TECHNICAL REPORT ECOM-01870-14-T

October 1967

OPTIMAL SPACE-TIME SIGNAL PROCESSING AND PARAMETER ESTIMATION WITH EMPHASIS ON ITS APPLICATION TO DIRECTION FINDING

SEL Technical Report No. 16

Contract No. DA 28-043-AMC-01870(E)

DA Project No. 1P021101 A042.01.02

Prepared by

Donald G. Mueller

The University of Michigan
SYSTEMS ENGINEERING LABORATORY
Department of Electrical Engineering
Ann Arbor, Michigan

for

UNITED STATES ARMY ELECTRONICS COMMAND, FORT MONMOUTH, N. J.

ERRATA

Page	Line	Correction
11	15	Change "where" to "were"
14	4	Insert the word "finding" between "direction" and "systems"
38	7	Delete "s" on "proceedings"
67	24	Insert the word "a" between "any" and "priori"
71	11	Insert the words "necessary and" between "is" and "sufficient"
72	4	Place quotation marks around the word "nondecreasing"
76	7	Change "of" to "or"
77	18	Insert the word "of" between "class" and "processes"
81	4	Change "zero mean" to "zero-mean"
82	3	Change "classify" to "clarify"
85	20	Insert the words "the "equivalence classes" of" between "includes" and "all"
102	18	Change "$\rightarrow$" to "="
106	14	Change "Let P_o and P_i" to "Let $\hat{P}_o$ and $\hat{P}_i$"
107	7	Change "Conversely, it can" to "But, by symmetry, it can also"
108	11	Change "on" to "to"
109	2	Change "B_j" to "$\mathcal{B}_j$"
115	10	Change "defined" to "define"
145	18	Change "λ_i^i" to "λ_j^i"
200	20	Change "$j \in K$" to "$k \in K$" and "$[Z^i]_k^T$" to "$[Z^i]_k$"

TABLE OF CONTENTS

TABLE OF CONTENTS (Concluded)

LIST OF FIGURES

LIST OF FIGURES (Concluded)

LIST OF SYMBOLS

General notes: Vectors in real Euclidean spaces will be underlined while vectors and matrices whose components are real-valued functions will be enclosed in brackets. The components of a vector $\underline{v}$ in J-dimensional Euclidean space will be denoted by $v_1, v_2, \ldots, v_J$ and the components of a K-dimensional vector of real-valued functions [v] will be denoted by $[v]_1, [v]_2, \ldots, [v]_K$. The components of a KxK matrix of real-valued functions [V] will be denoted by $[V]_{11}, [V]_{12}, \ldots, [V]_{KK}$. Symbols which are used with several indexes in the text will be represented below with the indexes i, j, and k. Also, symbols which depend on other variables will be written both with and without the dependence indicated. Finally, the symbols $\mathcal{B}$ and P with a variety of subscripts and superscripts will be reserved to denote σ-fields and probability measures.

Symbol	Description	First Reference
α	the true angle of arrival	14
α_e	an estimate of α	22
α_i	a possible angle of arrival	61
α^j	the true angle of arrival relative to the jth reference direction	55
α^o	a reference direction	14
$\hat{\alpha}$	an optimal estimate of α	16
β	the signal strength at the array site	16
Γ, Γ_j	the filter and control functions	37
γ^4	the fourth moment of α about its mean	26
$\Delta_\alpha, \Delta_\eta, \Delta_y$	small variations in the variables α, η, and y	44

Symbol	Description	First Reference
Δ_α^η	a change in an estimate of α due to the presence of noise	46
$\Delta_{\alpha_1}^\beta,\Delta_{\alpha_2}^\beta$	changes in an estimate of α due to imprecise knowledge of β	46
$\Delta_{\alpha_1}^{\beta_1},\Delta_{\alpha_1}^{\beta_2},\Delta_{\alpha_2}^{\beta_1},\Delta_{\alpha_2}^{\beta_2}$	components of $\Delta_{\alpha_1}^\beta$ and $\Delta_{\alpha_2}^\beta$	46
δ	a positive constant	59
$\delta(\ )$	the Dirac delta function	39
$\varepsilon_o,\varepsilon',\varepsilon_1,\varepsilon_2$	small positive constants	22
$\varepsilon(\alpha,u_j^1,u^2)$	the difference between $C(\alpha,u_j^1,u^2)$ and the first three terms of the power series expansion of $C(\alpha,u_j^1,u^2)$	22
ε	the error in a direction of arrival estimate	174
$\eta(t),\eta_j,\underline{\eta}$	the noise in the DF model, its jth sample, and a vector of such samples	18
$\{\eta_j\},\{\overline{\eta}_j\}$	sequences of random variables which allow a "decomposition" of the process $[Z^i]$	120
θ	an unknown phase angle	16
Λ_i	the ith element of a partition of the observation space	61
$\Lambda_i^*,\overline{\Lambda}_i^*$	the ith elements of optimal partitions of the observation space	110
Λ_k^i	the kth eigenvalue of the operator R^i	122
ρ_k	the kth eigenvalue of the operator $I-X^*X$	128
Σ_1,Σ_2	summations which allow the derivation of a bound on $E\{\varepsilon^2\}$	183
$\sigma_\alpha^2,\sigma_\beta^2,\sigma_\eta^2$	the variances of the random variables α,β, and η_j	20
τ	time variable	66

