsomeres, and signal extraction to determine the structure of a single capsomere.

A processing scheme has been proposed by Markham, et al. (1963) to increase the resolution by utilizing part of this redundancy. This is a superposition technique which averages sections of the micrograph that correspond to each other. This technique was found to have several pitfalls (Agrawal, et al., 1965) which could lead to misinterpretation of the data.

My interest in the virus micrograph processing problem began with the recognition that this technique can be described in precise mathematical terms. The pitfalls in the technique could then be anticipated, hence avoided (Norman, 1966). The analysis lead me to investigate the possibility of using more sophisticated signal processing methods on these photographic data.

The power of signal theory in describing data processing rests largely on the idea of the Fourier transform, or spectrum, of a function. The data is represented as a sum of known standard functions with simple properties. The process of finding this representation is the transform, the coefficients in the representation are the spectrum. An especially important class of data manipulations, the linear operations, have a particularly simple form when expressed in terms of the transform; the data processing involves multiplying the data spectrum by the "system transfer function" of the manipulation.

A simple lens system can be used to find the transform of photographic data: the spatial distribution of light in the back focal plane of a lens is the transform of the light distribution in the front focal plane. Multiplication by a fixed function can be implemented by passing the light through a suitable photographic transparency.

Through certain tricks (which are themselves derived from communications ideas!) using coherent light (see Cutrona, et al., 1960) it is possible to multiply the light amplitude by negative or even complex values. These coherent optical techniques seemed to be particularly appropriate for processing the micrograph data.

If the original data has translational symmetry, corresponding to the periodic functions of signal theory, the transform is simple. The light energy in the transform plane is concentrated into discrete points, with no light energy appearing in between. The helical symmetry of the rod shaped viruses gives rise to a more complex spectrum, but one which has been determined for X-ray diffraction purposes (Cochran, et al., 1952). Klug and Berger (1964) took advantage of this fact and

used the optical transform of micrographs of helical viruses to make
certain measurements, although they did not attempt any actual processing.
Berger, et al., (1966) proposed using coherent light in the optical
system, but only for practical convenience, again not utilizing the pro-
cessing capabilities. Besides this optical transform technique and
Markham's superposition technique, no processing method has been proposed
to utilize the redundancy in the virus micrograph.

One of the difficulties in applying transform methods is that
if the original symmetry is rotational, the transform also has rotational
symmetry and no simplification is effected. The superposition technique
succeeds because, in polar coordinates, rotational symmetry is reduced
to translational symmetry in the angular coordinate. If the original
symmetry is more complex, e.g., icosahedral, no general expression for
the structure of the transform has yet been obtained. Moreover, no
simple coordinate change will reduce the problem to translational
symmetry.

My original intention was to perform both the signal class-
ification and the signal extraction problems using the coherent light
techniques. The classification problem was to be handled using the results
of signal detectability theory -- the matched filter (Middleton, 1960;
Kozma and Kelly, 1965). Filters were made from models of the appro-
priate arrangements so that the correlation coefficient between the
data and each arrangement could be determined. These models were based
on structureless capsomeres (Fig. 1.3). The transforms of some of the
virus models are shown in Fig. 1.4. Although the patterns are complex,
it can be seen that different arrangements have their energy concentrated
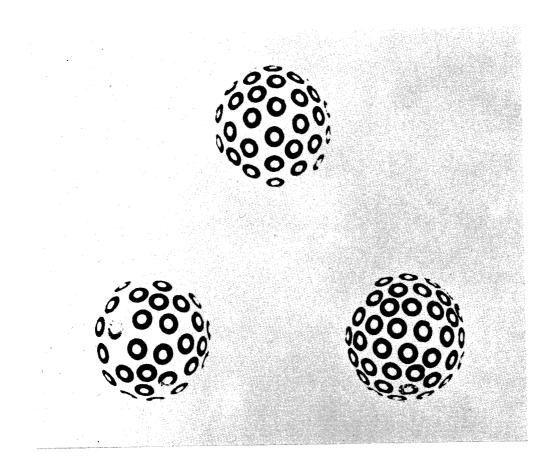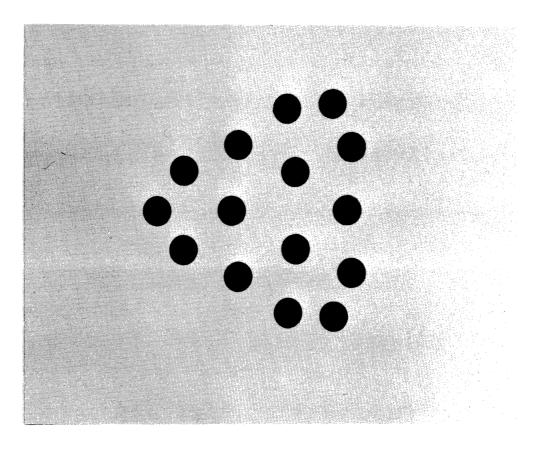
Figure 1.3. Virus Models. Constructed from flat
capsomers placed on a spherical surface.
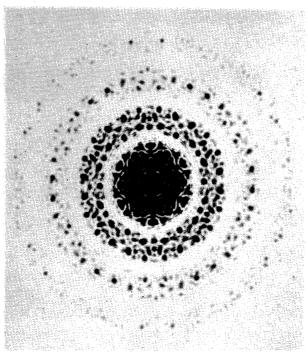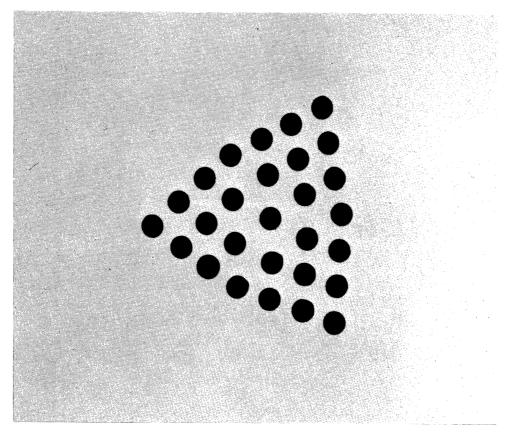
Figure 1.4a. Optical Transforms of Virus Models. Upper,
central region of model. Lower, optical
transform of the model. (a) Arrangement
with 42 capsomeres. Three-fold rotation
axis central. Model based on spherical
surface.
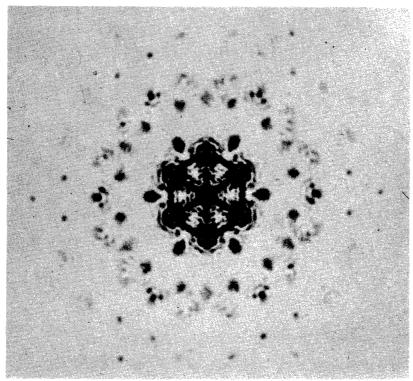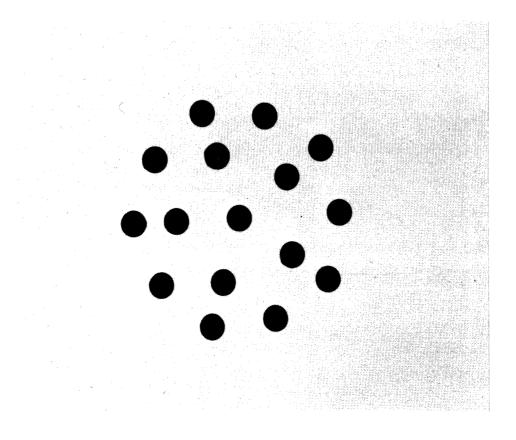
Figure 1.4b. Optical Transforms of Virus Models. (b) Arrange-
ment with 92 capsomeres. Three-fold rotation
axis central. Model based on icosahedral surface.
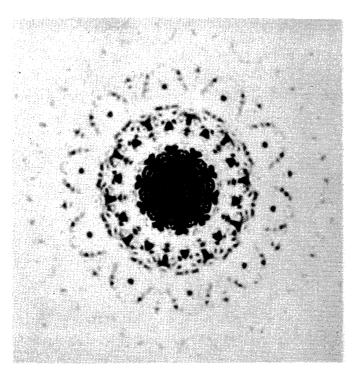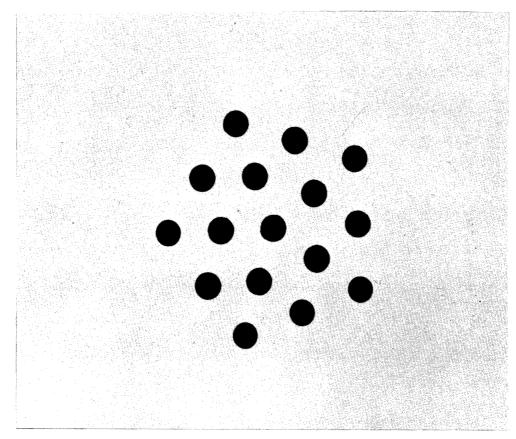
Figure 1.4c. Optical Transforms of Virus Models. (c) Arrange-
ment with 42 capsomeres. Five-fold rotation axis
central. Model based on spherical surface.
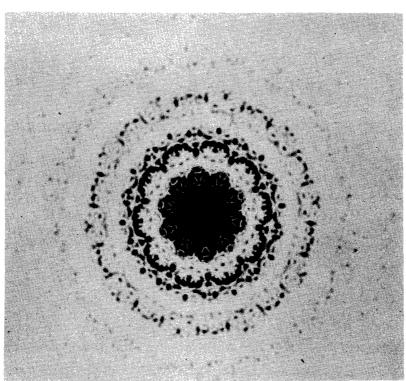
Figure 1.4d. Optical Transforms of Virus Models. (d) Arrange-
ment with 42 capsomeres. Five-fold rotation axis
central. Model based on icosahedral surface.

in different places. A working hypothesis was assumed: the structure

of the capsomere does not influence the basic pattern of the spectrum,

only the relative intensities of the different parts of the pattern.

On this basis, a filtering operation suggested itself. Since the

transform of the model shows that no light should be expected in certain

areas, any light actually appearing at those places should be due to

noise. The appropriate filter should then block light from passing

where it is not expected, allowing it to pass where it is expected. This

filtering process is very similar to an "optimal" filter derived by

Wiener (1949), the so-called Wiener smoothing filter. I expected to

be able to justify my procedures, first by proving the working hypo-

thesis, then by demonstrating the applicability of the matched and

the Wiener smoothing filters in this application.

Two mathematical problems had to be solved. First the spectrum

or transform of icosahedrally symmetric structures had to be determined.

And second, the assumptions on which the Wiener filter is derived had to

be reconciled with the knowledge of the symmetry of the virus data.

In considering the first problem, I came to realize that the

icosahedral symmetry of the data was not consistent with the concept

of the Fourier transform. The representation in terms of the exponen-

tial function is intimately linked with translational symmetry, as seen

by the very simple structure of the transform of periodic functions.

Although each data function has a transform, two functions almost

identical under icosahedral symmetry may have vastly different transforms.

In fact, the regularities that exist between regions of icosahedrally

symmetric data lack some of the mathematical prerequisites for performing

the Fourier transformation; the domain of the data is not a topological
group.

Considering the second problem, I recognized that the difference
between the formulation of the virus processing and the Wiener proces-
sing problems is in the nature of the prior knowledge. The Wiener filter
is derived using the statistical structure of the signal and the noise
as the prior knowledge. In the virus problem, the prior knowledge is
the symmetry relation in the underlying structure. There is no a priori
reason to believe that a Wiener filter, derived under such different
considerations, would be at all appropriate for my purposes.

In the process of solving these two mathematical problems, I
came to realize that the optical processing techniques were not appro-
priate for the virus problem. The full analysis of the micrographs would
require digitizing the data and performing the analysis by computer.
The results would probably not justify the tremendous effort required.
However the concepts required in the solution represented a distinct
generalization of signal theoretic ideas. The present thesis is a
consideration of this generalization; the specific problem, the virus
micrograph, will no longer be considered.

The General Processing Problem

Two general problems have to be answered: (1) What kind of
signal processing can be performed on data which is defined on a mathe-
matical structure weaker than a topological group? (2) What is the
relation between such processing and the Wiener smoothing filter?

The development of these questions from the virus problem
has led to the notion of symmetric functions and to a generalization

of the types of spaces these functions are defined on. Since the data

are spatial, not temporal, the ideas of prediction and of realizability

do not arise. Because these latter concepts create some of the most

difficult aspects of the general processing problem in signal theory,

the generalization in the statement of the problem has been obtained at

the expense of the demands on the solution.

The idea of a symmetric function in the present context is

novel, so that techniques of processing had to be developed. These

techniques are essentially parallels of those involved in classical

signal theory, particularly the introduction of abstract function spaces.

It is thus natural to sketch some of the history of signal processing,

following the development of the concepts, which I will be using. The

concepts themselves will be defined at the time they are used.

The ideas of processing random signals and noise begin with

Wiener's work (1949) on stationary time signals, where he applies his

development of random harmonic analysis (Fourier analysis of random

signals) to the engineering problems of prediction and smoothing. The

engineering or communication theory aspect of processing was generalized

to include certain kinds of deterministic signal functions (Zadeh and

Ragazzini, 1950) until now it is a standard part of signal theory

(Middleton, 1960; Davenport and Root, 1958). Particular specializations

of interest include multivariate analysis, or simultaneous processing

of several signals (Masani and Wiener, 1957, 1958), and "distortionless

filtering" (Bendat, 1957; Stewart and Parks, 1957).

The mathematical development of similar ideas is intimately

bound up in the representation of random or stochastic processes in an

abstract mathematical space, a Hilbert space (Karhunen, 1947). Hilbert space notions had been developed to handle linear operations in classical analysis as well as for the development of Fourier analysis in a generalized and abstract manner (see Murray, 1941; Halmos, 1957).

Although Wiener did not stress Hilbert space methods in his 1949 work, it was implicit in the relation between abstract harmonic analysis, generalized Fourier analysis, and the Hilbert space representation of the noise process and the linear manipulations on the data. The explicit nature of abstract space ideas in prediction and filtering was brought out by Doob (1953), Grenander and Rosenblatt (1957), and Loève (1960).

Meanwhile, development of Hilbert space representations of random processes continued (Cramér, 1951; Doob, 1953; Loève, 1960; Kawata, 1965), particularly on weakening the requirements on the stationarity of the process; the parameters of the random process could change with time. In addition, the random process could be defined as arbitrary mathematical functions, not only as functions of time.

The introduction of the reproducing kernel Hilbert space (Aronszajn, 1950) lead to new applications of the abstract space ideas to signal theory (Parzen, 1961 and 1962). These developments were brought into the engineering literature just recently (Capon, 1965).

The "classical" approach of Wiener and followed by the engineering texts derives integral equations for the processing operations. The processing is optimized using techniques of calculus of variations. In the abstract space approach, the data functions are represented as points in a space and the optimization is done geometrically, by minimizing

the distance between two points. The latter approach is the one that will be used here.

The Solution

The first step in this thesis is the exact formulation of the notion of symmetry and of symmetric functions. The data will be considered to be a function defined on a mathematical space, the data space. The nature of this space is left unspecified, but it must have some mathematical structure to allow data manipulation. The notion of symmetry is expressed in terms of geometric relationships in this data space.

The observed data is represented as a point in an abstract function space, the data space. The data is assumed to be the sum of some specific, but unknown, symmetric function and a random noise function. The processing problem can thus be expressed in geometric terms: given the observed data point, find the signal estimate that lies closest to the actual signal point.

In order to carry out the geometric interpretation, a notion of distance is needed in the function space. This distance notion is to correspond to our intuitive ideas of how difficult it is to distinguish between two functions. Two functions that are close together are easily confused, while functions distant from each other are easily resolved. Because the difficulty in resolving two functions is caused by the noise, which may be different in different parts of the space, the distance function must depend on the noise process. Optimizing the solution by selecting the signal estimate nearest the actual signal point is interpreted as minimizing the mean square error between the estimate and the signal.

Finding the estimate depends on the notion of the signal, or symmetric function, subspace. This subspace consists of those functions which satisfy the symmetry requirement. The actual signal point must lie in this subspace.

The processing proceeds by recognizing that the signal subspace can be segregated from the non-symmetric functions in a natural way. The problem of finding the optimum signal estimate is then shown to be simply the problem of finding the point in this signal subspace which lies closest, according to the distance measure, to the data point. The estimate is found by the projection operation, whereby the data point is projected perpendicularly down into the signal subspace, the foot of the perpendicular being the estimate. Integral equations are then derived which completely determine the processing procedure.

Results and Conclusions

The optimal processing procedure is completely specified by three conditions: symmetry, idempotency on the symmetric function subspace, and self-adjointness. The first condition states that the signal estimate is itself a symmetric function -- if P is the projection operator and d is the data function, then the signal estimate $\hat{s} = P[d]$ is symmetric. The second condition states that if s is any symmetric function, then $P[s] = s$. The term "idempotent" is used to describe the fact that $P^2 = P$. The third condition expresses the perpendicularity of the projection. If R is the noise covariance function, interpreted as a linear operator, then $P R = R P^{ad}$, where P^{ad} is the adjoint of P.

A special case of the general processing problem arises if the data are a finite set of numbers. The three processing conditions can then be expressed as a matrix equation, the function space being a finite-dimensional vector space. The optimal processing operator may be found explicitly by solving a finite set of simultaneous linear algebraic equations.

In another special case, the data consist of a finite set of functions. This case may arise in considering a replicated experiment, each trial producing a data function. The processing operator again may be expressed as the solution to a matrix equation, with the matrix entries now being functions, rather than numbers. The operator may be found by solving a finite set of simultaneous linear integral equations. In particular, the Wiener multivariate smoothing problem may be expressed readily in this context and the smoothing filter obtained as the solution to a symmetric function smoothing problem.

Another special case arises when the data space is a topological group, for then Fourier analysis is possible. However, the power of Fourier analysis is lost unless the symmetry structure is compatible with the group structure of the data space. The only type of symmetry compatible with topological group structure is translational symmetry. Although many important types of symmetry are translational -- periodic functions of time, for example -- the virus symmetry is not, so that Fourier analysis is not appropriate for processing the virus micrographs.

The idea of a symmetric signal may be expressed, loosely, by saying the signal is repeated over the domain of the data. A naïve approach to estimating the signal -- here called the "trivial" processing

technique -- is simply to average over all the repetitions. In general,
the optimal processing solution weights the repetitions differently,
depending on the relative amplitude of the noise over the various repeti-
tions. Moreover, the optimal estimate of the signal at one point may
depend on all the data, even data values which appear at first to be irre-
levant to the estimate since they occur at points which are not related
to the estimated point by the symmetry.

Under many important special cases, the optimal processing
procedure reduces to this "trivial" operation. These cases include
white noise and noise for which the covariance function is symmetric.
If the data space is a topological group, the concept of stationary or
wide-sense stationary noise can be defined; for these cases the "trivial"
processing is again optimal. Furthermore, if the processing technique
is restricted to be stationary, or translation invariant, the "trivial"
processing is the best that can be done. This latter result has
important practical consequences, since the optical technique, as well
as ordinary electronic devices, have the stationarity property.

The processing procedure is solved for an example in which a
periodic signal function is observed over three periods. The noise is
assumed to be uncorrelated between points, but varies in amplitude over
the observation period, reaching a minimum of zero at one point. The
estimate of the signal at this point is, of course, the data value.
The data with the least noise is weighted most heavily in computing the
signal estimate but, interestingly, the data with the most noise is not
weighted least.

The results of the study indicate that the optical processing
technique is inappropriate for the processing of virus micrographs on

two counts. First, the optical technique is based on the Fourier transform, and the virus symmetry is not compatible with the transform. Second, the processing operation is restricted to the "trivial" one, since the optical processing technique is translation invariant.

The idea of a symmetric signal covers more than just photographs or diffraction patterns of geometrically symmetric structures. It includes statistical sampling, multivariate or replicated data, and periodic time functions. Under the unifying mathematical development presented here, it may be possible to apply results from one to another, seemingly unrelated area. In particular, the solved example derived a processing filter for time-varying noise in a relatively easy way. This approach may prove to be useful in the analysis of other time-varying problems concerned with periodic functions.

CHAPTER II

THE PROCESSING PROBLEM

A. The Structure of the Data

The Data Space

The signal, the noise, and the observed data are the functions which will be considered in this thesis. A function, $f(x)$, is a rule which assigns a real or complex value to each point x in a data space X. In the fullest generality, the space X is left undefined, although it must have some mathematical structure in order to perform the desired manipulations on the data.

> Definition 2.1. The data space, X, is a countable union of disjoint connected Hausdorff spaces with a Borel measure.

It is convenient to take the measure of the whole space as one. For bounded data spaces this poses no problem. In unbounded cases this may be accomplished by a procedure illustrated by the distinction between functions of finite energy, $\int_{-\infty}^{\infty} |f(t)|^2 \, dt < \infty$, and those of finite power, $\lim_{T \to \infty} (1/2T) \int_{-T}^{+T} |f(t)|^2 \, dt < \infty$. The latter type of integral corresponds to making the measure of the entire real line equal to one.

An important type of data space arises when an experiment is replicated several times. Consider the data obtained on the first trial. This data will be defined over a space, X_i. Similarly the data from the i-th trial is defined over the space X_i. Assuming that no experimental conditions change between trials, all the X_i spaces should have the same structure. The data space X corresponding to

the totality of the data is the union of all the spaces from the separate replications, $X = \bigcup_i \{X_i\}$.