Symbol	Description	First Reference
τ_j	the time of propagation of a signal from the jth antenna site to the origin	59
$[\phi_j],[\overline{\phi}_j]$	an eigenvector of the operator I-X*X and an element of a basis for the null space of I-X*X	120
$[\psi_j^i]$	the jth eigenvector of the operator R^i	122
$\Omega(\;)$	the region of variation of the variables listed in the parentheses	34
Ω	the observation space	96
ω	the "frequency" variable of a Fourier transform	63
ω_o	the center frequency of a band-limited signal	14
$[A_1(\alpha_i)],[A_2(\alpha_i)]$	matrices which are deterministic functions of α_i	153
a_o	a constant related to the bandwidth of the noise	166
a'_k	a positive number	66
$B(\alpha_i)$	a bias of the statistic ℓ_i	164
B_j	represents a device which delays and amplifies signals received by the jth array element	14
$\mathcal{B}_i$	a σ-field of subsets of Ω	99
$b_o,\hat{b}_o,\tilde{b}_o$	the filter bias, its optimal and suboptimal values	22
$b_j,\underline{b}$	the "weight" applied to the jth sample by the filter and a vector of such "weights"	22
$\hat{\underline{b}},\tilde{\underline{b}},\tilde{\underline{b}}_1,\tilde{\underline{b}}_2$	the optimal and several suboptimal forms of the vector $\underline{b}$	22
b_o	a constant related to the bandwidth of the signal	166

Symbol	Description	First Reference
$C'(\alpha,\beta,u^1,u^2)$	a function which summarizes the effects of the variables α,β,u^1, and u^2 on the signal $s'(t)$	16
$C''(\alpha,u^1,u^2)$	the amplitude antenna pattern	16
$C(\alpha,u^1,u^2)$	an approximation for $C''(\alpha,u^1,u^2)$	18
C_1,C_2	variables which depend on the design parameters of the DF system	50
C	a positive constant	59
$\underline{c}_o,\underline{c},\underline{c}'$	constant vectors	23
c	velocity of propagation of signals	55
D,D_ε	the covariance function of the random vector $\underline{y}$ and the error that results when it is approximated	25
D_i	a constant necessary for the evaluation of the Radon-Nikodym derivative $d\overline{P}_i/d\overline{P}_o$	133
D	the separation of array elements in the example	167
det{ }	the determinant of the quantity within the brackets	163
$d\overline{P}_i/d\overline{P}_o$	the Radon-Nikodym derivative of $\overline{P}_i$ with respect to $\overline{P}_o$	110
$\underline{d},\underline{d}_\varepsilon$	the expectation of a particular vector and the error that results when it is approximated	25
E{ }	the expectation of the quantity within the brackets	23
E_i{ }	the expectation of the quantity within the brackets when P_i is the probability measure	108
E_S,E_N	constants which indicate signal and noise levels present	166
$[F(\omega)]$	form of signal and noise power spectrums in special cases	63

Symbol	Description	First Reference
$F_N(\omega), F_S(\omega), F_h(\omega)$	the Fourier transforms of $r_N(\tau)$, $r_S(\tau)$, and $h(\tau)$	148
$F_S^+(\omega), F_N^+(\omega)$	positive portions of the spectrums $F_S(\omega)$ and $F_N(\omega)$	150
$F_{S_o}(\omega), F_{N_o}(\omega)$	the Fourier transforms of $r_{S_o}(\tau)$ and $r_{N_o}(\tau)$	151
$F_{h_1}(\omega), F_{h_2}(\omega)$	functions of ω which allow the determination of $F_h(\omega)$	151
$\mathfrak{F}\{\ \}$	the Fourier transform of the quantity within the brackets	151
$G(\alpha_e), G_j(u_j)$	expressions which are minimized by an optimal estimate and the optimal jth control	35
H_i	a bounded self-adjoint operator which allows the evaluation of the Radon-Nikodym derivative $d\overline{P}_i/d\overline{P}_o$	133
H_i^T, H_i^∞	integral operators whose kernels are $h_i^T(t,s)$ and $h_i^\infty(t-s)$	145
H	an integral operator whose kernel is $h(\tau)$	164
$[h_\alpha^T(t,s)], [h_\alpha^\infty(t-s)]$	kernels of linear filters which allow the evaluation of the statistic $\ell(\alpha)$	65
$h(\tau)$	an impulse response from which $[h_\alpha^T(t,s)]$ can be constructed	147
$h_1(\tau)$	an impulse response related to $h(\tau)$	152
$h_{1r}(\tau), h_{1i}(\tau)$	the real and imaginary parts of $h_1(\tau)$	152
$\hat{h}_1(\tau), \hat{h}_2(\tau)$	realizable impulse responses	155
I	the identity matrix and the identity operator	26
I_i	the value of the Radon-Nikodym derivative $d\overline{P}_i/d\overline{P}_o$ at the observation $[y]$	139