In communications theory the functions are usually functions of time, functions of a single real variable. Sampled data are defined only at discrete points in time, the data space consisting of the set of these discrete points. If each trial of a replicated experiment produces a time function as data, the data space consists of a union of lines. The same data space occurs if a single signal is monitored simultaneously on separate channels, the multivariate signal case.

In optical data processing, the functions are defined on two-dimensional spaces, usually the surfaces of photographs. Three-dimensional data spaces may also enter into optical data processing by a consideration of emulsion thickness (van Heerden, 1963), by photography of three-dimensional objects as in holography (Upatnieks and Leith, 1964), or by considering time varying two-dimensional figures. Three-dimensional data spaces also occur in X-ray diffraction problems (Buerger, 1956). The functions involved in the structure of viruses are defined on a sphere or a cylinder. The virus micrographs, being photographs, are functions defined over the face of the micrograph plate.

Symmetrical Functions

The structure to be assumed for the signals is a generalization of the notion of periodicity. A periodic function of a real variable, say $\sin(x)$, is left unaltered if the data space, the real line, is moved a certain distance to the left or right: $\sin(x + 2\pi) = \sin(x)$.

Similarly a cube is left unaltered if it is rotated 120° through a body diagonal. These are rigid body motions of Euclidean data spaces.

The concept of motion is now generalized to include all transformations that do not involve bending or tearing.

> Definition 2.2. An automorphism, $a: X \to X$, is a 1:1 continuous mapping of X onto itself. If x' is the image of x under a, then $a: x \to x'$ or $x' = x^a$.
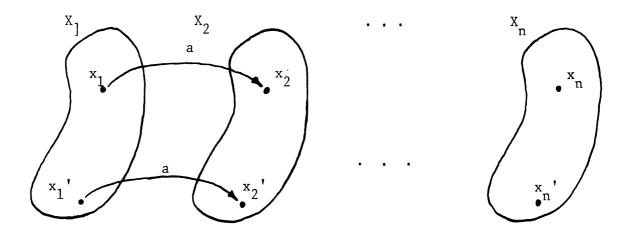
The automorphism $a: x \to x'$ acts on a function $f(x)$ defined on X to form a new function, $f'(x) = f(x')$. It may happen, as was the case with $f(x) = \sin(x)$ and $x' = x + 2\pi$, that $f'(x) = f(x)$.

> Definition 2.3. A function $f(x)$ on X is invariant with respect to a, an automorphism of X, if $f(x) = f(x^a)$ for all x in X.

Two automorphisms can be applied in succession to form a third, the product. If a function is invariant with respect to each of the two, it is also invariant with respect to the product. A function invariant with respect to an automorphism is also invariant with respect to the inverse automorphism. All functions are invariant with respect to the identity. Thus a function invariant with respect to a set of automorphisms is invariant with respect to the group generated by that set. Conversely, the set of all automorphisms which leave a function invariant is a group.

> Definition 2.4. A symmetry group, G, is a discrete group of automorphisms of a space, X. The Borel measure of X is assumed to be invariant with respect to the elements of the symmetry group.

By this definition, the idea of symmetry is expressed in terms of the algebraic structure of the data space. The relation between this imposed structure and any pre-existing algebraic structure of the

The whole data space is the union of these subregions.

Figure 2.1. Data Space Corresponding to the Replicated Experiment.

data space will be explored in a later chapter. The types of groups

that are to be studied are all discrete, so that no processing capability

is lost by restricting the symmetry group to be discrete at this stage.

By making the Borel measure invariant under the symmetry group, the

automorphisms are, in effect, rigid motions that do not distort distance.

> Definition 2.5. A function is symmetric with respect
> to a symmetry group, or more simply is symmetric, if
> it is invariant with respect to all the automorphisms
> of the group.

This concept of symmetry as invariance under a group of auto-

morphisms is important in virtually all branches of mathematics. The

concepts of physical law and prior knowledge, the foundations of

modern physics, are based on just this concept of symmetry (Weyl, 1952).

It is not surprising, then, to find symmetry arising from such diverse

situations as periodic functions or replicated experiments, as well

as measurements made on regular structures such as crystals and viruses.

The Unit Cell

Symmetry has been defined as an algebraic concept. In this

section, symmetry will be expressed in terms of the geometry of the

data space. The goal is to find a way of expressing the data space in

terms of "building blocks", each block being identical to every other

block.

The data space in the case of a replicated experiment has

already been described as the union of subspaces, X_i, each resulting

from a single trial. Let x_i and x_j be corresponding points in the

subspaces X_i and X_j (see Fig. 2.1). The signal function $s(x)$
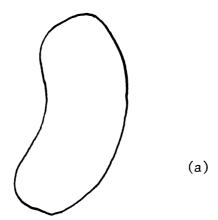
must have the same value on x_i and x_j, $s(x_i) = s(x_j)$, so that the automorphisms of the symmetry group must be of the form, $a:x_i \to x_j$. Since all of the spaces X_i are equivalent, the automorphism group consists of all the ways of shuffling these subspaces, leaving the structure of the subspaces unchanged. If there are m replications, the symmetry group is the full symmetry[1] on m items, the group of all permutations of the subscripts.

Any one of these subspaces can be taken as a building block for the whole space in that X is composed of a series of these blocks which are completely equivalent by the symmetry. Any symmetric function on X is completely specified merely by defining its values on this building block, since the function must repeat these values on all the other blocks.

In the replicated experiment situation there is a natural way of choosing the building block. In general, though, there is no reason why the natural block, shown in Fig. 2.2a, should be chosen rather than the chopped up block of Fig. 2.2b. Either selection is adequate for describing the symmetry and structure of the data space. There is no a priori way of choosing a single building block description out of all the possibilities.

In addition to not being able to select a unique representation, two further difficulties arise in the building block concept. The first arises if the data space contains "singular" points under the symmetry group.

---

[1] The word "symmetric" is used here in a technical sense (Hall, 1959). In this case, the symmetry group is the symmetry group.

(a)

(b)
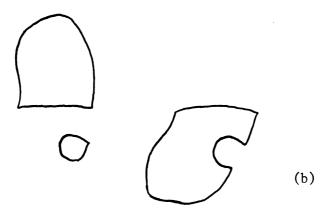
Figure 2.2. Two Possible Building Block Configurations for the
Data Space of Figure 2.1.

Definition 2.6. A point x of X is singular with
respect to the symmetry group G if there exists in
G an automorphism, a, not the identity, such that
a:x → x.

All the points on a rotation axis or on a plane of reflection are singular

points. The difficulty is caused by the fact that two building blocks

must overlap at singular points. The topological structure of the

building block at these singular points becomes very complex. The

second difficulty is caused by the fact that the overall geometry of

the building block can be rather different from that of the data space.

This situation occurs when two different points on one building block

are equivalent and will be illustrated later.

To avoid these complications, the building block idea is

dropped in favor of the mathematical idea of a quotient space, the unit

cell space. This space may be considered to represent all the equivalent

building blocks superimposed; it has all the properties and only those

properties that all the building blocks share.

Definition 2.7. Two points, x and x', of X are
equivalent under the symmetry group G, or simply
equivalent, if there exists an automorphism in the
symmetry group such that a:x → x'.

Definition 2.8. The unit cell space, Y, of X
under the symmetry group G is the set of equi-
valence classes of X under G. The topology of
Y is that induced from the topology of X.

In topological terms, the space X forms a covering space

for Y. Several topological structure results are of interest. If

X is connected, then Y is not simply connected. Conversely, if Y

is simply connected, then X is composed of disjoint subregions and

the building block type of structure of X is valid. Further topolo-
gical structural relations will be considered in a later chapter.
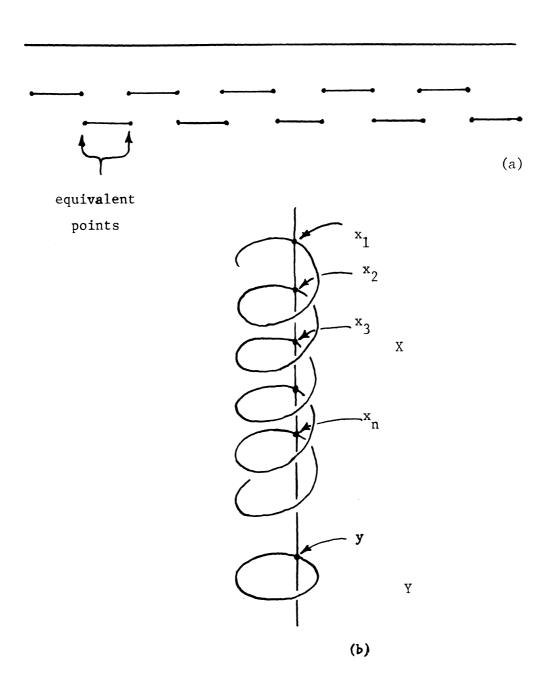
The unit cell space for the replicated experiment situation
looks superficially the same as the "natural building block" of Fig. 2.2a.
A point y in Y, though, represents an equivalence class of points of
X, rather than a single point of X.

The unit cell space corresponding to the periodic functions
of a real variable is a circle whose circumference is the period. The
data space is the real line. According to the building block concept,
the data space is composed of a sequence of segments, each of length
equal to the period, joined end to end (see Fig. 2.3a). But the end
points of any segment are equivalent, so that topologically they may be
considered joined, forming the circle. In this sense, the geometry of
the building block differs from that of the data space. This case also
illustrates the point that a connected data space gives rise to a non-
simply connected unit cell space.

A more accurate description of the formation of the unit cell
in the periodic function case is illustrated in Fig. 2.3b. The data
space is drawn as a helix so that the distance around one turn is equal
to the period, and equivalent points lie on a vertical line. The unit
cell is the set of these vertical lines, a cylinder. Since the vertical
dimension has no significance, it is dropped and the set of lines becomes
the set of points shown, the circle.

B.   The Noise Process

The noise is assumed to be a sample function from a set of
random variables, one for each point of the data set. At any particular

(a)

equivalent
points

(b)

Figure 2.3. Periodic Function Data Spaces.
(a) "Building block" representation of periodic function
data space.
(b) Quotient space representation of periodic function
data space and unit cell space.

point x of the data space, the noise value n(x) is a random variable

taken from a probability distribution which depends on x and possibly

also depends on the values of the noise at other points. It is assumed

that the expected value of n(x), that is the mean, is zero at all points

and that the probability distribution is independent of the signal func-

tion. The noise must have finite second moment at all points, but it is

not assumed to be stationary, either in the full or the wide sense, nor

are ergodic properties assumed.

The assumption concerning the independence of signal and noise

is necessary to develop the processing procedure without undue compli-

cations. It is recognized that non-linearities in the channel may cause

larger noise amplitudes at small signal amplitudes than at large, or

vice versa. In a photographic system the noise may well depend on the

intensity of the signal in that two different noise mechanisms may

apply. At small signal values the noise will be a background of dark

granules, hence the mean is positive, while at high signal levels it

is "drop-out", with a negative mean. This situation will apply to all

data which are restricted in range -- at extreme values of signal, the

noise will have a non-zero mean tending to bring the data back to the

center of the range. In electron microscopy, a large noise component

is the shot effect of individual electrons in the electron beam, so

the noise amplitude again depends on the strength of the beam, hence

the signal.

The prior knowledge that is assumed about the noise, in

addition to the assumptions of independence from the signal and zero

mean, is the covariance function, $R(x_1, x_2) = E[n(x_1) n^*(x_2)]$, where

the asterisk denotes complex conjugation. The expected value is taken across the probability distribution, not over the data space.

The covariance function is assumed, for convenience, to be strictly positive definite. This means that no linear combination of noise samples will cancel to zero, not only in the mean but for every sample.

> Assumptions on the noise. The noise is a sample function from a set of random variables with zero mean, finite second moment, independent of the signal. The covariance function is assumed known and strictly positive definite.

In signal theory, great emphasis is placed on stationary or wide sense stationary noise processes. If the noise is stationary, the probability distribution underlying the noise is independent of location in the data space -- the noise has the same structure at all points. For wide sense stationarity, only the first two moments need be stationary, other noise parameters may vary at different locations in the data space. In particular, either type of stationarity implies that $R(x_1,x_1) = R(x_2,x_2)$ for any x_1 and x_2 in X.

It is not possible to express completely the notion of either type of stationarity in the data space without some additional algebraic structure. Either condition implies that $R(x_1,x_2) = R(x_3,x_4)$ if $x_2 - x_1 = x_4 - x_3$. In the general data spaces so far considered, it is not possible to subtract two points of the data space.

It is possible, though, to define the notion of periodic noise:

> Definition 2.9. The noise process is symmetric with respect to the symmetry group G, or more simply symmetric, if $R(x_1,x_2) = R(x_1{}^a,x_2{}^a)$ for all a in G.

The noise process is symmetric if it has the same structure at all equivalent points. Note that this is not the same as saying that each sample function is symmetric. It will be shown later that symmetry is a weaker condition than weak sense stationarity, in that if the noise is weak sense stationary it must necessarily be symmetric.

C. Data Processing

The Function Space

The basis for the data processing is the construction of an abstract space, the function space $\mathscr{H}$. The possible data functions, the signal functions, and the noise sample functions are all included in this space as separate points. The space must also include all linear combinations of points within it: if $f(x)$ and $g(x)$ are represented as points in the function space, then the function $\alpha f(x) + \beta g(x)$ must also be a point, where α and β are complex numbers. It is also necessary that limits of Cauchy sequences of functions exist in the function space. Integration is then possible, so that the function:

$$f'(x) = \int_X h(x,x') \, f(x') \, dx'$$

is a point in the space.

The functions in the space are assumed to have the properties necessary for convergence of all integrals and limits. This assumption is usually expressed by limiting the functions to those that are integrable square, or L_2. No great emphasis is placed on these mathematical niceties since the data in practice must satisfy the assumptions.

The data is bounded, of bounded variation, and at least piecewise contin-
uous, being derived from physically occurring variables. The data space
is actually compact in any real situation, since it must be a bounded
region of a finite-dimensional space. The resolution of observation is
limited so that data occurring at one point is related to the data
occurring at nearby points. It is possible to make only a finite number
of independent observations on the data. The actual number of these
"degrees of freedom" in the data is not important, it is sufficient that
the number be countable for all the ensuing mathematical operations to
be valid without further restrictions.

Some geometric structure is needed in this function space.
First, the notion of distance is required, so that it is possible to
say how similar or dissimilar two functions are. Furthermore, the notion
of an inner product is needed. With an inner product it is possible
to say, not only that two functions are so far apart, but whether the
dissimilarity is due to differences in the nature of the functions or
in their amplitudes. Mathematically, the inner product provides for
the notion of direction in the space: two points f and g are
orthogonal if their inner product is zero. The inner product between
two functions $f(x)$ and $g(x)$, written[2] $(f(x),g(x))$, is a complex
valued bilinear, skew-symmetric, function on the function space. That
is, it must satisfy the conditions:

$$(\alpha f_1 + \beta f_2, g) = \alpha(f_1, g) + \beta(f_2, g), \quad \text{where} \quad \alpha \quad \text{and} \quad \beta$$

are complex numbers, and $(f,g) = (g,f)^*$.

A "norm", written[2] $\|f(x)\|$, can be defined from the inner

---

[2] When several different spaces are being considered, a subscript is
used to identify which inner product or norm is meant; i.e., $(f,g)_{\mathcal{H}}$
or $\|f\|_{\mathcal{H}}$.

product for all functions in the space: $\|f(x)\|^2 = (f(x),f(x))$. The

norm must be a positive definite quadratic form: $\|f(x)\| \geq 0$, with

equality holding if and only if $f(x) = 0$. The distance between two

functions can now be defined as the norm of their difference:

$\|f(x) - g(x)\|$.

> Definition 2.10. The function space, $\mathscr{H}$, is a complete
> linear inner product space; a Hilbert space. The points
> in $\mathscr{H}$ are functions defined on the data space.

In the processing problem, we are given an observed data

function which is the sum of a particular, but unknown, signal and

noise. The signal is a symmetric function; the noise a sample function

from the random process. We are to find the "best" estimate of the

actual signal. To optimize the estimate, a quantitative error value

is assigned to every possible estimate, the best being the one with

the smallest mean square error value. The idea now is to choose the

function space so that the distance between the signal function and

the signal estimate is the desired quantitative error value. The

problem then becomes a geometric one: given a data function point,

find that point of the function space which lies closest to the signal

point and which can be found by knowing only the data.

The notion of the distance between two functions is inter-

preted to mean the ease with which the two deterministic functions

can be resolved in the presence of the noise. In the situation illus-

trated in Fig. 2.4, the difference between f and g_1 appears to be

smaller than the difference between f and g_2. Thus f should be

more easily confused with g_1 than with g_2. However the discrepancy,
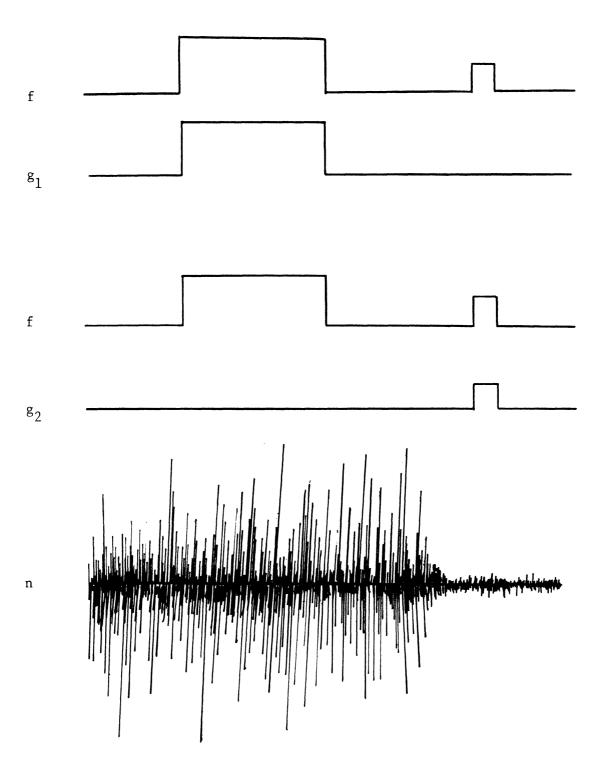
Figure 2.4. Resolving Signals in the Presence of Noise.

or error, between f and g_1 occurs at a place where the noise is
small, while the error between f and g_2 occurs where the noise is
large. Thus f and g_1 are more easily resolved than f and g_2.
The notion of distance, hence of inner product, must depend on the mag-
nitude of the noise at each point.

If the noise has the same magnitude at each point, and if
the noise at one point is uncorrelated with the noise at any other point,
then the noise is said to be white. The noise amplitude may be normalized
to have unit variance at each point, $R(x,x) = 1$, without loss of
generality. With white noise, the distance between two functions depends
only on the difference between them, not on where in the data space the
difference occurs. The square of the distance can be taken to be:

$$\|f(x) - g(x)\|_{\mathcal{W}}^2 = \int_X E\left\{|f(x) - g(x)|^2\right\} dx$$

corresponding to the inner product:

$$(2.1) \qquad (f,g)_{\mathcal{W}} = \int_X E[g^*(x)\ f(x)]\ dx.$$

The subscript, $\mathcal{W}$, indicates that this inner product is derived from the
white noise process. The function space consisting of the functions on
X with this inner product will be denoted $\mathcal{W}$. Note that this inner
product definition, without the expected value operation, is the usual
one for L_2 function spaces.

Two mathematical concepts will be useful for the further
development of the function space: the notion of a coordinate system
or a "basis", and the "associated linear operator" of an inner product.

A set of linearly independent points, b_i, in a space is said

to form a basis for that space if any point, p, in the space can be

written as a linear combination of the basis elements:

(2.2) $\qquad p = \Sigma_i \, \alpha_i \, b_i .$

The vector from the origin to b_i can be considered to be a coordinate

axis, with α_i the coordinate of p along this axis.

The data space X can be thought of informally as forming a

"span" for $\mathcal{H}$. Any point f in $\mathcal{H}$ is a function f(x); the coordinate

in the x_i direction being the value of the function $f(x_i)$. The actual

span element corresponding to x_i is the function:

(2.3) $\qquad f_i(x) = \begin{cases} 0, & x \neq x_i \\ 1, & x = x_i \end{cases}$

The span is reduced to a basis by considering only linearly independent

points of the form given in Eq. 2.3.

A basis is called orthonormal if:

(2.4) $\qquad (b_i, b_j) = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$

Interpreting Eq. 2.4 geometrically, a basis is orthonormal if two different

coordinate axes are perpendicular (orthogonal) to each other and if the

unit length along each axis (the norm of the basis element) is one. If

the coordinate system is orthonormal, the inner product is that of

Eq. 2.1. The white noise space, $\mathcal{W}$, has an orthonormal coordinate system.

The second concept is that of the associated linear operator,

corresponding to an inner product (cf. Friedrichs, 1960). The functions

f(x) and g(x) are points, f and g, in the white noise space $\mathcal{W}$

as well as the function space $\mathcal{H}$. The idea of the associated operator

states that the inner product of $\mathcal{H}$ can be expressed in terms of a bounded, self-adjoint, linear operator in $\mathcal{W}$:

$$(2.5) \qquad (f,g)_{\mathcal{H}} = (f,Hg)_{\mathcal{W}} = (Hf,g)_{\mathcal{W}}$$

In terms of integral operators, H becomes H(x,x'):

$$(2.6) \qquad (f,g)_{\mathcal{H}} = \iint_X E\,[g^*(x)\,H(x,x')\,f(x')]\,dx\,dx'.$$

In terms of a coordinate system, the inner product of two basis elements is $(b_i,b_j)_{\mathcal{H}} = H(x_j,x_i)$. The coordinate system is orthonormal if and only if H is the identity operator.[3]

The remainder of this section is concerned with deriving the form of the inner product for $\mathcal{H}$. It will be shown that, if the covariance function, R(x,x'), is considered to be a linear operator, then the operator H is the inverse of R: $H = R^{-1}$.

This derivation is easily provided if a formulation for the space $\mathcal{H}$ can be found in which the noise becomes white. This formulation will be given by a transformation of variables, a change to an orthonormal coordinate system. Assume for the moment that such a transformation, T, is found so that n(x) is tranformed into a white noise process N(t):

$$N(t) = \int T(t,x)\,n(x)\,dx$$

$$\text{and} \quad E[N(t)\,N^*(t')] = \begin{cases} 0, & t \neq t' \\ 1, & t = t' \end{cases}$$

---

[3] This result is valid even for spaces of uncountable dimension. The delta function is really the integral operator representation of the identity operator.

This same transformation is applied to the deterministic functions:

$$F(t) = \int T(t,x) \, f(x) \, dx.$$

The functions $F(t)$ form a new function space, $\mathcal{H}'$. Since the noise in $\mathcal{H}'$ is white, the inner product is given by Eq. 2.1:

$$(F,G)_{\mathcal{H}'} = (F,G)_{\mathcal{W}} = \int E \, [G^*(t) \, F(t)] \, dt.$$

The inner product in $\mathcal{H}$ is now defined to make it consistent with that in $\mathcal{H}'$:

$$(2.7) \qquad (f,g)_{\mathcal{H}} = (F,G)_{\mathcal{H}'}$$

Writing T^{ad} for the adjoint of T:

$$(f,g)_{\mathcal{H}} = (F,G)_{\mathcal{W}} = (Tf,Tg)_{\mathcal{W}} = (f,T^{ad}Tg)_{\mathcal{W}}$$

But the associated operator notation gives $(f,g)_{\mathcal{H}} = (f,Hg)_{\mathcal{W}}$, so that the desired inner product is obtained by setting:

$$(2.8) \qquad H = T^{ad} \, T, \quad \text{or} \quad H(x,x') = \int T^*(t,x) \, T(t,x') \, dt.$$

Since the noise is assumed to be strictly positive definite, $R(x,x)$ is bounded away from zero and the operator T is bounded and has an inverse:

$$n(x) = \int T^{-1}(x,t) \, N(t) \, dt, \quad n(x') = \int T^{-1}(x',t) \, N(t) \, dt,$$

so that

$$(2.9) \qquad R(x,x') = E \, [n(x) \, n^*(x')] = \int T^{-1}(x,t) \, T^{*-1}(x',t) \, dt.$$

Interpreting Eq. 2.9 as an equation in linear operators, $R = T^{-1}(T^{ad})^{-1}$.

Thus

(2.10) $R^{-1} = T^{ad} T = H.$ Q.E.D.

Finding the transformation to represent the noise in terms of a white noise process is a complex problem. Cramér (1951) derives a condition very similar to Eq. 2.9 using the assumption that $R(x,x')$ be of bounded variation in every finite domain. It should be noted that a white noise representation is always possible if the data is defined on a countable set of values (Loève, 1960).

The assumption of strict positive definiteness on R guarantees the existence of R^{-1} as an operator, hence as an inner product measure (Friedrichs, 1960; Thm. 31.1). For the general function space, the inner product of Eq. 2.5 using $H = R^{-1}$ can be simply posited, the formulation being justified by the actual construction of R^{-1} on any subspace of countable dimension.

The function space $\mathcal{H}$ with inner product measure given by R^{-1} is very closely related to the idea of the reproducing kernel Hilbert space introduced by Aronszajn (1950) and applied to signal processing by Parzen (1961, 1962; cf. Capon, 1965). The noise process forms a Hilbert space (Karhunen, 1947), η, where:

$$(f,g)_\eta = (f,Rg)_{\mathcal{W}} = \iint g^*(x)\, R(x,x')\, f(x)\, dx\, dx'.$$

An operator $B = B(x,x')$ is a reproducing kernel if, when considered as a function of its first variable, it satisfies:

$$(f(x),\, B(x,x'))_\eta = f(x'); \quad \text{or} \quad (f, B\cdot 1)_\eta = f.$$

Using the associated linear operator, R:

$$(f,B\cdot 1)_{\eta} = (f,RB\cdot 1)_{W} = f.$$

Thus RB = I, the identity operator, and the reproducing kernel is R^{-1}. Conversely, the space for which R is the reproducing kernel must have R^{-1} for the inner product operator. The space $\mathcal{H}$ that we shall use is this space with reproducing kernel R. Using this viewpoint, the process of changing inner products replaces the process of finding a transformation between two spaces.

Still a different approach to finding the inner product on $\mathcal{H}$ is the "classical" technique: finding the spectral representation of the operator, R. Friedrichs (1960) shows that every bounded Hermitian operator, even on a space of uncountable dimension, has a spectral representation. Since R is positive definite, all its eigenvalues are positive. The spectral representation is worked out in detail when X is a finite line segment by Rosenblatt (1962, Ch. VIII). In general it involves finding the eigenvalues and the eigenfunctions of the covariance function; that is, solving:

$$\lambda\phi(x) = \int_{X} R(x,x')\ \phi(x')\ dx'.$$

The expression of the noise as a white noise process can be found from the eigenfunction expansion (Loève, 1960). The transformation T corresponds to the operator whose eigenvalues are the reciprocals of the square roots of the eigenvalues of R. The existence of a reproducing kernel means that the eigenfunctions span the whole space $\mathcal{H}$. In finite dimensional terms, it means the rank of the

covariance matrix R is equal to the dimensionality of the function space. In this case, R is non-singular and has an inverse, R^{-1}.

It has thus been shown that R^{-1} exists and that the inner product using R^{-1} as the associated operator is the appropriate one for $\mathcal{H}$. The transformation T need not actually be found. Yet the idea of the space $\mathcal{H}'$ in which the inner product is the simple one of Eq. 2.1, in which the noise is white, and in which the coordinate system is orthonormal, has sufficient intuitive appeal that it will be retained in the following discussions.

The Signal Subspace

The ability to process the data is based on the fact that the symmetric functions lie in a subspace of the space of all functions. Furthermore, a coordinate system can be given to $\mathcal{H}$ in which the symmetric functions are isolated from the non-symmetric ones. Naturally, the distance notion in $\mathcal{H}$ must be preserved in this new coordinate system.

The symmetric functions clearly form a subspace since linear combinations and limits of sequences of symmetric functions are also symmetric. More precisely, the symmetric function subspace, $\mathcal{S}$, is the function space of functions defined on Y, the unit cell space. The inner product on $\mathcal{S}$ is defined as:

$$(2.11) \qquad (s_1,s_2)_{\mathcal{S}} = \iint_Y s_2{}^*(y)\, R_s{}^{-1}(y,y')\, s_1(y')\, dy\, dy'$$

$$\text{where } R_s^{-1}(y,y') = \Sigma_i\, \Sigma_j\, R^{-1}(x^{a_i}, x'^{a_j}),$$

where y is the equivalence class of x, y' of x'. The summations in the definition of R_s^{-1} are taken over all the automorphisms in the

symmetry group.

> **Definition 2.11.** The symmetric function subspace, $\mathcal{S}$, is the function space of functions defined on Y, with inner product given by Eq. 2.11.

The symmetric function space is defined as a space separate from $\mathcal{H}$.

A symmetric function can be considered to be either a function on Y or a special type of function on X. In geometric terms, the space $\mathcal{S}$ may be considered to be a separate space or a subspace of $\mathcal{H}$: the space $\mathcal{S}$ may be embedded in $\mathcal{H}$. The inner product for $\mathcal{S}$ has been defined to make it consistent with this embedding:

$$(s_1(x), s_2(x))_{\mathcal{H}} = (s_1(y), s_2(y))_{\mathcal{S}} \text{ , or}$$

$$\iint_X s_2^*(x) \ R^{-1}(x,x') \ s_1(x') \ dx \ dx' = \iint_Y s_2^*(y) \ R_s^{-1}(y,y') \ s_1(y') \ dy \ dy' .$$

This equality is easily verified, since the integration measure, as well as s_1 and s_2, are invariant under the automorphisms of the symmetry group.

A method is now needed to determine which points of $\mathcal{H}$ are in the subspace $\mathcal{S}$. This determination is made by extending the notion of an automorphism of X to a transformation of $\mathcal{H}$. It is recalled that, given a function $f(x)$, the automorphism, a, produced a new function $f'(x) = f(x^a)$. The automorphism, a, can thus be considered to be a linear operator which transforms the point f in $\mathcal{H}$ to the new point $f' = a \ f$. This operator is an automorphism of $\mathcal{H}$ in the strict mathematical sense, being a 1:1 continuous mapping of $\mathcal{H}$ onto itself which preserves all its mathematical structure. A function $s(x)$ is symmetric if and only if it is invariant under all the automorphisms of the

symmetry group. Thus a point s in $\mathcal{H}$ is in the symmetric function

subspace if and only if it is invariant under all the operators, a, in

the symmetry group.

> The symmetric function subspace, $\mathcal{S}$, considered as a sub-
> space of $\mathcal{H}$, is the set of all points of $\mathcal{H}$ which are
> singular points under all the operators, a, in the sym-
> metry group.

In order to simplify the processing procedure, it is convenient

to have a coordinatization of the function space in which the symmetric

functions are expressed very simply. As an example, consider a two-

dimensional function space[4] -- the x-y plane. A function in this space

will be described by its coordinates[5] (α, β). Suppose the symmetric

function subspace is the set of all points with equal coordinates,

(α, α). The subspace is thus the line described by the equation $y = x$.

If the plane is rotated through $45°$, a new coordinate system $x' - y'$

is found, where the x' axis coincides with the symmetric function

subspace. The symmetric function now has coordinates $(\alpha', 0)$. In the

new coordinate system, a point is in the symmetric function subspace

if and only if its second coordinate is zero.

Returning to the general case, a new coordinate system is to

be found in which the function $f(x)$ becomes $\underline{F(z)}$. The set of points,

z, forms the new basis for this space $\mathcal{H}''$ of the functions $\underline{F}$. The idea

is to choose the coordinates z in such a way that $\underline{F(z)}$ is in the

signal subspace if and only if $\underline{F(z)} = 0$ for certain values of z.

That is, there is a set, Z_S, such that $\underline{S(z)}$ is a symmetric function

if and only if $\underline{S(z)} = 0$ for $z \notin Z_S$.

---

[4] This space will be considered in detail in Chapter III.
[5] This notation should not be confused with the inner product.

Such a coordinate system does exist since the points in Y

form a basis for the subspace $\mathcal{S}$. The orthogonal complement of $\mathcal{S}$ in

$\mathcal{H}$ is the space $\mathcal{T}$ of all points t that satisfy $(s,t)_{\mathcal{H}} = 0$ for all

s in $\mathcal{S}$. A basis for $\mathcal{T}$ combined with the basis for $\mathcal{S}$ forms a basis

for $\mathcal{H}$ with the desired property.

The basis can be constructed by a spectral representation

technique. Each transformation, a, is a permutation of the points of

X, the basis elements of $\mathcal{H}$. Therefore they are unitary transformations,

with all their eigenvalues having absolute value one. The invariant

functions for each transformation are those eigenfunctions with eigen-

value one. The basis for the symmetric function subspace is taken from

those functions that have eigenvalue one for all the transformations.

The basis for the orthogonal complement is taken from those functions

that do not have eigenvalue one for any transformation except the

identity.

Summary of Function Spaces

Several types of spaces have so far been introduced. It will

be useful to survey the spaces and their relations in order to provide

a fuller understanding and motivation for the processing procedures.

The data is defined on the data space X. The data is composed

of a structured signal and a sample function of a noise process. The

signal is a function defined on the unit cell space Y, which may be

considered to be a representative structural unit of X. The space X

is composed of equivalent building blocks, each identical to Y (Fig.

2.5a), though these blocks may themselves have some interrelations.

Finally, equations are derived which involve only the automorphisms of the symmetry group and the covariance function R, and which have as their unique solution the optimum processing operation. These equations do not involve either the intermediate transformations or the inverse covariance function R^{-1}.

The space $\mathcal{H}$. -- A point in $\mathcal{H}$, written f, represents the function f(x). The covariance function is $R(x,x') = E[n(x)\ n^*(x')]$. The inner product operator is the inverse of the covariance operator:

$$(f,g)_{\mathcal{H}} = (f,R^{-1}g)_{\mathcal{W}} = \iint E[g^*(x)\ R^{-1}(x,x')\ f(x')]\ dx\ dx'$$

Symmetric functions satisfy:

$$s = a\ s,\quad \text{or}\quad s(x) = s(x^a),\quad \text{for all}\quad a\quad \text{in}\quad G.$$

The space $\mathcal{H}'$. -- A point in $\mathcal{H}'$, written F, represents the function F(t):

$$F = T\ f,\quad \text{or}\quad F(t) = \int T(t,x)\ f(x)\ dx.$$

The covariance in $\mathcal{H}'$ is $R'(t,t') = E[N(t)\ N^*(t')]$. In operator terms, the covariance is transformed as:

$$R' = T\ R\ T^{ad}.$$

T is chosen to make R' = I, giving the inner product operator $R'^{-1} = I$:

$$(F,G)_{\mathcal{H}'} = \int E[G^*(t)\ F(t)]\ dt.$$

Symmetric functions satisfy $S = T\ a\ T^{-1}\ S$:

$$S = A \ S, \text{ where } A = T \ a \ T^{-1} \quad \text{for all } A \text{ in } G$$

$$\text{or } S(t) = \int A(t,t') \ S(t') \ dt', \text{ where } A(t,t') = \int T(t,x) \ T^{-1}(x^a,t') \ dx.$$

The space $\mathcal{H}''$. -- A point in $\mathcal{H}''$, written $\underline{F}$, represents the function $\underline{F(z)}$:

$$\underline{F} = U \ F, \quad \text{or } \underline{F(z)} = \int U(z,t) \ F(t) \ dt.$$

U is a unitary transformation: $U^{ad} = U^{-1}$. The covariance in $\mathcal{H}''$ is R'':

$$R'' = U \ R' \ U^{ad} = I$$

since $R' = I$. The inner product operator is $R''^{-1} = I$:

$$(\underline{F},\underline{G})_{\mathcal{H}''} = \int E[\underline{G^*(z)} \ \underline{F(z)}] \ dz$$

Symmetric functions satisfy $\underline{S} = U \ A \ U^{-1} \ \underline{S}$:

$$\underline{S} = \underline{A} \ \underline{S}, \text{ where } \underline{A} = U \ A \ U^{-1} = U \ T \ a \ T^{-1} \ U^{-1}, \text{ for all } \underline{A} \text{ in } G.$$

U is chosen so that the $\underline{A}$ operator takes the form:

$$\underline{A} \ \underline{F} = \underline{A(z)} \ \underline{F(z)}.$$

Since the operator, a, is unitary, $|\underline{A(z)}| = 1$. Let Z_i be the set of z such that $\underline{A_i(z)}$ is actually 1:

$$Z_i = \left\{ z \,|\, \underline{A_i(z)} = 1 \right\}$$

Define Z_s as the set of z such that all the $\underline{A_i}$ are 1:

$$Z_s = \bigcap_i Z_i \ ,$$

taking the intersection over all the elements of the symmetry group.

Then the symmetric functions are exactly those functions that satisfy:

$$\underline{S(z)} = 0 \quad \text{for} \quad z \notin Z_s.$$

D. The Optimization

Data Processing

There is now enough mathematical structure to consider the data processing operation itself. The manipulations of the data function that are considered here are restricted to linear, continuous operations. If the data is $d(x)$, the signal estimate is the result of applying a linear operation, L, on the data: $L[d(x)]$. The linearity and continuity conditions provide that:

$$L\left[\alpha d_1 + \beta d_2\right] = \alpha L[d_1] + \beta L[d_2], \text{ where } \alpha \text{ and } \beta \text{ are complex numbers}$$

$$L\left[\int h(x,x') \, d(x') \, dx'\right] = \int h(x,x') \, L[d(x')] \, dx'.$$

Since the transformations between the various function spaces are linear, the processing manipulation can be expressed equally well in all of them: L' operates in $\mathcal{H}'$ on $D(t)$, L" operates in $\mathcal{H}"$ on $\underline{D(z)}$. Specifying the manipulation in any one space completely determines it in all of them. It must be remembered that, in the most general processing problem, the transformation T may not be available. The processing solution must be expressed solely in terms of the $\mathcal{H}$ space, and any results derived using concepts from the other spaces must also be valid completely within $\mathcal{H}$.

The error function between the signal, s, and the signal

estimate, L[d], is s-L[d]. The error magnitude is the norm of the

error function, $\|s-L[d]\|_{\mathcal{H}}$. The optimal processing operation L_{opt}

must minimize the magnitude of this error over the class of all linear

operations.

> Definition 2.12. The optimum data processing manipula-
> tion, L_{opt}, is the linear, continuous operator that
> satisfies: $\|s-L_{opt}[d]\|_{\mathcal{H}}$ = inf $\|s-L[d]\|_{\mathcal{H}}$, taking the
> infimum over all linear, continuous operations, L.
> Write $\hat{s} = L_{opt}[d]$, and $\Delta = \|s-\hat{s}\|_{\mathcal{H}}$.

It is apparent that the signal estimate $\hat{s}$ must be in the

signal subspace. For suppose the contrary. Then $\hat{s}$ can be separated

into two components, one in $\mathcal{S}$ and one in $\mathcal{T}$, the orthogonal complement

of $\mathcal{S}$: $\hat{s} = s_1 + t$, where $s_1 \in \mathcal{S}$ and $(t,f) = 0$ for all f in $\mathcal{S}$. Then:

$$\Delta^2 = \|s-\hat{s}\|^2 = \|s-s_1 - t\|^2$$

$$= \|s-s_1\|^2 + \|t\|^2$$

since $(s-s_1,t) = 0$. The error may be decreased by setting $t = 0$,

contrary to the assumption that $\hat{s}$ is the optimal estimate.

This condition on the optimal processing operator is called

the _symmetry_ condition, since $\hat{s}$ must be a symmetric function.

> Processing condition (1): The optimal processing oper-
> ator L_{opt} is symmetric: $L_{opt} [f] \in \mathcal{S}$ for all f in $\mathcal{H}$.

This condition can also be expressed in the other spaces: $L'_{opt}[F] \in \mathcal{S}'$,

$L''_{opt} [\underline{F}] \in \mathcal{S}''$.

The only prior knowledge that we have concerning the signal

is that it is located in the symmetric function subspace. The optimization

must therefore proceed for all possible s in $\mathcal{S}$ and the optimum operator must be independent of the signal. As a consequence, the signal subspace points are all invariant under the processing operator: $L_{opt}[s] = s$, for all s in $\mathcal{S}$. To prove this statement, first note that, since L_{opt} is linear, $L_{opt}[d] = L_{opt}[s] + L_{opt}[n]$. Thus:

$$\Delta^2 = \|s - \hat{s}\|^2 = \|s - L_{opt}[s] - L_{opt}[n]\|^2$$

(2.12) $$\Delta^2 = \|s - L_{opt}[s]\|^2 + \|L_{opt}[n]\|^2 + (s - L_{opt}[s], L_{opt}[n])$$

$$+ (L_{opt}[n], s - L_{opt}[s]).$$

Now consider the first inner product term.

$$(s - L_{opt}[s], L_{opt}[n]) = \iint E\left[L_{opt}^*[n(x)] \; R^{-1}(x,x') \left\{ s(x') - L_{opt}[s(x')] \right\} \right] dx \; dx$$

The signal is a deterministic function, as is $L_{opt}[s]$. Since the noise is assumed uncorrelated with the signal, it is also uncorrelated with any linear transformation of the signal. Thus the expected value operation applies only to the noise term. But the noise has been assumed to have zero mean. Thus the whole inner product is zero. The other inner product is the complex conjugate of this one, so it, too, is zero. Equation 2.12 simplifies to:

(2.13) $$\Delta^2 = \|s - L_{opt}[s]\|^2 + \|L_{opt}[n]\|^2.$$

To make the minimization of error independent of the signal, set:

(2.14) $$L_{opt}[s] = s, \text{ for all } s \text{ in } \mathcal{S}, \qquad Q.E.D.$$

Since the result of processing any data is a symmetric function, no further improvement can be obtained by passing the data through the

processing scheme twice. The second processing manipulation will give

the same result as the first: $L_{opt}^2[f] = L_{opt}\left[L_{opt}[f]\right] = L_{opt}[f]$, or

$L_{opt}^2 = L_{opt}$. Any operator that satisfies $L^2 = L$ is called <u>idempotent</u>,

and for every idempotent operator there is an invariant subspace. To

show that the invariant subspace for the optimal processing operator is

the symmetric function subspace, the condition given by Eq. 2.14 will be

called <u>idempotency on</u> $\mathscr{S}$.

> Processing condition (2): The optimal processing opera-
> tor L_{opt} is idempotent on $\mathscr{S}$: $L_{opt}[s] = s$, for all s
> in $\mathscr{S}$.

Similar idempotency conditions apply to the other function spaces:

$L'_{opt}[S] = S$, for all S in $\mathscr{S}$ '; $L''_{opt}[\underline{S}] = \underline{S}$, for all $\underline{s}$ in $\mathscr{S}$ ".

Two consequences of idempotency are of interest. First,

Eq. 2.13 reduces to:

(2.15) $\qquad \Delta = \|L_{opt}[n]\|_{\mathscr{H}}$

The processing error is simply the mean amplitude of the result of

processing pure noise. The final processing condition will be derived

from this result.

As a second consequence, the signal estimate is unbiased:

$$E[\hat{s}] = E\left[L_{opt}[s] + L_{opt}[n]\right] = L_{opt}[s] + E\left[L_{opt}[n]\right]$$

(2.16) $\qquad E[\hat{s}] = s.$

The condition of idempotency requires L_{opt} to have only

1 and 0 as eigenvalues. Idempotency <u>on S</u> requires the eigenfunctions

of eigenvalue 1 to be the symmetric functions. The eigenfunctions of

eigenvalue 0 form a subspace, $\mathcal{L}$, which is annihilated by L_{opt}:

> Definition 2.13. The space $\mathcal{L}$ is the subspace of $\mathcal{H}$ which is annihilated by L_{opt}: $L_{opt}[f] = 0$, for all f in $\mathcal{L}$. The orthogonal complement of $\mathcal{L}$ in $\mathcal{H}$ is denoted $\mathcal{M}$.

The subspaces $\mathcal{L}'$ and $\mathcal{L}''$, which are annihilated by L'_{opt} and L''_{opt}, respectively, are the analogs of $\mathcal{L}$ in $\mathcal{H}'$ and $\mathcal{H}''$.

Two facts about $\mathcal{L}$ will be of use later: (1) the spaces $\mathcal{L}$ and $\mathcal{M}$ span $\mathcal{H}$, and (2), $\mathcal{L}$ and $\mathcal{M}$ are disjoint. The first follows since any point f in $\mathcal{H}$ can be written as a sum of a point, $L_{opt}[f]$, in $\mathcal{M}$ and a point, $f - L_{opt}[f]$, in $\mathcal{L}$. The second follows since any point f simultaneously in $\mathcal{M}$ and $\mathcal{L}$ must satisfy $L_{opt}[f] = f$ and $L_{opt}[f] = 0$, hence f = 0.

Noise Components in Subspaces

The first two processing conditions are independent of the noise process. The noise covariance enters by means of Eq. 2.15, which states that the processing error amplitude is the mean processed noise amplitude. Since the norm of any function is preserved by the transformations:

$$(2.15') \qquad \Delta = \|L_{opt}[n]\|_{\mathcal{H}} = \|L'_{opt}[N]\|_{\mathcal{H}'} = \|L''_{opt}[\underline{N}]\|_{\mathcal{H}''}$$

The optimal processing operation has been shown to reduce the whole function space to a subspace. Some results concerning noise amplitudes in subspaces will be derived in this section. The main idea to be developed is that the mean noise amplitude in any subspace depends only on the dimension of the subspace in comparison with the

dimension of the whole function space.

The mean amplitude of the pure noise is one, in any function space:

$$(2.17) \qquad \|n\|_{\mathcal{H}} = \|N\|_{\mathcal{H}'} = \|\underline{N}\|_{\mathcal{H}''} = 1.$$

This need only be demonstrated in one of the spaces, since the norm is preserved by the transformations:

$$\|n\|_{\mathcal{H}}^2 = \iint_X E[n^*(x)\ R^{-1}(x,x')\ n(x')]\ dx\ dx'$$

$$= \iint_X R(x',x)\ R^{-1}(x,x')\ dx\ dx'$$

$$= \int_X 1\ dx = 1$$

since the measure of the data space is 1. If Z is the "data space" for the $\mathcal{H}''$ function space, that is the set of points z,

$$\|\underline{N}\|_{\mathcal{H}''}^2 = \int_Z E[\underline{N}^*(z)\ \underline{N}(z)]\ dz$$

$$= \int_Z R''(z,z)\ dz = \int_Z 1\ dz = 1$$

since the covariance function in $\mathcal{H}''$ is the identity. The measure of Z is 1.

Now suppose that $\mathcal{H}''$ has finite dimension, k. Then Z is a finite set and the integrals become finite summations:

$$\|\underline{N}\|_{\mathcal{H}''}^2 = (1/k)\ \Sigma_i\ R(z_i,z_i) = 1.$$

The weighting factor $(1/k)$ normalizes the measure of Z to one.

Now consider the subspace of dimension 3, formed by z_2, z_5, and z_7, say. The noise on this subspace will have three components:

$$\underline{N}_3 = (\underline{N}(z_2), \ \underline{N}(z_5), \ \underline{N}(z_7))$$

and the mean square noise amplitude will be:

$$\|\underline{N}_3\|^2_{\mathcal{H}''} = (1/k) \ [R''(z_2, z_2) + R''(z_5, z_5) + R''(z_7, z_7)]$$

$$= 3/k$$

since the covariance $R''(z, z) = 1$ for all z. It is clear that the mean square noise amplitude on <u>any</u> subspace of dimension 3 will also be $3/k$.

The noise on any subspace of arbitrary dimension will be composed of several components, one for each dimension. The mean square amplitude of each of these components is equal to $1/k$, since the noise is uniform in every direction. Taking these components to be mutually orthogonal, the Pythagorean theorem says that the total mean square noise amplitude is the sum of the mean square amplitudes of the components, thus for any subspace of dimension j, $\|\underline{N}_j\|^2_{\mathcal{H}''} = j/k$.

For spaces of infinite dimension, a simple counting procedure will not work, and this result must be modified. Still referring to the space $\mathcal{H}''$, consider the noise $\underline{N}_{Z_1}$ in the subspace of functions which vanish outside Z_1: $\underline{F}(z) = 0$ for $z \notin Z_1$. The set Z_1, which we have been calling a basis for the subspace, is the <u>support</u> for the subspace.

$$\|\underline{N}_{Z_1}\|^2_{\mathcal{H}''} = \int_{Z_1} R''(z, z) \ dz = \int_{Z_1} 1 \ dz = \mu(Z_1)$$

where $\mu(Z_1)$ is the measure of Z_1. The mean square noise is given by measuring the support of the subspace, rather than by counting basis elements. It should be recalled that the measure of the support of the whole function space is one. The concept of the measure of the

support of a subspace replaces the idea of the number of basis elements

for infinite dimensional spaces. Thus, if two subspaces are supported

by sets of equal measure, the mean noise amplitude in each subspace will

be equal.

The key to finding the mean square noise amplitude in a sub-

space is finding an orthonormal basis for the subspace, then counting

the basis or measuring its support. Orthonormality is required in

order for each component to contribute equally and independently to the

total.

The inner product function in $\mathcal{H}$ has been defined to make the

noise in orthogonal directions uncorrelated and of equal amplitude.

By fiat, the notion of dimension in $\mathcal{H}$ has been equated with mean square

noise amplitude. Two subspaces are of equal dimension if they have equal

noise amplitude. If an orthonormal basis for a subspace is found, so

that the support of the functions in it can be expressed, the dimension-

ality of the subspace equals the measure of the support.

For our own purposes, we need not actually count or measure

the dimension of any subspace in $\mathcal{H}$. We need only compare certain sub-

spaces to verify that they are of equal dimension and thus have equal

mean noise. As a convenient notation, write $\mu(\mathcal{S})$ for this "dimension"

or measure of $\mathcal{S}$.

The first result we need follows from the additivity of measure

or dimension:

Lemma 2.1. If $\mathcal{S}_1$ and $\mathcal{S}_2$ are disjoint subspaces
that span $\mathcal{H}$, then $\mu(\mathcal{S}_1) + \mu(\mathcal{S}_2) = \mu(\mathcal{H}) = 1$.

The second result follows from the generalized Pythagorean

Theorem: orthogonal components of noise combine by adding the mean

square amplitudes.

Lemma 2.2. Suppose $\mathcal{H}$ has two orthogonal decompositions: $\mathcal{H} = \mathcal{S}_1 + \mathcal{T}_1 = \mathcal{S}_2 + \mathcal{T}_2$, where $\mathcal{S}_1 \perp \mathcal{T}_1$ and $\mathcal{S}_2 \perp \mathcal{T}_2$. Then write $n = n_{\mathcal{S}_1} + n_{\mathcal{T}_1} = n_{\mathcal{S}_2} + n_{\mathcal{T}_2}$, where the subscript indicates to which space that noise component belongs. Suppose further that $\mu(\mathcal{S}_1) = \mu(\mathcal{S}_2)$. Then $\|n_{\mathcal{S}_1}\|_{\mathcal{H}} = \|n_{\mathcal{S}_2}\|_{\mathcal{H}}$.

Minimizing the Error

The space $\mathcal{H}$ can be decomposed into disjoint subspaces in

three ways. We have already used the decomposition into the symmetric

function subspace $\mathcal{S}$ and its orthogonal complement $\mathcal{T}$. The discussion

following Definition 2.13 gives a decomposition into the spaces $\mathcal{S}$ and

$\mathcal{L}$. Finally, $\mathcal{H}$ can also be decomposed into $\mathcal{L}$ and its orthogonal

complement $\mathcal{M}$.

These three decompositions are illustrated in Fig. 2.6, using

the $\mathcal{H}$" space in which orthogonal axes can be drawn geometrically per-

pendicular. The algebraic decompositions can be written:

$$(2.18) \quad (1) \quad n = n_{\mathcal{S}_o} + n_{\mathcal{T}} \qquad \underline{N} = \underline{N}_{\mathcal{S}_o\text{"}} + \underline{N}_{\mathcal{T}\text{"}}$$

$$(2) \quad n = n_{\mathcal{S}_L} + n_{\mathcal{L}_L} \qquad \underline{N} = \underline{N}_{\mathcal{S}_L\text{"}} + \underline{N}_{\mathcal{L}_L\text{"}}$$

$$(3) \quad n = n_{\mathcal{M}} + n_{\mathcal{L}_o} \qquad \underline{N} = \underline{N}_{\mathcal{M}\text{"}} + \underline{N}_{\mathcal{L}_o\text{"}}$$

The subscript indicates to which subspace that component of the noise

belongs. The subscript $\mathcal{S}_L$ indicates that the component is in the space

$\mathcal{S}$, and that it is obtained using the L_{opt} operator; the subscript $\mathcal{S}_o$
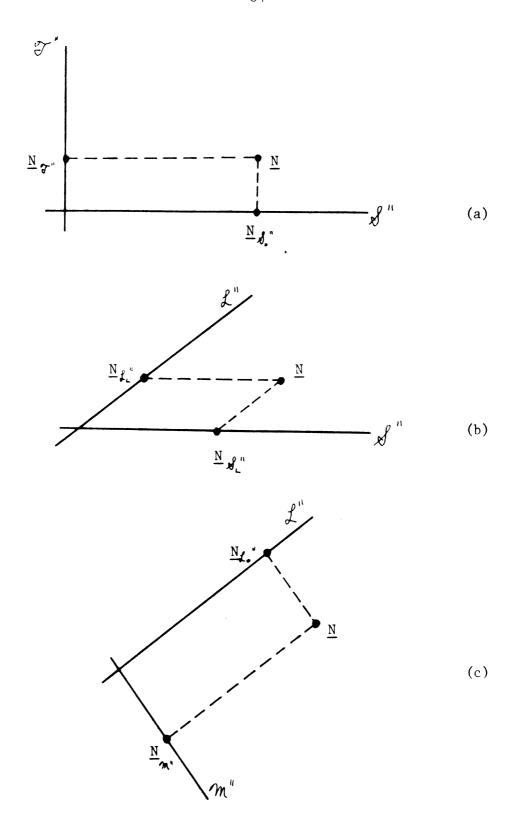
Figure 2.6. Three Ways of Representing the Noise in Terms of Components.

indicates that the decomposition is orthogonal.

Apply Lemma 2.1:

$$(2.19) \qquad \mu(\mathcal{S}) + \mu(\mathcal{T}) = \mu(\mathcal{S}) + \mu(\mathcal{L}) = \mu(\mathcal{M}) + \mu(\mathcal{L}) = 1,$$

$$\mu(\mathcal{S}'') + \mu(\mathcal{T}'') = \mu(\mathcal{S}'') + \mu(\mathcal{L}'') = \mu(\mathcal{M}'') + \mu(\mathcal{L}'') = 1.$$

$$(2.20) \qquad \mu(\mathcal{L}) = \mu(\mathcal{T}); \qquad \mu(\mathcal{L}'') = \mu(\mathcal{T}'')$$

$$\mu(\mathcal{S}) = \mu(\mathcal{M}); \qquad \mu(\mathcal{S}'') = \mu(\mathcal{M}'').$$

It then follows from Lemma 2.2:

$$(2.21) \qquad \left\| n_{\mathcal{S}_o} \right\|_{\mathcal{H}} = \left\| n_{\mathcal{M}} \right\|_{\mathcal{H}}; \quad \left\| \underline{N}_{\mathcal{S}_o''} \right\|_{\mathcal{H}''} = \left\| \underline{N}_{\mathcal{M}''} \right\|_{\mathcal{H}''}.$$

Furthermore, applying the operator L_{opt} to the noise as expressed in the second decomposition of Eq. 2.18:

$$L_{opt}[n] = L_{opt}[n_{\mathcal{S}_L}] + L_{opt}[n_{\mathcal{L}}]$$

$$= n_{\mathcal{S}_L}$$

since L annihilates $\mathcal{L}$ and leaves $\mathcal{S}$ invariant. Thus:

$$(2.22) \qquad \Delta = \left\| n_{\mathcal{S}_L} \right\|_{\mathcal{H}} = \left\| \underline{N}_{\mathcal{S}_L''} \right\|_{\mathcal{H}''}.$$

Now referring to Fig. 2.7, it is obvious that

$$\left\| \underline{N}_{\mathcal{M}''} \right\|_{\mathcal{H}''} \leq \left\| \underline{N}_{\mathcal{S}_L''} \right\|_{\mathcal{H}''}$$

since a leg of a right triangle cannot be longer than the hypotenuse. Applying Eq. 2.21 and 2.22:

$$\left\| \underline{N}_{\mathcal{S}_o''} \right\|_{\mathcal{H}''} \leq \Delta = \left\| \underline{N}_{\mathcal{S}_L''} \right\|_{\mathcal{H}''}$$
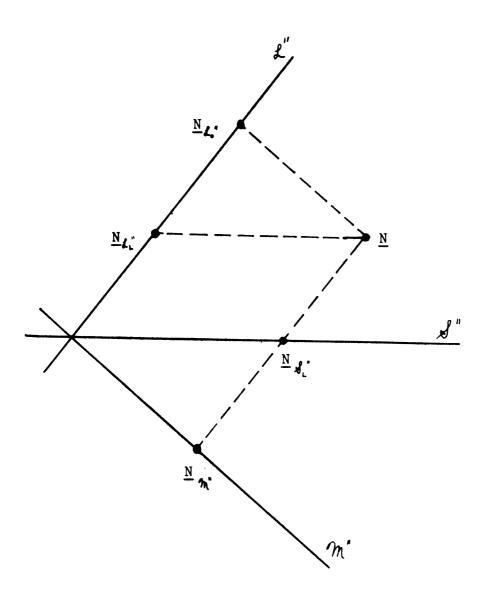
Figure 2.7. Relationship of the Noise Components in Two Decompositions.

But Δ is required to be the smallest error, so the equality must hold and $\|\underline{N}_{\mathcal{L}_L}"\| = \|\underline{N}_{\mathcal{L}_o}"\|$. This can be accomplished by setting $\mathcal{L}" = \mathcal{T}"$. The optimal processing operator must project the data point orthogonally to the symmetric function subspace.

Essentially the same argument can be carried out algebraically in the $\mathcal{H}$ space by noting:

$$n_{\mathcal{M}} = n - n_{\mathcal{L}_o} \qquad n_{\mathcal{L}_L} = n - n_{\mathcal{L}_L}$$

from Eq. 2.18. Since $(n-n_{\mathcal{L}_o}) \in \mathcal{M}$ and $(n_{\mathcal{L}_o} - n_{\mathcal{L}_L}) \in \mathcal{L}$,

$$(n-n_{\mathcal{L}_o}, n_{\mathcal{L}_o}-n_{\mathcal{L}_L})_{\mathcal{H}} = 0.$$

Thus
$$\Delta^2 = \|n_{\mathcal{L}_L}\|^2_{\mathcal{H}} = \|n - n_{\mathcal{L}_L}\|^2_{\mathcal{H}}$$

$$= \|(n-n_{\mathcal{L}_o})-(n_{\mathcal{L}_o}-n_{\mathcal{L}_L})\|^2_{\mathcal{H}}$$

$$= \|n-n_{\mathcal{L}_o}\|^2_{\mathcal{H}} + \|n_{\mathcal{L}_o}-n_{\mathcal{L}_L}\|^2_{\mathcal{H}}$$

$$= \|n_{\mathcal{M}}\|^2_{\mathcal{H}} + \|n_{\mathcal{L}_o}-n_{\mathcal{L}_L}\|^2_{\mathcal{H}}$$

Since $\|n_{\mathcal{L}_o}-n_{\mathcal{L}_L}\| \geq 0$ and, from Eq. 2.21, $\|n_{\mathcal{M}}\| = \|n_{\mathcal{L}_o}\|$

$$\|n_{\mathcal{L}_o}\|_{\mathcal{H}} \leq \Delta = \|n_{\mathcal{L}_L}\|_{\mathcal{H}}$$

Again, Δ is not the minimum error unless $\|n_{\mathcal{L}_L}\| = \|n_{\mathcal{L}_o}\|$, which is accomplished by setting $\mathcal{L} = \mathcal{T}$.

> Processing condition (3): The optimal processing
> operator L_{opt} is obtained only if $\mathcal{L} = \mathcal{T}$.

The Projection Operation

It has been shown in the previous section that the optimal

operator L_{opt} transforms the data orthogonally onto the signal subspace. The best estimate of the signal is that point in $\mathscr{S}$ that lies at the foot of the perpendicular from the data point to $\mathscr{S}$. Of course, in the space $\mathscr{H}$ this "perpendicular" is not geometrically perpendicular, but is orthogonal in the sense of the inner product. Thus the processing operation depends on the noise covariance function. The processing problem has been expressed so that the signal estimate is that point in the symmetric function subspace which lies closest to the actual signal point. It turns out that the signal estimate is also that point in the symmetric function subspace which lies closest to the observed data point.

The optimal processing operator is the operator that gives the $\mathscr{S}$ component in the orthogonal decomposition of the data into the $\mathscr{S}$ and $\mathscr{T}$ subspaces. This operator is called the "projector" or "projection operator" from $\mathscr{H}$ to $\mathscr{S}$. The optimal processing operator shall henceforth be written P (or P' or P") to indicate that it is this projection operator. The existence and uniqueness of this operator is guaranteed by the completeness of the function space (see Murray, 1941).

This projection operator satisfies the conditions of symmetry, since it projects onto the symmetric function subspace, and idempotency, since it is idempotent on its range.

> The optimal processing solution. The optimal processing
> operator is the projection operator, P, which projects
> the function space $\mathscr{H}$ onto the symmetric function sub-
> space $\mathscr{S}$.

This operator may be explicitly constructed if the transformations T and U are found. In the coordinate system of $\mathscr{H}$ ", the

symmetric function subspace has the simple representation:

$$\underline{S} \in \mathscr{J}\text{''} \quad \text{if and only if} \quad \underline{S(z)} = 0 \text{ for } z \notin Z_S.$$

The projection operator in $\mathscr{H}\text{''}$ is found by inspection: it sets to zero all values of the data function $\underline{D(z)}$ which correspond to $z \notin Z_S$. In this coordinate system, $P\text{''}$ has its spectral representation and is represented by a <u>multiplicative</u> operation, not the usual integral:

$$(2.23) \qquad \overset{\wedge}{\underline{S(z)}} = P\text{''}\,[\underline{D(z)}] = \underline{P(z)}\,\underline{D(z)}$$

$$\text{where} \quad \underline{P(z)} = \begin{cases} 0, & z \notin Z_S \\ 1, & z \in Z_S \end{cases}$$

The manipulation may be expressed on the original data function $d(x)$ by transforming through $\mathscr{H}\text{'}$ to $\mathscr{H}\text{''}$, obtaining $\underline{D(z)}$; performing the projection using $P\text{''}$ to get $\overset{\wedge}{\underline{S(z)}}$; then transforming back through $\mathscr{H}\text{'}$ to $\mathscr{H}$, obtaining $\hat{s}(x)$:

$$(2.24) \qquad \hat{s} = P[d], \quad \text{where} \quad P = T^{-1}\,U^{-1}\,P\text{''}\,U\,T;$$

$$\hat{s}(x) = \int P(x,x')\,d(x')\,dx', \quad \text{where}$$

$$P(x,x') = \iiiint T^{-1}(s,t)U^{-1}(t,z)\underline{P(z)}U(z,t')T(t',x') \quad dx'\,dt'\,dz\,dt$$

An alternate procedure uses some properties of the projection operator. The projection operator to $\mathscr{J}$ is idempotent and self-adjoint: $P^2 = P$ and $(f,Pg)_{\mathscr{H}} = (Pf,g)_{\mathscr{H}}$ for all f,g in $\mathscr{H}$. Conversely, any idempotent, self-adjoint operator is a projection. Furthermore, any projection that has values in $\mathscr{J}$ (is symmetric) and holds $\mathscr{J}$ invariant (is idempotent on $\mathscr{J}$) must be the desired optimal operator.

Lemma 2.3. Any self-adjoint, symmetric operator that is idempotent on $\mathcal{J}$ is the optimal processing operator, P, the projection from $\mathcal{H}$ to $\mathcal{J}$.

To prove this statement, let L be a self-adjoint, symmetric operator, idempotent on $\mathcal{J}$. Any point f in $\mathcal{H}$ has the unique decomposition, f = s + t, with s $\in \mathcal{J}$, t $\in \mathcal{T}$. Then for any s' in $\mathcal{J}$:

$$(s',Lt) = (Ls',t) \quad \text{by self-adjointness of L}$$

$$= (s',t) \quad \text{by idempotency on } \mathcal{J}$$

$$= 0 \quad \text{since } s' \in \mathcal{J} \text{ and } t \in \mathcal{T} .$$

But L t $\in \mathcal{J}$ by symmetry of L. Thus Lt = 0 and:

$$L\ f = L\ s + L\ t = L\ s = s.$$

But the projection operator also satisfies P f = s so that P f = L f, for all f in $\mathcal{H}$. Thus L = P, Q.E.D.

The concept of self-adjointness is expressed in terms of the inner product in $\mathcal{H}$, hence depends on the inner product associated operator R^{-1}.

$$(f,Pg)_{\mathcal{H}} = (f,R^{-1}Pg)_{\mathcal{W}}$$

$$= (Pf,g)_{\mathcal{H}} = (Pf,R^{-1}g)_{\mathcal{W}} = (f,P^{ad}R^{-1}g)_{\mathcal{W}}$$

In operator terminology, self-adjointness can be expressed:

$$(2.25) \qquad P^{ad}\ R^{-1} = R^{-1}\ P.$$

Applying the R operator to the left and right of Eq. 2.25 gives:

$$(2.26) \qquad R\ P^{ad} = P\ R.$$

The three conditions of symmetry, idempotency on $\mathcal{J}$, and self-adjointness (as expressed in Eq. 2.26) will uniquely determine the operator P.

> The processing operator P which minimizes the mean square error between signal and signal estimate is given uniquely by the three conditions:
>
> (1) symmetry: P f $\in \mathcal{J}$, for all f in $\mathcal{H}$
> (2) idempotency on $\mathcal{J}$:   P s = s, for all s in $\mathcal{J}$
> (3) self-adjointness: P R = R P^{ad}
>
> If the noise process can be represented by a transformation of a white noise process, the optimal processing operator is given explicitly by Eq. 2.24.

CHAPTER III

FINITE DATA SPACES

A. The Processing Problem

The Data Space and Symmetry

In this chapter, the data is assumed to consist of a finite set of N numbers. The data space is a set of points, x_i, $i = 1,\ldots,N$, with the discrete topology. The i-th data value is the value of the data function evaluated at x_i: $d_i = d(x_i)$. Since all the function spaces and all the processing operations are linear, matrix algebra notation and the ideas of finite dimensional vector spaces are natural. Functions defined on X will be written as column vectors:

$$(3.1) \qquad f = (f_i) = (f(x_i)) = \begin{bmatrix} f(x_i) \\ \vdots \\ f(x_N) \end{bmatrix}$$

The automorphisms of the symmetry are shufflings, or permutations, of the points x_i and can be represented as permutation matrices. Each row and each column of a permutation matrix contains exactly one entry with value 1, all other entries are 0. The permutation matrix A corresponding to the automorphism $a: x_i \rightarrow x_j$ has a 1 in the j-i position.

The automorphism, a, transforms the function $f(x)$ to the new function $f'(x) = f(x^a)$. In matrix notation, the permutation matrix A transforms the vector f to the new vector $f' = A f$, the operation being performed by matrix multiplication. A vector, s, is symmetric if and only if:[1]

---

[1] This concept is meaningful since there always exist symmetric functions, namely those functions all of whose components are equal.

(3.2) $s = A \, s$, for all A in G.

The following notation is adopted for convenience. If x_i is equivalent to x_j under the symmetry group G, write $i \equiv j$. The equivalence class of x_j is written y_j and contains μ_j members. Let ν be the number of points in the unit cell space. Then the total number of points in the X space is $N = \sum\limits_{i=1}^{\nu} \mu_i$.

This type of data space can arise from a situation in which an experiment is replicated μ times, each replication producing ν data values. The signal ν-tuple obtained at each trial should remain the same from trial to trial, only the noise should vary. The unit cell space has ν points and each equivalence class has the same number, μ, of points, there being $\mu\nu$ data points in all.

An example of such a space is the statistical problem of estimating the mean of a multivariate, ν-dimensional probability distribution given μ samples. The first moments, the means, correspond to the unknown signal, while the central second moments, the covariances, are known. Of course when the samples are uncorrelated, the best estimate of the population mean must be the sample mean. The present formulation includes, not only the case of correlated samples, but also the case in which the data are derived by correlated samples from different probability distributions.

Of particular interest, is the case in which all the points are equivalent, the case $\nu = 1$. The data space has μ points and the symmetry group is the "full symmetry group" on μ items, the group of all $\mu \times \mu$ permutation matrices.

The Function Space and the Noise

The function space, $\mathscr{H}$, will be a vector space with as many dimensions as there are points in X. Points in $\mathscr{H}$ are column vectors and will be denoted by lower case letters: f represents the vector (f_i) with components $f_i = f(x_i)$. Linear transformations can be written as finite sums:

$$(3.3) \qquad f'(x_i) = \Sigma_j \; h_{ij} \; f(x_j)$$

Using matrix notation, let $H = (h_{ij})$:

$$(3.4) \qquad f' = H \; f.$$

The prior knowledge of the noise is given by the covariance function $R(x_i, x_j) = E \; [n(x_i) \; n^*(x_j)]$. This, too, can be written in matrix notation, setting $R = (r_{ij})$, where $r_{ij} = R(x_i, x_j)$. If n is the noise sample column vector:[2]

$$(3.5) \qquad R = E[n \; n^*]$$

The noise process is known only by the first moments, which are assumed to be zero, and the second moments, which are given by R. The noise probability distribution thus may be approximated by the "ellipse of concentration" (Cramér, 1946). A probability distribution which is uniform inside this ellipse and zero outside has the same first and second moments as the given noise process. The boundary of the ellipse is the locus of points f that satisfy $f^* \; R^{-1} \; f = c$,

---

[2] In this chapter, the superscript asterisk denotes the operation of conjugate transposition; n* is a row vector whose entries are the complex conjugates of the elements of n.

where R^{-1} is the inverse of the covariance matrix and c is a real number whose value depends only on the number of dimensions in $\mathcal{H}$.

Furthermore, the mean square noise amplitude in the direction u, where u is a vector of unit geometric[3] length, is $E[n_u{}^* n_u] = 1/(u^* R^{-1} u)$.

The desired notion of distance in the $\mathcal{H}$ space is the geometric distance divided by the mean noise amplitude in the direction the distance is measured. The unit of length in any direction is thus set equal to the mean noise amplitude in that direction. Let ℓ be the geometric length of the vector f, so that the unit vector in the f direction is $v = f/\ell$. The desired distance notion is:

$$\|f\|_{\mathcal{H}}^2 = \frac{\ell^2}{(v^* R^{-1} v)^{-1}} = (\ell v)^* R^{-1} (\ell v) = f^* R^{-1} f.$$

The inner product in the $\mathcal{H}$ space is thus defined to be:

$$(3.6) \qquad (f,g)_{\mathcal{H}} = E[g^* R^{-1} f].$$

The Transformations

The optimal processing operator can be derived directly from the three conditions given at the end of the previous chapter. The transformations from one function space to another arise so naturally, though, and are so easily visualized in the finite dimensional case that it is worth while considering this approach first.

In two dimensions, the function space may be illustrated by the plane, as shown in Fig. 3.1a. The coordinate axes for $\mathcal{H}$ are x_1

---

[3] The geometric norm is $|u|^2 = u^* u$, the geometric inner product is $u \cdot v = v^* u$, the vector scalar product.
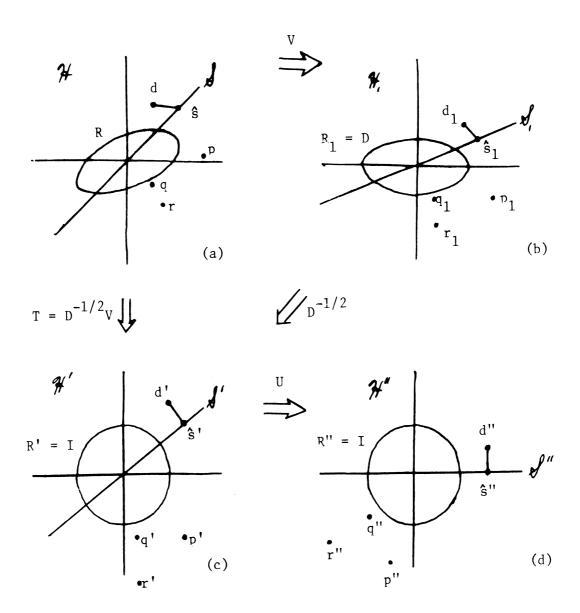
Figure 3.1. Two-dimensional Function Spaces.

and x_2; the data space forms the basis for the function space. The ellipse of concentration represents the noise process: the noise shown is positively correlated between the two data space points. The points p and q are geometrically farther apart than the points q and r. In the $\mathscr{H}$ space distance, though, the two distances should be equal because the trend in the noise makes them equally hard to resolve.

The two points x_1 and x_2 are equivalent under the symmetry group. The permutations are:

$$A_1 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

with the multiplication rules: $A_2^2 = I$, $A_2^{-1} = A_2$. A symmetric function must have equal components, $s(x_1) = s(x_2)$, so the symmetric vector s must lie on the line $x_2 = x_1$, the symmetric function subspace, $\mathscr{S}$.

Returning to the general finite-dimensional case, apply the non-singular matrix T to the function space as a coordinate transformation, obtaining the space $\mathscr{H}'$. The point f in $\mathscr{H}$ transforms to the point f' in $\mathscr{H}'$ by the rule $f' = T f$. The covariance in $\mathscr{H}'$ is given by $R' = E[n'(n')^*] = E[(Tn)(Tn)^*] = E[T(nn^*)T^*] = T R T^*$. The symmetric functions in $\mathscr{H}'$ are transforms of the symmetric functions in $\mathscr{H}$, so that s' is symmetric if and only if $T^{-1} s' = A T^{-1} s'$, or $s' = T A T^{-1} s' = A' s'$ for all A in G, where A' is the transformed permutation matrix. These rules can be summarized:

(3.7) (1) $f' = T f$

  (2) $R' = T R T^*$

  (3) $A' = T A T^{-1}$

Since R is a positive definite matrix, there is a unitary[4] matrix V such that $V R V^* = D$, where D is a real diagonal matrix with positive entries. Thus V can be considered to be a coordinate transformation, taking the $\mathcal{H}$ space to the new function space $\mathcal{H}_1$, which diagonalizes the covariance. A unitary transformation preserves the distance and inner product notions, $(Vf, Vg) = (f, g)$. It is therefore a "rigid body motion", the algebraic representation of a rotation or reflection. Returning to the two-dimensional example, the transformation V rotates the coordinates until the ellipse of concentration has its axes aligned with the coordinate axes (Fig. 3.1b). The covariance matrix corresponding to such an ellipse is diagonal. Since distance is preserved, the relations between p_1, q_1, and r_1 are the same as in $\mathcal{H}$. The symmetric function subspace is no longer the line with slope 1, but must now be found from the transformed permutation matrices.

Since the covariance matrix R is assumed to be strictly positive definite, the matrix D will have no zeroes along the diagonal -- in no direction is the noise amplitude zero. Geometrically, the area inside the ellipse is not zero. By expanding the ellipse along some coordinate axes and shrinking it along others, all the semi-axes can be brought to have unit length; the ellipse becomes a circle, and the noise becomes white (Fig. 3.1c). In the general case, this change in scale factors along the axes is provided by a diagonal transformation matrix, specifically the matrix $D^{-1/2}$. Applying $D^{-1/2}$ as a transformation to $\mathcal{H}_1$ gives the space $\mathcal{H}'$. The covariance in $\mathcal{H}'$ is

---

[4] A matrix V is unitary if $V^* = V^{-1}$.

$R' = D^{-1/2} R_1 (D^{-1/2})^*$. But D is diagonal and positive, so $D^{-1/2}$ is diagonal and real; $(D^{-1/2})^* = D^{-1/2}$. Thus $R' = D^{-1/2} D D^{-1/2} = I$. The noise has been transformed to a white noise process. The inner product in $\mathcal{H}'$ thus has the white noise or geometric form:

(3.8) $(f',g')_{\mathcal{H}'} = E[(g')^* f]$.

In Fig. 3.1c, the distance between p' and q' is now equal to the distance between q' and r'.

The successive action of the transformations V and $D^{-1/2}$ is, in the notation of the previous chapter, the transformation T. In the $\mathcal{H}'$ space:

(3.9) $f' = T f, \quad R' = T R T^* = I, \quad A' = T A T^{-1}$.

Defining the inner product in $\mathcal{H}$ in terms of that in $\mathcal{H}'$:

$$(f,g)_{\mathcal{H}} = (Tf,Tg)_{\mathcal{H}'} = E[(Tg)^*(Tf)] = E[g^*(T^*T)f]$$

$$(f,g)_{\mathcal{H}} = E[g^* H f], \quad \text{where} \quad H = T^* T.$$

Thus the associated linear operator for the inner product is $H = T^* T$. This result is simply a restatement of Eq. 2.8, recognizing that the representation of the adjoint of an operator is the conjugate transpose of a matrix.

From Eq. 3.9, $T R T^* = I$. Therefore, $R = T^{-1}(T^*)^{-1}$ and $R^{-1} = T^* T = H$:

(3.10) $(f,g)_{\mathcal{H}} = E[g^* R^{-1} f]$.

The optimal signal processing operation in $\mathcal{H}'$ is found by

projecting the data point d' orthogonally to the symmetric function

subspace $\mathcal{J}$', as shown in Fig. 3.1c. This projection is most easily

performed by rotating the coordinate axes once more so that the symmetric

function subspace lies oriented with the coordinate axes. In Fig. 3.1d,

the rotation brings $\mathcal{J}$" to lie on the horizontal axis. Rotating the

covariance circle leaves it a circle -- the noise remains white.

In the general case, suppose the unit cell space consists of

ν points. The symmetric function subspace is then ν-dimensional. A

unitary transformation U will rotate the coordinates until $\mathcal{J}$" lies

along the first ν coordinates in the new space $\mathcal{H}$". The symmetric

functions are those vectors whose coordinates are all zero after the

first ν.

The projection operation in $\mathcal{H}$" consists of retaining the

first ν coordinates of the data vector, setting the remainder to zero:

(3.11) $\hat{s}$" = P" d", where P" = $(p"_{ij})$ is given by

$$p"_{ij} = \begin{cases} 1, & i = j \leq \nu \\ 0, & i = j > \nu \\ 0, & i \neq j \end{cases}$$

It is easily verified that **P"** is idempotent and self-adjoint.

The existence of the transformation U is guaranteed by a

theorem in van der Waerden (1955; v. II, p. 134) which can be rephrased

in the present notation:

Given an inner product operator $(f,g)_{\mathcal{H}} = g^* R^{-1} f$ and a
self-adjoint operator P: $(f,Pg)_{\mathcal{H}} = (Pf,g)_{\mathcal{H}}$, there
exists a coordinate transformation UT such that if
f" = UT f, g" = UT g, then $(f",g")_{\mathcal{H}"} = g"^* f"$ and
$(f,Pg)_{\mathcal{H}} = g"^* P" f"$, where P" is a diagonal matrix
whose diagonal entries are the eigenvalues of P.

The eigenvalues of a projection operator must be either zero or one, by the idempotency condition, and the eigenvalue one has multiplicity equal to the dimension of the symmetric function subspace, by the idempotency on $\mathcal{S}$ condition. The operator P'' of the theorem is the desired projection operator in the $\mathcal{H}''$ plane. For simplicity, in our formulation of the $\mathcal{H}''$ space, the non-zero entries on the diagonal of P'' preceded the zero entries. This theorem proves the existence of the solution. The uniqueness proof follows that in Lemma 2.3.

The projection operation in the $\mathcal{H}$ space is found by transforming the data to $\mathcal{H}''$, performing the projection with P'', then retransforming the solution back to $\mathcal{H}$:

$$(3.12) \qquad \hat{s} = P\, d, \text{ where } P = T^{-1}\, U^{-1}\, P''\, U\, T.$$

The P'' operator in $\mathcal{H}''$ is symmetric, idempotent on $\mathcal{S}''$, and self-adjoint by construction. It is now verified that the P operator in $\mathcal{H}$ also satisfies these conditions. Note that $P'' = UTPT^{-1}U^{-1}$ from Eq. 3.12.

Symmetry. -- $A'' = UA'U^{-1} = UTAT^{-1}U^{-1}$. By the symmetry of P'', $P'' = A''P''$ for all A'' in G. Then $UTPT^{-1}U^{-1} = (UTAT^{-1}U^{-1})(UTPT^{-1}U^{-1})$ or $P = AP$ for all A in G.

Idempotency on $\mathcal{S}$. -- $Ps = T^{-1}U^{-1}P''UTs = T^{-1}U^{-1}P''s'' = T^{-1}U^{-1}s'' = s$.

Self-adjointness. -- Note, first, that $R^{-1} = T^*T$, $U^* = U^{-1}$, $(T^*)^{-1} = (T^{-1})^*$, and $P''^* = P''$. Then $P^*R^{-1} = T^*U^*P''^*(U^{-1})^*(T^{-1})^*T^*T = T^*U^*P''UT$. Also, $R^{-1}P = T^*TT^{-1}U^{-1}P''UT = T^*U^{-1}P''UT = T^*U^*P''UT$. Thus $P^*R^{-1} = R^{-1}P$.

$(f, Pg)_{\mathcal{H}} = E\,[g^*P^*R^{-1}f] = E\,[g^*R^{-1}Pf] = (Pf, g)_{\mathcal{H}}$.

B. The Processing Solution

Obtaining the Solution

In practice, it is easier to find a matrix that is simultaneously symmetric[5], idempotent on $\mathcal{S}$, and self-adjoint, than to find the transformations U and T. The equations which must be solved can be expressed in matrix form:

(3.13)   symmetry: $A P = P$, for all A in G

           idempotency on $\mathcal{S}$: $P s = s$, for all s in $\mathcal{S}$

           self-adjointness: $P R = (P R)^* = R P^*$

Since A is a permutation matrix, it is unitary. Thus if A is in G, $A^{-1} = A^*$ is also in G and $A^* P = P$. By self-adjointness, then:

$$PR = RP^* = R(A^*P) = RP^*A = PRA, \text{ for all } A \text{ in } G.$$

By symmetry, $PR = APR$, for all A in G, proving:

(3.14)   If $C = PR$, then $AC = C = CA$, for all A in G.

The elements of C are therefore symmetric in both subscripts:

(3.15) $C = (c_{ij})$, where $c_{ij} = c_{i'j'}$ for $i \cong i'$, $j \cong j'$,

and there are only v^2 different coefficients to be determined.

Since the projection operator P can be obtained from C:

(3.16) $P = C R^{-1}$

---

[5] In the sense that $AP = P$ for all A in G.

it, too, has only ν^2 linearly independent entries.

Define a series of "standard" symmetric functions, s^i, such that the function s^i has value 1 on all points equivalent to x_i and value 0 on all other points. These ν standard functions form a basis for $\mathcal{J}$.

$$(3.17) \qquad s^i = (s^i(x_j)) = \begin{cases} 1, & i \equiv j \\ 0, & i \not\equiv j \end{cases} \qquad \text{for} \quad i = 1, 2, \ldots, \nu$$

The idempotency condition for P requires $s^i = P s^i$, for all i. Applying Eq. 3.16 expresses this condition in terms of the C matrix:

$$(3.18) \qquad s^i = C R^{-1} s^i \qquad \text{for} \quad i = 1, \ldots, \nu$$

Allowing for the symmetry requirement, Eq. 3.15, for the coefficients of C, each value of i in Eq. 3.18 yields ν linearly independent equations. In all, then, Eq. 3.18 expresses the ν^2 equations needed to determine the ν^2 different elements of C. The matrix P is then found by applying Eq. 3.16.

In addition to solving these ν^2 simultaneous linear equations, this solution requires inverting the $N \times N$ covariance matrix. A shorter procedure involves expressing idempotency in terms of P directly:

$$(3.19) \qquad \sum_{j' \equiv j} p_{ij'} = \begin{cases} 1, & i \equiv j \\ 0, & i \not\equiv j \end{cases} \qquad \text{for} \quad i = 1, \ldots, \nu$$

For each value of i, Eq. 3.19 expresses ν conditions on the p_{ij}, so each row of P has only $N-\nu$ independent elements. Because of the symmetry condition, $p_{ij} = p_{i'j}$ for $i \equiv i'$, P has only ν distinct rows. There are $\nu(N-\nu)$ coefficients left to be determined for P.

These may be found by applying Eq. 3.14 and 3.15. In fact, the solution decomposes into the solution of ν separate sets of $N-\nu$ equations which can be chosen to have the same $(N-\nu)\times(N-\nu)$ kernel. The entire solution is then reduced to inverting one $(N-\nu)\times(N-\nu)$ matrix, a significant saving in computational effort.

As an example, let $x_1 \equiv x_3$, $x_2 \equiv x_4$. Applying the symmetry and idempotency requirements to P reduces it to only four coefficients:

$$(3.20) \qquad P = \begin{bmatrix} \alpha, \beta, & 1-\alpha, & -\beta \\ \gamma, \delta, & -\gamma, & 1-\delta \\ \alpha, \beta, & 1-\alpha, & -\beta \\ \gamma, \delta, & -\gamma, & 1-\delta \end{bmatrix}$$

Obtaining C from Eq. 3.14 and applying Eq. 3.15 reduces the problem of finding these coefficients to the solution of the two sets of equations:

$$(3.21) \qquad M \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} R_{33} - R_{31} \\ R_{34} - R_{32} \end{bmatrix} \qquad M \begin{bmatrix} \gamma \\ \delta \end{bmatrix} = \begin{bmatrix} R_{43} - R_{41} \\ R_{44} - R_{42} \end{bmatrix}$$

$$\text{where} \quad M = \begin{bmatrix} R_{11} - R_{13} - R_{31} + R_{33} & R_{21} - R_{41} - R_{23} + R_{43} \\ R_{12} - R_{14} - R_{32} + R_{34} & R_{22} - R_{42} - R_{24} + R_{44} \end{bmatrix}$$

All four coefficients in P may be found by inverting the single 2x2 matrix M. It is interesting to note that M is the covariance matrix of the noise <u>differences</u>.

$$(3.22) \qquad M = \begin{bmatrix} E[\,|n_1 - n_3|^2\,] & E[\,(n_2 - n_4)(n_1 - n_3)^*\,] \\ E[\,(n_1 - n_3)(n_2 - n_4)^*\,] & E[\,|n_2 - n_4|^2\,] \end{bmatrix}$$

<u>Full Symmetry</u>

If all the N data points are equivalent, the symmetry is the full symmetry group on N elements. The unit cell space has only one

point and the symmetric function subspace is one-dimensional. There is

little computational difference between the short and the long solutions,

but the latter can be expressed rather simply, involving the computation

of a single constant.

The matrix $C = PR$ must have all its entries equal, by Eq.

3.15. Writing J for the $N \times N$ matrix with all elements equal to one:

$C = cJ$. Once R^{-1} is determined, the constant c is the only para-

meter left to compute. Equation 3.18 then becomes:

$$
\begin{bmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{bmatrix} = c \begin{bmatrix} 1 & 1 & . & . & 1 \\ 1 & 1 & . & . & 1 \\ . & & & & . \\ . & & & & . \\ 1 & 1 & . & . & 1 \end{bmatrix} R^{-1} \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{bmatrix}
$$

This equation can be solved to give

$$
c = (s^{1*} R^{-1} s^1) = (\sum_{i,j} r^{-1}_{ij})^{-1},
$$

where s^1 is the vector with all N components equal to one. The

reciprocal of c is the sum of all the entries in the R^{-1} matrix.

Applying Eq. 3.16:

$$
(3.23) \qquad P = \frac{J R^{-1}}{s^{1*} R^{-1} s^1} = \frac{s^1 (s^{1*} R^{-1})}{(s^{1*} R^{-1}) s^1} ,
$$

since $J = s^1 s^{1*}$ (cf. Grenander and Rosenblatt, 1957). The coefficients

of P can be written:

$$
(3.24) \qquad p_{ij} = \frac{\sum_k r^{-1}_{jk}}{\sum_{j,k} r^{-1}_{jk}}
$$

The contribution of the j-th data value to the signal estimate is the

ratio of the sum of the elements in the j-th column of the inverse

covariance matrix to the sum of all the elements in the inverse covar-

iance matrix.

In the two-dimensional case, the solution becomes:

$$(3.25) \qquad \hat{s} = \frac{(r_{22}-r_{21})d_1 + (r_{11}-r_{12})d_2}{r_{11} - r_{12} - r_{21} + r_{22}}$$

Estimating the Noise

Some interesting properties of the solution can be derived from

Eq. 3.19 without explicitly solving for P. The estimate of the i-th

signal value can be written, in general:

$$(3.26) \qquad \hat{s}(x_i) = \hat{s}_i = \Sigma_j p_{ij}\, d(x_j)$$

The estimate of the signal at x_i depends on all the data values. The

coefficient p_{ij} is the weight of the contribution of $d(x_j)$ to the

estimate of $s(x_i)$.

Grouping the data into equivalence classes, Eq. 3.26 becomes:

$$(3.27) \qquad \hat{s}_i = \Sigma_j p_{ij}\, d_j + \Sigma_k p_{ik}\, d_k + \ldots$$

The first term on the right is summed over the μ_i points in the equi-

valence class of x_i. The first summation is the contribution of the

data values as measured at points equivalent to the estimate point:

the contribution of "relevant" data values. The remaining summations

are each over a separate equivalence class. They represent the contri-

bution of points not equivalent to the estimate point: the contribution

of "irrelevant" data values.

Equation 3.19 states that the sum of the weights, p_{ij}, in the first summation is 1, the sum of the weights in each other summation is 0. In particular, if x_O is an isolated singular point of the data space, being invariant under all the automorphisms, the equivalence class y_O contains only the one element x_O. This datum value makes no contribution to the estimate of any signal value other than $s(x_O)$.

The question arises: why should any irrelevant data value contribute at all to the estimate? The answer, of course, is that the irrelevant <u>signal</u> does not. The <u>noise</u> at the irrelevant point is correlated with the noise on the desired signal, though, and knowing this noise value aids in removing the noise obscuring the signal. But the noise values, "uncontaminated" by the irrelevant signals, are not known. If the data at two equivalent points are subtracted, the signal will cancel out, leaving the noise difference. These noise differences are used in making the estimate. The datum at an isolated point cannot contribute to the estimate because the noise cannot be separated from the signal.

The contribution of these noise differences is illustrated by rewriting Eq. 3.27 in terms of the signal and noise separately. Because of the constraints of Eq. 3.19, the signal term in all the summations after the first disappears:

$$\hat{s}_i = s_i + \Sigma_j p_{ij}\, n_j + \Sigma_k p_{ik}\, n_k + \ldots$$

The dependence of the estimate on the noise differences is tacitly contained in the constraints on the weights. This dependence may be emphasized:

$$\hat{s}_i = s_i + n_{j_0} + \Sigma_j p_{ij}(n_j - n_{j_0}) + \Sigma_k p_{ik}(n_k - n_{k_0}) + \ldots$$

where x_{j_0} is any point in the equivalence class y_i, x_{k_0} any point in y_k, etc. Since $s_i + n_{j_0} = d(x_{j_0})$, the problem can be restated: The signal s_i is estimated by the data value $d(x_{j_0})$. The coefficients p_{ij} are then chosen, under the constraints of Eq. 3.19, to best cancel out the noise n_{j_0} with the remaining noise differences.

The presence of the irrelevant data does not automatically lead to an improvement in the estimate. Even though the noise at points x_k and $x_{k'}$ may be correlated with the noise at x_{j_0}, the noise difference $(n_k - n_{k'})$ may be uncorrelated with n_{j_0}. In this case, the knowledge of the data at irrelevant points is, in fact, irrelevant to the signal estimate.

In the four point example (Eq. 3.20) the estimate $\hat{s}_1$ depends on all four data values. If the noise difference is uncorrelated between equivalence classes, the matrix M, the covariance of the noise differences, becomes diagonal, as seen from Eq. 3.22. The solution then reduces to

$$\hat{s}_1 = \alpha \, d_1 + (1-\alpha) \, d_3$$

which is the solution to the two-point problem, given by Eq. 3.25.

C. Special Covariance Matrices

The "trivial" Solution

In this section, two special types of noise processes are considered. In these cases, not only are the irrelevant data values

indeed irrelevant, but all the relevant data are weighted equally. The estimate of s_i is just the arithmetic mean of the data at the relevant points:

$$(3.28) \qquad s_i = (1/\mu_i) \sum_{i' \equiv i} d_{i'}$$

where μ_i is the number of points in the equivalence class of x_i. This solution will be called the "trivial" solution, since it seems to ignore all the information about the inter-relationships between the noise samples.

Symmetric Noise

The noise process is symmetric if the properties of the noise process are invariant under all the automorphisms of the symmetry group. For our purposes, only the first two moments of the noise process are considered, so a weaker condition is sufficient. The noise process is "wide sense" symmetric if the covariance function satisfies:

$$R(x^a, (x')^a) = R(x,x') \quad \text{for all} \quad a \quad \text{in} \quad G.$$

In matrix terms, $R = E[n \, n^*]$, so the symmetry requirement becomes: $R(x^a, (x')^a) = E[(An)(An)^*] = E[Ann^*A^*] = ARA^*$. The noise process is wide sense symmetric if and only if:[6]

$$(3.29) \qquad R = A R A^*, \text{ for all } A \text{ in } G.$$

The self-adjoint property can then be expanded:

----

[6] Since the permutation matrices, A, are unitary, the white noise process, $R = I$, is wide sense symmetric.

$$PR = R \ P^*$$

$$P \ ARA^* = ARA^* \ P^* \quad \text{for all} \quad A \quad \text{in} \quad G.$$

By symmetry, $\quad AP = A \quad$ and $\quad P^*A^* = (AP)^* = P^* \quad$ so

$$AP \ ARA^* = ARA^* \ P^*A^*$$

$$PAR = RA^*P^*$$

$$(PA)R = R(AP)^*$$

Thus the matrix PA is self-adjoint. PA is symmetric since $A(PA) = APA = PA$ for all A in G. PA is also idempotent on $\mathcal{S}$, since $(PA)s = P(As) = Ps = s$. Therefore PA is a solution to the processing problem. Since P is the unique solution, $PA = P$ for all A in G, so that $p_{ij} = p_{ij'}$ for $j \cong j'$. Thus all the summands in the sum of Eq. 3.19 are equal:

$$\sum_{j' \cong j} p_{ij'} = \mu_j \ p_{ij} = \begin{cases} 1, & i \cong j \\ 0, & i \not\cong j \end{cases}$$

The solution is therefore the trivial one of Eq. 3.28.

A special case of symmetric noise is the statistical problem arising by taking uncorrelated samples from a single multivariate probability distribution. Let each sample produce ν data values. The $\nu \times \nu$ covariance matrix R_ν describes the relationships between the noise in any one sample. Since different samples are assumed uncorrelated, the total covariance matrix is simply:

$$R = \begin{bmatrix} R_\nu & 0 & . & . & 0 \\ 0 & R_\nu & . & . & 0 \\ . & & . & & . \\ . & & & . & . \\ 0 & 0 & . & . & R_\nu \end{bmatrix}$$

where O represents a $v \times v$ block of zeroes. Since the covariance matrix R is symmetric, the solution is the "trivial" one: given uncorrelated samples from a multivariate probability distribution, the sample mean vector is the best estimate of the distribution mean vector in the sense of minimizing mean square error. This estimate is unbiased.

Stationary Noise

The concept of stationarity is meaningless unless the data space is a topological group. It will thus be considered in full detail in a later chapter. Because this type of noise does lead to the "trivial" processing, and because it can be discussed with the mathematical structure that has already been developed, it will be considered briefly.

The finite data space, x_i for $i = 0,1,...N-1$, is now assumed to form a group by defining addition: $x_i + x_j = x_{i+j}$, where the summation on the subscripts is taken modulo N. In essence, this means that x_O is "adjacent" to x_{N-1}, just as x_O is adjacent to x_1.

The noise process is stationary if the joint probability distribution of n_i and n_j depends only on the difference $i-j$. The condition of "wide sense" stationarity is sufficient for our purposes: the noise is wide sense stationary if the covariance function $R(x_i, x_j)$ depends only on the difference $i-j$. In matrix terms: $R = (r_{ij}) = (r_{i-j})$.

$$
R = \begin{bmatrix}
r_o, & r_1, & r_2, & \cdot & \cdot & \cdot & r_{N-1} \\
r_1^*, & r_o, & r_1, & \cdot & \circ & \cdot & r_{N-2} \\
r_2^*, & r_1^*, & r_o, & \cdot & \cdot & \cdot & r_{N-3} \\
\cdot & & & & & & \cdot \\
\cdot & & & & & & \cdot \\
\cdot & & & & & & \cdot \\
r_{N-1}^*, & r_{N-2}^*, & r_{N-3}^*, & \cdot & \cdot & \circ & r_o
\end{bmatrix}
$$

Each row of R is identical to the row above, but shifted one unit to the right. The topological group structure means that an entry shifted off the right end re-appears on the left. Such a matrix is called a "circulant," with eigenvalues λ_j and eigenfunctions ϕ_j given by (Marcus, 1960):

$$\phi_j R = \lambda_j \phi_j ,$$

where $\lambda_j = \Sigma_k R_k \exp(2\pi \underline{i} jk/N),$ $\underline{i}^2 = -1$

and $\phi_j = (1, \exp(2\pi \underline{i} j/N), \exp(4\pi \underline{i} j/N), \ldots , \exp[2(N-1)\pi \underline{i} j/N])$

Since the eigenfunctions and eigenvalues are known, the transformations V and D are easily found. The matrix V is the matrix whose j-th row is the j-th eigenfunction. The matrix D is the diagonal matrix whose j-th diagonal entry is the j-th eigenvalue. V is given by:

$$V = (\phi_{j,k}), \text{ where } \phi_{j,k} = \exp(2\pi \underline{i} jk/N)$$

The transformation $\mathbf{g} = V f$ becomes:

$$\mathbf{g}_j = \Sigma_k d_k \exp(2\pi \underline{i} jk/N).$$

The j-th coefficient of $\mathbf{g}$ is exactly the coefficient of the j-th term in the Fourier expansion of f. In fact, the eigenfunction λ_j is the j-th coefficient in the power spectral density function of the noise. The mathematical manipulations we have created reduce to the Fourier transform when the problem is specialized enough for Fourier analysis to apply!

Continuing with the solution, assume the symmetry to be full, so that all the points are equivalent. A function is symmetric if and

only if it is a constant. Therefore its spectrum consists only of the

"DC term." The signal subspace in either $\mathcal{H}_1$ or in $\mathcal{H}'$ is already

aligned along the $j = 0$ axis. The processing therefore consists of

computing the transform of the data, ignoring all terms except the zero

frequency term, then retransforming. This procedure, of course, is

exactly the "trivial" processing of Eq. 3.28.

CHAPTER IV

REPLICATED DATA SPACES

A. The Processing Problem

Reduction to the Finite Case

In this chapter, the data are assumed to be derived from a

series of replicated experiments. Each trial produces a data function,

rather than the single datum value of Chapter III. The symmetry is

assumed to be full -- the signal function is identical from trial to

trial. The first part of this chapter shows how this replicated data

problem reduces to the problem of the finite data space, so that the

solution involves the inversion of a finite matrix or the solution of

a finite number of simultaneous linear equations. In contrast with the

previous chapter, the matrices will have functions, rather than numbers,

as entries. The simultaneous equations will be linear integral, rather

than algebraic, equations.

The data space for this type of data is illustrated in Fig.

2.1. The data obtained from the i-th trial are a function defined on

the space X_i. This space is a "building block" for X, $X = \bigcup_i X_i$.

Under the symmetry group, each X_i is equivalent to every other X_j

and also to the unit cell space Y. Any point x in X must be

located in one of the building blocks, $x = y_i$ for some y in Y,

$i \leq N$. Any function $f(x)$ defined over X can thus be represented as

a series of functions, each defined over a separate block:

$f(x) = f(y_i) = f_i(y)$, $y \in Y$, $i = 1, 2, \ldots, N$. This series of functions

can be written as a column vector:

$$f = (f_i(y)) = \begin{bmatrix} f_1(y) \\ f_2(y) \\ \cdot \\ \cdot \\ \cdot \\ f_N(y) \end{bmatrix}$$

Two points x and x' in X are equivalent if and only if $x = y_i$ and $x' = y_j$: equivalent points can differ only in their subscripts. The automorphisms of the symmetry group permute the subscripts. In the vector notation, the permuted function $f'(x) = f(x^a)$ has the same components as $f(x)$, but in different order. Thus $f' = A f$, where A is an $N \times N$ permutation matrix as in Chapter III. The symmetric functions are those with equal components:

(4.1) $\qquad f \in \aleph$ if and only if $f_i(y) = f_j(y)$ for all $y \in Y$; $i,j = 1,\ldots,N$.

In order to proceed with the machinery of Chapter III, it is necessary to represent transformations of coordinates and linear operators as matrices with functions as entries. The operators that will be needed include the transformations $V, T,$ and U; the projection operators P and P'', and the linear operators associated with the inner product notion, R^{-1} and R. The permutation matrix A is also an example of such an operator matrix.

Consider a linear transformation, represented by the integral operator $h(x,x')$ as in Chapter II.

(4.2) $\qquad f'(x) = \int_X h(x,x') \, f(x') \, dx'$

Writing x as y_i and noting that the integral over X can be broken down into a sum of integrals, each over a subspace X_i:

$$f'(y_i) = \sum_{j=1}^{N} \int_{X_j} h(y_i,x') \, f(x') \, dx'$$

The variable of integration, x', in each integral is limited to one subspace, so the integral can be rewritten:

$$f'(y_i) = \sum_{j=1}^{N} \int_Y h(y_i,y'_j) \, f(y'_j) \, dy'$$

The kernel function $h(y_i,y'_j)$ is now broken up into a series of functions:

$$(4.3) \qquad h(y_i,y'_j) = h_{ij}(y,y')$$

so that the transformation becomes:

$$f'(y_i) = \sum_{j=1}^{N} \int_Y h_{ij}(y,y') \, f(y'_j) \, dy'$$

This result is formally similar to Eq. 3.3, but rather than multiply the datum value $f(x_j)$ by the constant h_{ij} as in Eq. 3.3, the data function $f(y'_j)$ is <u>composed</u> with the function $h_{ij}(y,y')$:

$$(4.4) \qquad h_{ij} * f_j = \int_Y h_{ij}(y,y') \, f_j(y') \, dy'$$

The operator matrix H can now be defined as the matrix of functions $h_{ij}(y,y')$ as given in Eq. 4.3. The transformation of Eq. 4.2 becomes:

$$f' = H \, f$$

where the multiplication of elements is given by the composition of Eq. 4.4.

In general, the multiplication of two matrices is given by the rule: If $U = (u_{ij})$ and $V = (v_{ij})$, the product $U\ V = W = (w_{ij})$ is given by:

$$(4.5) \qquad w_{ij} = \Sigma_k\ u_{ik} * v_{kj},$$

where $u_{ik}* v_{kj} = \int_Y u_{ik}(y,y'')\ v_{kj}(y'',y')\ dy''.$

In order to extend this definition to include the product of a matrix with a vector or the scalar product of two vectors, the definition of composition is extended to include:

$$f_i\ * h_{ij} = \int_Y f_i(y')\ h_{ij}(y\ ,y)\ dy'$$

$$f_i\ * g_i = \int_Y f_i(y)\ g_i(y)\ dy$$

As an illustration of these composition methods using the matrix notation, consider the desired inner product notation:

$$(f,g)_{\mathcal{H}} = E[g^* R^{-1} f].$$

This can be written out in full as:

$$(f,g)_{\mathcal{H}} = \Sigma_{i,j} E[g_i^* * R_{ij}^{-1} * f_j]$$

$$= \Sigma_{i,j} \iint_Y E[g_i^*(y)\ R_{ij}^{-1}(y,y')\ f_j(y')]\ dy\ dy'$$

$$= \Sigma_{i,j} \int_{X_i} \int_{X_j} E[g^*(x)\ R^{-1}(x,x')\ f(x')]\ dx\ dx'$$

$$= \int_X E[g^*(x)\ R^{-1}(x,x')\ f(x')]\ dx\ dx'$$

Thus the matrix notation for the inner product corresponds exactly with the integral formulation of Chapter II. Note that the asterisk has been used in three different senses in these equations: as a superscript on a constant or a function it denotes complex conjugation; as a superscript on a vector it denotes conjugate transposition; and as an operation between two functions it denotes composition. These usages are always distinguished by the context.

In order to utilize fully the matrix methods of Chapter III, the matrix entries, the operator functions, must form a field of characteristic not equal to 2 (Van der Waerden, 1955; v. II, p. 128). Three conditions must therefore be satisfied by the operator functions: (1) There must be an <u>identity</u> function, I_{ij}, with the property that $I_{ij} * f_j = f_i$ and $f_i * I_{ij} = f_j$ for all functions f_i, f_j. (2) Each operator h_{ij} must have an <u>inverse</u>, h_{ij}^{-1}, such that $h_{ij} * h_{jk}^{-1} = h_{ij}^{-1} * h_{jk} = I_{ik}$. And (3) it must not be the case that $h_{ij} * h_{jk} = I_{ik}$ for all operator functions h.

Condition (3) is automatically satisfied if the data values can be numbers other than zero and one -- the processing techniques do not necessarily apply to binary data. Condition (2), the existence of inverses, need not apply if we are satisfied merely with writing down the set of simultaneous linear equations that the processing operator must satisfy. The techniques involved in actually solving the equations will not be discussed here.

Condition (1), the existence of the identity function, cannot be dismissed so easily. In fact, the entries of the permutation matrix A are the operators O_{jk} and I_{jk}, not the numbers 0 and 1. The

issue of the identity function was avoided by defining symmetric functions in Eq. 4.1 without reference to the permutation matrices. The required identity is the delta function, widely used in signal theory even though mathematically it is not a function. Rigor may be maintained, for the present purposes, by including the identity operator as a possible entry in the operator matrices. Thus the operator matrix H has, as entries, the function of Eq. 4.3 or the identity operator. Matrix multiplication is given by the rule in Eq. 4.5 except in the case of the identity operator, when the properties given above in condition (1) are used. The identity operator need occur only in the operator matrices. The function vectors (the signal, the noise, the observed data, and the estimated signal) never contain the identity as an entry. Although the idempotency condition in Chapter III was expressed with the aid of the "standard" symmetric functions (Eq. 3.17) which contain these identity operators as elements, the idempotency condition in this chapter can be expressed in other terms.

The Solution

The optimal processing operator is an operator matrix, P, which must satisfy the three conditions:

$$(4.6) \qquad \begin{array}{lll} \text{symmetry:} & A\,P = P, \text{ for all } A \text{ in } G \\ \text{idempotency on } \mathscr{S}: & P\,s = s, \text{ for all } s \text{ in } \mathscr{S} \\ \text{self-adjointness:} & P\,R = R\,P^* \end{array}$$

The symmetry requirement is satisfied by setting $p_{ij} = p_{i'j}$ for all i, i'. Since full symmetry is assumed, only one estimate is needed and the first subscript can be dropped, the processing operator matrix reducing to a single row:

(4.7) $\qquad \hat{s} = \Sigma_j \; p_j * d_j$

$$\hat{s}(y) = \Sigma_j \int_Y p_j(y,y') \; d_j(y') \; dy'$$

In order to avoid inverting the R matrix, the "short" solution is used, following Eq. 3.19. The solution is thus expressed in terms of N-1 simultaneous linear equations. Since these equations involve the composition operator, they are integral equations. The idempotency condition, Eq. 3.19, becomes:

(4.8) $\qquad \Sigma_j \; p_j = I,$

that is, $\Sigma_j \; p_j * f = I f = f$, for all f in $\mathcal{H}$.

By Eq. 3.14, the matrix C = P R has all its entries equal:

(4.9) $\qquad \Sigma_j \; p_j * r_{jk} = c = c(y,y'), \quad k = 1,\dots,N.$

Since it is not necessary to compute the function c, set:

$$c = \Sigma_j \; p_j * r_{j1}$$

so that Eq. 4.9 reduces to:

(4.10) $\qquad \displaystyle\sum_{j=1}^{N} p_j * (r_{jk} - r_{j1}) = 0, \quad k = 2,\dots,N.$

Equation 4.8 can be rewritten: $p_1 = I - \displaystyle\sum_{j=2}^{N} p_j$, so that Eq. 4.10 becomes:

$$\sum_{j=2}^{N} p_j * (r_{jk} - r_{j1}) = - p_1 * (r_{1k} - r_{11})$$

$$= - (r_{1k} - r_{11}) + \sum_{j=2}^{N} p_j * (r_{1k} - r_{11})$$

(4.11) $\qquad \displaystyle\sum_{j=2}^{N} p_j * (r_{11} - r_{1k} - r_{j1} + r_{jk}) = (r_{11} - r_{1k}), \quad k = 2,\dots,N.$

Equations 4.8 and 4.11 give the N equations necessary to determine the N entries of the processing vector. Since all rows of the matrix P are identical, the processing operator P is completely determined.

B. Wiener Multivariate Smoothing

The Dummy Signal

The Wiener filtering problem is derived under the assumption that the signal, as well as the noise, is a sample function from a random process. The data space is assumed to be the real line, all functions being functions of time, t. The signal is assumed to have zero mean, $E[s(t)] = 0$, for all t, and the signal covariance is known: $R_s(t,t') = E[s(t)s^*(t')]$. The covariance between the signal and each noise term is known: $R_{sn_i}(t,t') = E[s(t)n_i^*(t')]$, as well as the noise covariance: $R_{n_i n_j} = E[n_i(t)n_j^*(t')]$. Finally, it is assumed that the N data functions, d_i, $i = 1,\ldots,N$, are given and that each is the sum of a particular signal sample function and the corresponding noise sample: $d_i(t) = s(t) + n_i(t)$.

The problem is the same as the symmetric function problem: find a processing operator that produces, from the data, the best estimate of the signal in the sense of minimizing the mean square error. The solution can be obtained as the solution to an N+1 dimensional symmetric function problem. Define the "noise" function n_0 to be the negative of the random signal: $n_0(t) = -s(t)$. The remaining N noise terms, $n_i(t)$, in the symmetric function problem are equal to the corresponding N noise terms of the Wiener problem. Define a "dummy" signal function $s_d(t)$ and N+1 dummy data functions: $f_i(t) = s_d(t) + n_i(t)$, $i = 0,1,\ldots,N$. The noise covariance matrix will be

(N+1)-dimensional, $R = (r_{ij})$, $i,j = 0,1,\ldots,N$:

(4.12)
$$r_{oo}(t,t') = R_s(t,t')$$

$$r_{oj}(t,t') = -R_{sn_j}(t,t')$$

$$r_{ij}(t,t') = R_{n_i n_j}(t,t')$$

The negative sign in the definition of r_{oj} is due to the fact that the noise n_o is the negative of the random signal.

The solution to this $N+1$ dimensional symmetric function problem is the solution of the N equations given in Eq. 4.11, where now the summation index runs from $j = 1$ to N, and the special role of subscript 1 is replaced with subscript 0:

(4.13)
$$\sum_{j=1}^{N} p_j * (r_{oo}-r_{ok}-r_{jo}+r_{jk}) = (r_{oo}-r_{ok}), \quad k = 1,\ldots,N.$$

Assume that the solution has been obtained, so that:

$$\hat{s}_d = \sum_{j=0}^{N} p_j * f_j$$

Since $n_o = f_o - s_d$, the best estimate of n_o is $\hat{n}_o = f_o - \hat{s}_d = n_o + s_d - \hat{s}_d$:

$$\hat{n}_o = n_o + s_d - \sum_{j=0}^{N} p_j * (s_d + n_j)$$

By Eq. 4.8, $\sum_j p_j * s_d = s_d$, so that the dummy signal cancels out, leaving:

$$\hat{n}_o = n_o - \sum_{j=0}^{N} p_j * n_j$$

Again using Eq. 4.8, write $p_o = I - \sum_{j=1}^{N} p_j$, so that:

$$\hat{n}_o = n_o - p_o * n_o - \sum_{j=1}^{N} p_j * n_j$$

$$= n_o - n_o + \sum_{j=1}^{N} p_j * (n_o - n_j)$$

But $n_0 = - s$, so that $\hat{n}_0 = - \hat{s}$ and $n_0 - n_j = - (s+n_j) = - d_j$.

$$(4.14) \qquad \hat{s} = \sum_{n=1}^{N} p_j * d_j$$

Therefore the solution for the Wiener multivariate smoothing problem, given in Eq. 4.14, is obtained by solving the symmetric function problem Eq. 4.13.

Define two new covariance functions, the data covariance $R_{d_j d_k}(t,t') = E [d_j(t)d_k^*(t')]$ and the signal-to-data covariance $R_{sd_j}(t,t') = E [s(t)d_j^*(t')]$.

$$R_{d_j d_k} = E [(s+n_j)(s+n_k)^*] = E [ss^* + sn_k^* + n_j s^* + n_j n_k^*]$$

$$= R_s + R_{sn_k} + R_{n_j s} + R_{n_j n_k}$$

and $R_{sd_j} = E [s(s+n_j)^*] = E [ss^* + sn_j^*] = R_s + R_{sn_j}$

Using the covariance functions defined in Eq. 4.12, Equations 4.13 reduce to:

$$(4.15) \qquad \sum_{j=1}^{N} p_j * R_{d_j d_k} = R_{sd_k}, \quad k = 1, \ldots, N$$

as the equations determining the Wiener smoothing filter.

The Augmented Function Space

The Wiener multivariate smoothing problem may also be expressed in terms of the function space of the symmetric function problem, without the introduction of a dummy deterministic signal. In order to develop the relation between the two problems, they must both be expressed in the same language.

In the symmetric function problem, the function space $\mathcal{H}$ is an N-dimensional vector space.[1] The N data functions determine the data point d in this space. The assumption of full symmetry determines the symmetric function subspace $\mathcal{S}$ -- the line passing through the origin with the equation:

$$(4.16) \qquad x_i = x_j, \quad i,j = 1,2,\ldots,N$$

The actual noise sample vector n has coefficients $n_i = d_i - s$. Since $n_i - d_i = n_j - d_j = s$, this noise vector must lie on the <u>noise line</u> $\mathcal{L}_n$:

$$(4.17) \qquad x_i - d_i = x_j - d_j, \quad i,j = 1,2,\ldots,N$$

This noise line is the line that passes through the data point and is parallel to $\mathcal{S}$ (Fig. 4.1). By Eq. 2.15, the problem of finding the point $\hat{s}$ in $\mathcal{S}$ closest to d -- that is, minimizing $\|\hat{s}-d\|$ -- is equivalent to finding the point in $\mathcal{L}_n$ closest to the origin -- that is, minimizing $\|\hat{n}\| = \|d-\hat{s}\|$. This point is found by projecting the origin orthogonally to the line $\mathcal{L}_n$ (Fig. 4.1), according to the inner product notion in $\mathcal{H}$.

The Weiner problem involves the random signal process as well as the N random noise processes. The space $\mathcal{H}$ is thus augmented by the signal process, obtaining the N+1 dimensional vector space $\mathcal{H}_a$. The inner product and distance notions in $\mathcal{H}_a$ are given by the augmented (N+1)x(N+1) matrix R_a, defined in Eq. 4.12. A point in $\mathcal{H}_a$
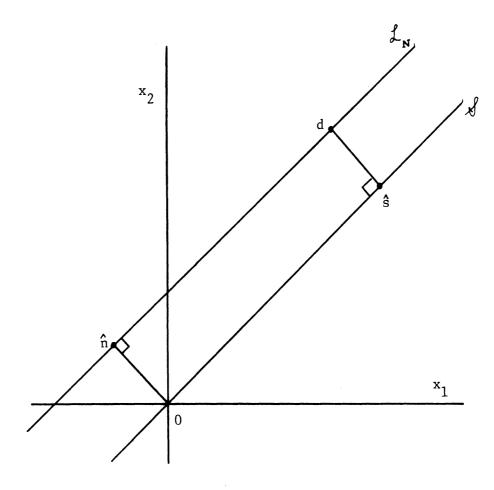
---

[1] over the ring of functions, $f(y)$.

Figure 4.1. The Symmetric Function Subspace and the Noise Line.

is determined only when the signal value s, as well as the N noise

values n_i, are known -- N+1 coordinates in all. Since the observed

data provide only N values, $d_i = s + n_i$, the data are not sufficient

to fix a data point in $\mathcal{H}_a$. They do determine a line, though, the <u>data

line</u> $\mathcal{L}_d$:

(4.18) $x_o + x_1 = d_i$, $i = 1,2,\ldots,N$

where x_o is the signal axis.

The signal point s_a in $\mathcal{H}_a$, given by the actual signal

value and the N noise samples, must lie on this data line. The Wiener

problem is to find the point $\hat{s}_a$ in $\mathcal{H}_a$ that lies closest to the actual

signal point s_a, using the notion of distance in $\mathcal{H}_a$. This problem is

equivalent, by an argument similar to that in Chapter II, to finding

the point in $\mathcal{L}_d$ which lies closest to the origin. Thus the signal

estimate $\hat{s}_a$ is found by projecting the origin orthogonally to the data

line.

The above discussion may be summarized as follows:

The symmetric function smoothing problem is equi-
valent to projecting the origin orthogonally onto the
<u>noise line</u> $\mathcal{L}_n$ in $\mathcal{H}$, the notion of orthogonality being
derived from the noise covariance R.

The Wiener smoothing problem is equivalent to project-
ing the origin orthogonally onto the <u>data</u> line $\mathcal{L}_d$ in $\mathcal{H}_a$,
the notion of orthogonality being derived from the aug-
mented covariance R_a.

These two problems are related by the fact that:

The noise line $\mathcal{L}_n$ is the geometric projection of the
data line $\mathcal{L}_d$ in $\mathcal{H}_a$ to the subspace $\mathcal{H}$.

The proof of this statement is simple. A point on $\mathcal{L}_d$ has coordinates,

by Equation 4.18: $(x_o, d_1-x_o, \ldots, d_N-x_o)$. The geometric projection of

this point to the $\mathcal{H}$ space is found merely by dropping the x_o coordinate: $(-, d_1 - x_o, d_2 - x_o, \ldots, d_N - x_o)$, where x_o is a constant. The coordinates of this point obviously satisfy Eq. 4.17, the equation of the line $\mathcal{L}_n$.

CHAPTER V

TOPOLOGICAL GROUP DATA SPACES

A. Data Processing on Groups

Topological Groups

If the data space forms a topological group, the powerful

techniques of Fourier analysis can be used to describe the data proces-

sing operations. In this chapter, the data space will be assumed to

be a locally compact, Abelian group. For simplicity, the data functions

are assumed to be real.

A topological group is a topological space whose elements

form a group. Thus, two points x_1 and x_2 in the space can be

added: $x_3 = x_1 + x_2$. This addition must be continuous -- loosely

speaking, if one of the addends is changed slightly, the sum is changed

slightly. There must be a zero element, 0, so that $x + 0 = 0 + x = x$,

for all x in X. Finally, there must be negatives, so that $x + (-x)$

$= (-x) + x = 0$. The group is Abelian, or commutative, if $x + x'$

$= x' + x$, for all x, x' in X. The condition of local compactness is

a type of "finiteness" condition -- finite-dimensional spaces are

locally compact.

The data space usually considered is the real line, the data

being a function of time. In practice, the data cannot extend over an

infinite duration, so the data space is really a finite line segment,

which is not a group. In order to maintain group structure, the ends

of the segment are joined to make a circle. The data are thus, in

effect, defined over the whole real line as a periodic function, one

period being the interval of observation.

The data space considered in optical data processing is usually the two-dimensional plane, again a topological group. Here, too, the data do not actually extend over an infinite region. The actual data function is extended over the whole plane by defining it to be zero outside the region of observation.

When the data space is a locally compact Abelian group, the function space $\mathcal{H}$ can be transformed -- using the Fourier transform -- to the new space $\mathcal{H}_F$. The function $f(x)$ is transformed into the function $F(w)$, the Fourier transform of $f(x)$. The transformed variable w forms a topological group W, the dual group of X.

The unit cell space Y is the quotient space X/G (cf. Definition 2.8). If the symmetry group G is a normal subgroup of X, the unit cell space is also a topological group and the functions on Y -- the symmetric functions -- will have Fourier transforms. The transformed variable will form the topological group Ω, the dual group of Y. The transformed symmetric function space can be embedded in the transformed function space $\mathcal{H}_F$ since the group Ω is a subgroup, as well as a subspace, of W. Specifically, Ω is the "annihilator"[1] of G in W. The symmetric functions in $\mathcal{H}_F$ are precisely those functions[2] $F(w)$ such that $F(w) = 0$ for $w \notin \Omega$. Thus the coordinate system of $\mathcal{H}_F$ is aligned with the symmetric function subspace.

The data processing operator can be represented as the integral:

---

[1] Pontrjagin, 1939; Theorem 34.

[2] Rudin, 1962; Theorems 2.7.2, 2.7.4.

$$(5.1) \qquad \hat{s}(x) = \int_X p(x,x') \, d(x') \, dx'$$

In the notation of the previous chapter (Eq. 4.4), the estimated data function is the composition of the processing operator with the data: $\hat{s} = p * d$. Since the data are assumed to be real, Parseval's theorem provides that:

$$(5.2) \qquad \hat{S}(w) = \int_W P(w,w') \, D(w') \, dw'$$

where $P(w,w')$ is the "two-dimensional" transform of $p(x,x')$, the result of transforming each variable separately. The processing operation in $\mathcal{H}_F$ is again a composition. Under two special conditions -- first, if the data space X is compact, and second, if the processing operator is translation invariant -- Eq. 5.2 represents a significant simplification over Eq. 5.1.

If the data space is a compact topological group, its dual group is discrete.[3] The integral in Eq. 5.2 becomes a summation:

$$(5.3) \qquad \hat{s}_j = \sum_k P_{jk} \, D_k$$

where the P_{jk} and D_k are complex numbers. The transformed space $\mathcal{H}_F$ now has a well-defined identity function:

$$(5.4) \qquad I_{jk} = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

that satisfies $f_j = \sum_k I_{jk} f_k$. This condition is explored further in the final part of this chapter.

---

[3] Pontrjagin, 1939; Theorem 31.

An important subclass of processing operations has the property

of being translation invariant. If P is a translation invariant

operator and $\hat{s}(x) = P[d(x)]$, then $\hat{s}(x + x_0) = P[d(x + x_0)]$. In signal

theory, the corresponding condition is time-invariance -- the processing

operation is independent of the time of occurrence of the signal. The

optical data processing technique of Cutrona, et al., (1960) is transla-

tion invariant. A translation invariant operator has the property that

$$(5.5) \qquad p(x,x') = p(x-x')$$

and the composition integral of Eq. 5.1 becomes the convolution integral:[4]

$$(5.6) \qquad \hat{s}(x) = \int_X p(x-x') \, d(x') \, dx'$$

In the transform plane, the processing operator is not the composition

of Eq. 5.2, but has a multiplicative form:[5]

$$(5.7) \qquad \hat{S}(w) = P(w) \, D(w)$$

where $P(w)$ is the transform of $p(x)$. The identity function is simply

$$(5.8) \qquad I(w) = 1, \quad \text{for all} \quad w \quad \text{in} \quad W$$

and the inverse of $F(w)$ is $1/F(w)$.

Data Space Structure and Symmetry

In order to utilize fully the advantages of the Fourier trans-

form, both the data space and the unit cell space have to be topological

---

[4] Rudin, 1962; Theorem 3.8.4.

[5] Rudin, 1962; Theorem 3.8.3.

groups. The symmetry group must therefore be compatible with the group structure of the data space. In this section, it will be proved that the only symmetry compatible with any topological group is pure translational symmetry. In a pure translation group, all the automorphisms have the form: $x^a = x + 0^a$.

The points x and x' are equivalent under the symmetry group (written $x \cong x'$) if there is an automorphism, a, in G such that $x' = x^a$ (cf. Def. 2.7). In order for the symmetry group to be compatible with the group structure of the data space, the automorphisms of G must preserve addition:

$$(5.9) \qquad (x + x')^a \cong x^a + x'^a, \quad (-x)^a \cong -(x^a)$$

It follows that the equivalence class of zero, the set of all points of the form 0^a, is a subgroup of X, the "kernel" of the symmetry group.

The unit cell space Y is the set of equivalence classes of X. Algebraically, Y is a homomorphic image of X. Thus Y can be written $Y = X/K$, where K, the kernel of the homomorphism, is the kernel of the symmetry group. In order for Y to be a group, equivalent points can differ only by an element in the kernel: $x \cong x'$ implies $x - x' \cong 0$.

If Y is to be a group, the addition of equivalence classes must be well-defined. Thus if $x_1 \cong x_1'$ and $x_2 \cong x_2'$, then $x_1 + x_2 \cong x_1' + x_2'$.

These two conditions, which must hold if Y is to be a topological group, are each equivalent to the condition that the symmetry group be purely translational.

Theorem 5.1: The following statements are equivalent:

(1) $x_1 \cong x_1'$ and $x_2 \cong x_2'$ implies $x_1 + x_2 \cong x_1' + x_2'$

(2) If $x \cong x'$, then $x - x' \cong 0$

(3) The symmetry group is a pure translation group.

Proof: (1) implies (2): Let $x_1 = x$, $x_2 = x_2' = -x'$, $x_1' = x'$. If $x \cong x'$ then $x_1 \cong x_1'$, so by (1), $x_1 + x_2 = x - x' \cong x_1' + x_2' = x' - x' = 0$.

(2) implies (3): The function $f(x) = x^a - x$ is a continuous function of x, since addition and subtraction, as well as the automorphisms of the symmetry group, are continuous. But $x^a \cong x$, so by (2), $f(x) = x^a - x \cong 0$. Since G is a discrete group, the kernel K is a discrete set. Thus $f(x)$ is a continuous function with discrete values and is therefore a constant. Since $f(0) = 0^a - 0 = 0^a$, $f(x) = 0^a$ for all x, or $x^a = x + 0^a$. The same argument applies for each a, so that the symmetry group is pure translational.

(3) implies (1): $x_1 \cong x_1'$ implies $x_1' = x_1^a$ for some a in G. Hence by (3), $x_1' = x_1 + 0^a$. Similarly $x_2 = x_2'$ implies $x_2' = x_2 + 0^b$ for some b in G. Then $x_1' + x_2' = x_1 + x_2 + (0^a + 0^b) \cong x_1 + x_2$.

B. Special Processing Solutions

The "trivial" Solution

In the case of a finite data space, the processing solution became a "trivial" averaging operator when the covariance function was either symmetric or stationary. These results are now extended to more general data spaces.

Assume the symmetry group is finite, of order μ. The trivial

processing technique is defined to be the operation:

$$(5.10) \qquad \hat{s}(x) = (1/\mu) \sum_i d(x^{a_i})$$

where the summation is over the whole symmetry group. In operator terminology, the automorphism $a: x \to x^a$ becomes the operator $A: f(x) \to f(x^a)$.

> Definition 5.1. The trivial processing operator is
> $P_t = (1/\mu) \sum_i A_i$, summing over the symmetry group.

Note that P_t is symmetric and idempotent on $\mathcal{S}$, and that $P_t^* = P_t$.

In particular, if $R = I$, then P_t is the optimal processing operator.

If P is the optimal processing operator, then $P P_t[f] = P_t[f]$, since $P_t[f]$ is a symmetric function. That is, $P P_t = P_t$.

> Lemma 5.1. If $P A = P$ for all A in G, then P is
> the trivial processing operator P_t.

Proof: $P P_t = P (1/\mu) \sum_i A_i = (1/\mu) \sum_i P A_i = (1/\mu) \sum_i P$, by the hypothesis. Thus $P P_t = P$. But $P P_t = P_t$, proving that $P = P_t$.

In terms of integral operators, the condition of the lemma is that $p(x, x'^a) = p(x, x')$, for all a in G.

The noise process is defined to be wide-sense symmetric if the covariance function is invariant under all automorphisms in G:
$R(x^a, x'^a) = R(x, x')$. In operator terminology:

> Definition 5.2. The noise process is wide-sense sym-
> metric if $A R A^* = R$, for all A in G.

As in the finite case, the optimal processing operator for symmetric noise is the trivial one:

Theorem 5.2. If the noise process is wide-sense symmetric, the optimal processing operator is P_t.

Proof: P_t is symmetric and idempotent on $\mathcal{J}$, as noted above. Also:

$P_t R = (1/\mu) \Sigma_i A_i R = (1/\mu) \Sigma_i (A_i R A_i^*) A_i = (1/\mu) \Sigma_i R A_i = R P_t$. Since

$P_t = P_t^*$, $P_t R = R P_t^*$, so P_t is also self-adjoint. Thus it is the

solution to the processing equations, and is the optimal operator.

The above results are valid for general data spaces -- they

do not require topological group structure. Now suppose that the data

space and the unit cell spaces are both topological groups. If the

processing operator is translation invariant, then $P(x,x') = P(x-x')$.

Theorem 5.3. If both X and Y are topological groups and the processing operator is translation invariant, then the optimal processing operator is P_t.

Proof: If P is translation invariant, then $P(x,x'^a) = P(x-x'^a)$.

But if Y is a group, then by Theorem 5.1, $x'^a = x' + 0^a$, so $P(x,x'^a)$

$= P(x-x'-0^a)$. But $-0^a \equiv (-0)^a = 0^a$, so that $-0^a = 0^b$ for some b in

G and $x-0^a = x+0^b = x^b$. Thus $P(x,x'^a) = P(x^b-x') = P(x^b,x')$. Since

P is symmetric, $P(x^b,x') = P(x,x')$. Therefore $P(x,x'^a) = P(x,x')$.

By Lemma 5.1, $P = P_t$.

The condition of wide-sense stationarity for the noise, given

in Chapter III, is simply the condition that the covariance function be

translation invariant.

Definition 5.3. The noise process is wide-sense stationary if the covariance function is translation invariant.

Theorem 5.4. If both X and Y are topological groups

and the noise process is wide-sense stationary, then
the optimal processing operator is P_t.

Proof: By hypothesis, $R(x,x') = R(x-x')$. Then $R(x^a, x'^a) = R(x^a - x'^a)$.
But if both X and Y are topological groups, $x^a = x + 0^a$ and x'^a
$= x' + 0^a$, so $x^a - x'^a = x - x'$. Thus the covariance function is sym-
metric and, by Theorem 5.2, the processing operator is P_t.

Replicated Data

Suppose now that the data space is obtained through a repli-
cated experiment, as in Chapter IV, and that the building block, the
data space of each trial, is a topological group. An example is the
case when each trial produces a function of time as data. The entries
in the matrices and vectors of Chapter IV are all functions defined
over Y, so that they have Fourier transforms. If the noise process is
wide-sense stationary within each trial, the processing operator within
a trial is translation invariant: $p_j(y,y') = p_j(y-y')$ (See Eq. 4.7).
The composition operation involved in matrix multiplication of the
transforms (Eq. 4.5) can be replaced by simple multiplication of the
transforms (Eq. 5.7), so that the matrix entries form a field. The
matrices can thus be inverted and the processing solution can be
obtained in closed form.

Processing on Compact Groups

Assume that the data space X is a compact topological group,
that the unit cell space is a group, and that the symmetry group is of
finite order, μ. The dual space W of X is discrete. The space W
can be thought of as a transformed data space, the transformed functions

being functions over W. The transformed symmetric functions are defined

over the dual group Ω of Y, which is a subgroup of index μ in W.

The space $\Omega = \Omega_O$ and its cosets Ω_i, $i = 1,\ldots,\mu-1$ form a type of

"building block" structure for W: $W = \bigcup_i \Omega_i$. Since W is discrete,

the Ω_i are disjoint. Furthermore, each Ω_i has the same algebraic

and geometric structure as Ω_O. Following a procedure analogous to

that in the previous chapter, write the function $F(w)$ as a column

vector:

$$(5.11) \qquad F(w) = (F_j(\omega)) = \begin{bmatrix} F_O(\omega) \\ F_1(\omega) \\ \cdot \\ \cdot \\ F_{\mu-1}(\omega) \end{bmatrix}$$

where each F_j is a function over Ω_j. The symmetric functions are

those for which $F_j(\omega) = 0$, $j \neq 0$.

The linear transformations and operators are defined as matrices

whose entries are functions of two variables, just as in Chapter IV:

$H = (H_{jk}(\omega,\omega'))$. Matrix multiplication involves the composition of H_{ij}

with F_j, but because Ω is discrete, the composition integral becomes

the summation:

$$(5.12) \qquad H_{ij} * F_j = \sum_{\omega'} H_{ij}(\omega,\omega') F_j(\omega') = F'_j(\omega)$$

The identity operator is well-defined:

$$(5.13) \qquad I_{ij}(\omega,\omega') = \begin{cases} 1, & \omega = \omega' \\ 0, & \omega \neq \omega' \end{cases}$$

If $F' = P F$, then in order for P to be symmetric, F'_i

must be zero for $i \neq 0$: $P_{ij} = 0$ for $i \neq 0$. Only the top row of the

P matrix is non-zero. Idempotency on $\aleph_F$ requires that

$\Sigma_j\ P_{oj} * S_j = P_{oo} * S_o = S_o$, so that $P_{oo} = I_{oo}$. Finally, self-adjointness requires that $P\ R = C$ be symmetric in both variables (cf. Eq. 3.14), so that $\Sigma_i\ P_{oi} * R_{ij} = C_{oj} = 0$, for $j \neq 0$. The optimal processing operator can then be written:

$$(5.14) \qquad \hat{S}_o = D_o + \sum_{i=1}^{\mu-1} P_{oj} * D_j$$

where the P_{oj} must satisfy the equations:

$$(5.15) \qquad \sum_{j=1}^{\mu-1} P_{oj} * R_{jk} = -R_{ok}$$

If the matrix entries form a field so that functions have inverses, the solution may be obtained very simply: $P = C\ R^{-1}$. Since $C_{ij} = 0$ unless $i = j = 0$, $P_{oj} = C_{oo} * R^{-1}{}_{oj}$. But $P_{oo} = C_{oo} * R^{-1}{}_{oo} = I_{oo}$, so that $C_{oo} = (R^{-1}{}_{oo})^{-1}$, and:

$$(5.16) \qquad P_{oj} = (R^{-1}{}_{oo})^{-1} * R^{-1}{}_{oj}$$

Processing Example

Let the signal be a periodic function of the real line, with period x_o. The data is observed for, say, three cycles -- over an interval of length $X_o = 3\ x_o$. Any function defined on an interval of length X_o can be written as the Fourier series:

$$(5.17) \qquad f(x) = \Sigma\ F(w)\ \exp\ (\underline{i}wW_ox), \text{ where } W_o = 2\pi/X_o$$

The function $F(w)$, defined for integral w, is the transform of $f(x)$. $F(w)$ can be written as the vector $(F_j(\omega))$, $j = 0,.,2$, where

$$(5.18) \qquad F_j(\omega) = F(3\omega + j).$$

The symmetric functions have period x_o, so they have the series form:

$$(5.19) \qquad s(x) = \Sigma \, S(\omega) \, \exp{(\underline{i}\omega W_o x)}, \text{ where } w_o = 2\pi/x_o$$

Since $w_o = 3W_o$, the variable ω can be written as $3w$:

$$(5.20) \qquad s(x) = S(3w) \, \exp{(\underline{i}3wW_o x)}.$$

Writing $S(w)$ as a vector gives $S_0(\omega) = S(3w)$, and $S_1(\omega) = S_2(\omega) = 0$, as expected.

Let the noise be white, for simplicity, but let it be time-varying so that its amplitude will fade in and out:

$$(5.21) \qquad R(x,x') = (1 + \cos W_o x) \, \delta(x - x')$$

The covariance function R has the Fourier series representation:

$$(5.22) \qquad R(x,x') = \Sigma_{w,w'} R(w,w') \, \exp{(\underline{i}wW_o x)} \, \exp{(-\underline{i}w'W_o x')}$$

$$\text{where } R(w,w') = \begin{cases} 1, & w'-w = 0 \\ 1/2, & |w'-w| = 1 \\ 0, & |w'-w| > 1 \end{cases}$$

Writing R in matrix form, $R = (R_{jk})$, where $R_{jk}(\omega,\omega') = R(3\omega+j, \; 3\omega'+k)$ gives:

$$(5.23) \qquad R_{00} = R_{11} = R_{22} = I = \begin{cases} 1, & \omega = \omega' \\ 0, & \omega \neq \omega' \end{cases}$$

$$R_{01} = R_{10} = R_{12} = R_{21} = (1/2)I = \begin{cases} 1/2, & \omega = \omega' \\ 0, & \omega \neq \omega' \end{cases}$$

$$R_{02} = \begin{cases} 1/2, & \omega = 1+\omega' \\ 0, & \text{otherwise} \end{cases}$$

$$R_{20} = \begin{cases} 1/2, & \omega' = 1+\omega \\ 0, & \text{otherwise} \end{cases}$$

To solve the processing equations, it may be noted that the entries in the covariance matrix are all translation invariant:

$R_{jk}(\omega,\omega') = R_{jk}(\omega-\omega')$. Thus they have Fourier transforms which combine multiplicatively, and inverses can be found. The matrix R can be inverted and the solution found from Eq. 5.16.

$$R = \begin{bmatrix} I, & I/2, & L/2 \\ I/2, & I, & I/2 \\ L'/2, & I/2, & I \end{bmatrix}$$

where L is the R_{02} function interpreted as an operator, L' the R_{20} function.

$$R^{-1} = \frac{1}{2I+(1/2)(L+L')-LL'} \begin{bmatrix} 3I, & L-2I, & I-2L \\ L'-2I, & 4I-\mathbf{LL'}, & L-2I \\ I-2L', & L'-2I, & 3I \end{bmatrix}$$

so that $P_{01} = \dfrac{L-2I}{3I}$, $P_{02} = \dfrac{I-2L}{3I}$, giving the processing formula:

(5.24) $\hat{S}_o = D_o + (1/3)(R_{02}-2I)*D_1 + (1/3)(I-2R_{02})*D_2$

The same result is obtained by solving the two simultaneous equations given in Eq. 5.15, without inverting R.

Writing out the compositions indicated in Eq. 5.24, using the rule of Eq. 5.12 with the values of R_{02} in Eq. 5.23, gives:

$$\hat{S}_o = D_o(w) - (2/3) D_1(\omega) + (1/3) D_1(\omega-1)$$
$$+ (1/3) D_2(\omega) - (2/3) D_2(\omega-1)$$

Rewriting according to Eq. 5.18 gives:

(5.25) $\hat{S}(3w) = D(3w) - (2/3) D(3w+1) + D(3w-1)$
$$+ (1/3) D(3w+2) + D(3w-2)$$

Using standard Fourier transform techniques, it is easily verified that:

$$(5.26) \qquad \hat{s}(x) = w(x)\, d(x) + w(x+x_0)\, d(x+x_0) + w(x+2x_0)\, d(x+2x_0)$$

where $w(x)$ is a weighting function given by

$$(5.27) \qquad w(x) = (2/9)(\cos 2W_0 x - 2 \cos W_0 x + 3/2)$$

$$= \left(\frac{2 \cos W_0 x - 1}{3} \right)^2$$

The value of the estimate at any point x is the weighted average of the data at the three equivalent points, x, $x+x_0$, and $x+2x_0$. Since the noise is uncorrelated between points, "irrelevant" data values do not enter into the estimate. The sum of the weights adds to one, as is required of an idempotent estimate. The trivial processing technique is given by Eq. 5.26, using the weighting function $W_t(x) = 1/3$. Both the trivial weighting and the optimal weighting functions are shown in Fig. 5.1.

The optimal weighting function weights the noise heavily when the noise amplitude (given by Eq. 5.21) is small. Note that the noise is zero for $x = 3x_0/2$, so the data at that point is taken as the signal estimate. Thus the equivalent points $x = x_0/2$ and $x = 5x_0/2$ are assigned weight zero. The smallest weight is not necessarily assigned to the points with the greatest noise: $x = 0$ and $x = 3x_0$.
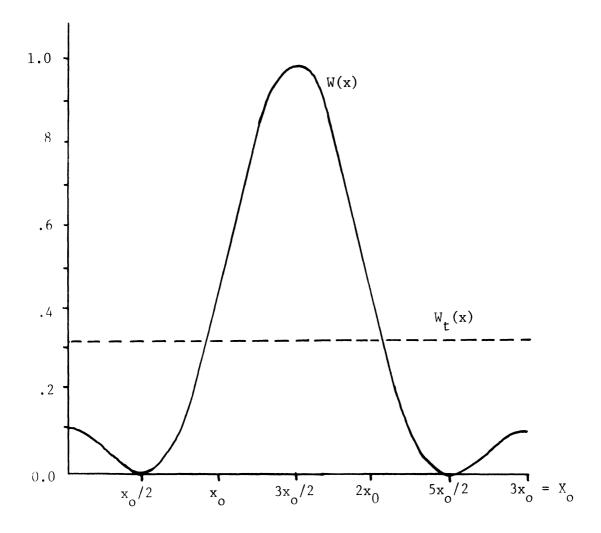
Figure 5.1. The Weighting Functions in the Processing Example.

BIBLIOGRAPHY

Agrawal, Hari Om, J. W. Kent, and D. M. MacKay. 1965. Rotation Technique in Electron Microscopy of Viruses. Science 148: 638-640.

Almeida, June D. 1963. A Classification of Virus Particles Based on Morphology. Can. Med. Ass. J. 89:787-798.

Aronszajn, N. 1950. Theory of Reproducing Kernels. Amer. Math. Soc., Trans. 68:337-404.

Bendat, Julius S. 1957. Optimum Filters for Independent Measurements of Two Related Perturbed Messages. Inst. Radio Eng., Trans. Circuit Theory CT-4:14-19.

Berger, J.E., C. Richard Zobel, and P. E. Engler. 1966. Laser as Light Source for Optical Diffractometers; Fourier Analysis of Electron Micrographs. Science 153:168-169.

Buerger, M.J. 1956. Elementary Crystallography. Wiley, N. Y.

Capon, Jack. 1965. Hilbert Space Methods for Detection Theory and Pattern Recognition. Inst. Electrical Electronics Eng., Trans. Information Theory 11:247-259.

Caspar, D.L.D. and A. Klug. 1962. Physical Principles in the Construction of Regular Viruses. Cold Spring Harbor Symp. Quant. Biol. 27:1-24.

Cochran, W., F.H.C. Crick, and V. Vand. 1952. The Structure of Synthetic Polypeptides. I. The Transform of Atoms on a Helix. Acta Cryst. 5:581-586.

Cramér, Harald. 1946. Mathematical Methods of Statistics. Princeton University Press, Princeton.

Cramér, Harald, 1951. A Contribution to the Theory of Stochastic Processes, p. 329 to 339. In Jerzy Neyman ed. Proc. Second Berkeley Symp. Math. Stat. and Prob.

Crick, F.H.C. and J. D. Watson. 1956. Structure of Small Viruses. Nature 177:473-475.

Crick, F.H.C. and J. D. Watson. 1957. Virus Structure: General Principles, p. 5 to 13. In G.E.W. Wolstenholme and E.C.P. Millar eds. The Nature of Viruses. Churchill, London.

Cutrona, L.J., E. N. Leith, C. J. Palermo, and L. J. Porcello. 1960. Optical Data Processing and Filtering Systems. Inst. Radio Eng., Trans. Information Theory IT-6:386-400.

Davenport, Wilbur B., Jr. and William L. Root. 1958. An Introduction to the Theory of Random Signals and Noise. McGraw-Hill, N.Y.

Doob, J. L. 1953. Stochastic Processes. Wiley, N. Y.

Friedrichs, K. O. 1960. Spectral Theory of Operators in Hilbert Space. New York University, Institute of Mathematical Sciences. (copyright, K. O. Friedrichs, 1961).

Grenander, Ulf, and Murray Rosenblatt. 1957. Statistical Analysis of Stationary Time Series. Wiley, N. Y.

Hall, Marshall, Jr. 1959. Theory of Groups. Macmillan, N. Y.

Halmos, Paul, R. 1957. Introduction to Hilbert Space and the Theory of Spectral Multiplicity. 2d ed. Chelsea Publishing Company, N. Y.

Heerden, P. J. van. 1963. Theory of Optical Information Storage in Solids. Appl. Optics $\underline{2}$:393-400.

Heidenreich, Robert D. 1964. Fundamentals of Transmission Electron Microscopy. Interscience, Wiley, N. Y.

Horne, R. W. 1963. The Structure of Viruses. Sci. Amer. $\underline{208}$:48-56.

Horne, R. W. 1964. Electron Microscopy of Viruses. Sci. Progress $\underline{52}$:525-542.

Horne, R. W. and P. Wildy. 1963. Virus Structure Revealed by Negative Staining. Advances Virus Res. $\underline{10}$:101-170.

Hosaka, Yasuhiro. 1965. A Criterion for Evaluating the Number of Capsomeres of Icosahedral Capsids. Biochim. Biophys. Acta $\underline{104}$:261-273.

Howatson, A. F. 1965. Structure of Viruses of the Papilloma-Polyoma Type. J. Mol. Biol. $\underline{13}$:959-960.

Karhunen, Kari. 1947. Über Lineare Methoden in der Wahrscheinlichkeits-rechnung. Ann. Acad. Sci. Fennicae (Suomalaisen Tiedeakatemian Toimituksia) I. Math.-Phys. $\underline{37}$:79pp.

Kawata, Tatsuo. 1965. Fourier Analysis of Nonstationary Stochastic Processes. Amer. Math. Soc., Trans. $\underline{118}$:276-302.

Klein, E. and G. Langner. 1963. Relations Between Granularity, Graininess and the Wiener-Spectrum of the Density Deviations. J. Phot. Sci. $\underline{11}$: 177-185.

Klug, A. and J. E. Berger. 1964. An Optical Method for the Analysis of Periodicities in Electron Micrographs, and Some Observations on the Mechanism of Negative Staining. J. Mol. Biol. $\underline{10}$:565-569.

Klug, A. and J. T. Finch. 1965. Structure of Viruses of the Papilloma-Polyoma Type. J. Mol. Biol. $\underline{13}$:961-962.

Kozma, Adam and David Lee Kelly. 1965. Spatial Filtering for Detection of Signals Submerged in Noise. Appl. Optics. 4:387-392.

Loève, Michel. 1960. Probability Theory. Van Nostrand, Princeton.

Marcus, Marvin. 1960. Basic Theorems in Matrix Theory. National Bureau of Standards Applied Mathematics Series - 57. U.S. Government Printing Office, Washington, D. C.

Markham, Roy, Simon Frey, and G. J. Hills. 1963. Methods for the Enhancement of Image Detail and Accentuation of Structure in Electron Microscopy. Virology 20:88-102.

Middleton, David. 1960. An Introduction to Statistical Communication Theory. McGraw-Hill, N. Y.

Murray, F. J. 1941. An Introduction to Linear Transformations in Hilbert Space. Princeton University Press, Princeton.

Norman, Richard S. 1966. Rotation Technique in Radially Symmetric Electron Micrographs: Mathematical Analysis. Science 152:1238-1239.

Parzen, Emanuel. 1961. An Approach to Time Series Analysis. Ann. Math. Stat. 32:951-989.

Parzen, Emanuel. 1962. Extraction and Detection Problems and Reproducing Kernel Hilbert Spaces. J. Soc. Ind. Appl. Math., Control Series A 1:35-62.

Pontrjagin, L. 1939. Topological Groups. (trans. Emma Lehmer) Princeton University Press, Princeton.

Rosenblatt, Murray. 1962. Random Processes. Oxford University Press, N. Y.

Rudin, Walter. 1962. Fourier Analysis on Groups. Interscience, Wiley, N. Y.

Stewart, R. M. and R. J. Parks. 1957. Degenerate Solutions and an Algebraic Approach to the Multiple-Input Linear Filter Design Problem. Inst. Radio Eng., Trans. Circuit Theory CT-4:10-14.

Thiry, H. 1963. Power Spectrum of Granularity as Determined by Diffraction. J. Phot. Sci. 11:69-77.

Upatnieks, Juris and Emmet N. Leith. 1964. Lensless, Three-dimensional Photography by Wavefront Reconstruction. J. Opt. Soc. Amer. 54:579-580.

Valentine, R. C. and N. G. Wrigley. 1964. Graininess in the Photographic Recording of Electron Microscope Images. Nature 203:713-715.

Waerden, B. L. van der. 1955. Algebra. 3rd ed. of Moderne Algebra. Springer-Verlag, Berlin. 2 vols.

Weyl, Hermann. 1952. Symmetry. Princeton University Press. Princeton.

Wiener, N. 1949. The Interpolation, Extrapolation, and Smoothing of Stationary Time Signals. Wiley, N. Y.

Wiener, N. and P. Masani. 1957. Multivariate Stochastic Processes. I. The Regularity Condition. Acta Math. 98:111-150.

Wiener, N. and P. Masani. 1958. Multivariate Stochastic Processes. II. The Linear Predictor. Acta Math. 99:93-137.

Williams, M. G., A. F. Howatson, and J. D. Almeida. 1961. Morphological Characterization of the Virus of the Human Common Wart (verruca vulgaris). Nature 139:895-897.

Zadeh, Lotfi A., and John R. Ragazzini. 1950. An Extension of Wiener's Theory of Prediction. J. Appl. Phys. 21:645-655.