Symbol	Description	First Reference
Im( )	the imaginary part of the quantity within the parentheses	152
J	the total number of samples of the observation interval	18
K	the total number of antennas in the array	14
K'	the value of (K-1)/2	183
L	a loss function	35
$\overline{L}_2[T]$	the space of all K-dimensional square integrable functions on $[0,T_f]$	87
$\ell'(\alpha_i)$	a quantity related to the array antenna pattern	178
$\ell_i,\ell(\alpha_i)$	the value of $(H_i[y],[y])$ which allows the evaluation of the Radon-Nikodym derivative $d\overline{P}_i/d\overline{P}_o$	142
$\underline{\ell}_k$	a vector drawn from the origin to the kth array element	55
M	a real integer and the set of integers {1,2,...,M}	61
M_o	the set {o,1,2,...,M}	97
m_α,m_β	the means of the variables α and β	21
[N(t)]	the corrupting noise process	59
$[N_1],[N_2]$	the noisy components of $[Y_1]$ and $[Y_2]$	175
N_i	the noise term corrupting the statistic $\ell(\alpha_i)$	180
$\overline{N}^1,\overline{N}^2$	quantities which are related to the noise energy in the system	182
N( )	the Hilbert-Schmidt norm of the operator within the parentheses	95

Symbol	Description	First Reference
$\overline{P}_i$	a complete probability measure on the σ-field $\mathcal{B}_i$	99
P_e	the probability of error	61
Pr( )	the a priori probability of the event within the parentheses	61
Pr_i( )	the probability of the event within the parentheses given α_i is the true direction of arrival	61
$p_j, \underline{p}$	a term in the power series expansion of $C(\alpha, u_j^1, u^2)$ and a vector of such terms	22
p( )	the probability density of the variables listed within the parentheses	22
$\underline{\tilde{p}}, \underline{\tilde{p}}$	the values of $\underline{p}$ for different values of the control $\underline{u}^1$	23
p_o( )	the probability density of the noise samples	34
$p(x_1/x_2)$	the conditional probability density of x_1 given x_2	36
$q_j, \underline{q}, \underline{\tilde{q}}_1, \underline{\tilde{q}}_2$	a term in the power series expansion of $C(\alpha, u_j^1, u^2)$, a vector of such terms, and the values of $\underline{q}$ for two forms of $\underline{u}^1$	22
q_{max}	the maximum length of the vector $\underline{q}$	23
R_T^k	the space of all real valued function on the interval $[0, T_f]$	96
R_i, R_i^*	an integral operator on $\overline{L}_2[T]$ and its adjoint	105
$R_i^{1/2}, R_i^{-1/2}$	the square root of R_i and its inverse	105
Re( )	the real part of the quantity within the parentheses	150
$r_j, \underline{r}, \underline{\tilde{r}}_1$	a term in the power series expansion of $C(\alpha, u_j^1, u^2)$, a vector of such terms, and the value of $\underline{r}$ for a particular $\underline{u}^1$	22

Symbol	Description	First Reference
$[r^{o}(t,v)],[r^{(\alpha)}(t,v)],[r^{\alpha}(t,v)]$	the covariance functions of the processes $[N(t)]$, $[S^{\alpha}(t)]$, and $[Y^{\alpha}(t)]$	59
$r_N(\tau)$	the covariance function of the noise in a special case	66
$r_S(t,v)$	the covariance function of the noise-free process driving the array element located at the origin	59
$r_{S_o}(\tau),r_{N_o}(\tau)$	the covariance function of processes related to the "slowly varying" portion of $r_S(\tau)$ and $r_N(\tau)$	150
S	a quantity related to the signal energy received by the array	178
$[S_1],[S_2]$	the pure signal portions of $[Y_1]$ and $[Y_2]$	176
Sf_i^+,Sf_f^-	"shift" operators	163
$[S^{\alpha}(t)],[s(t)]$	the signal process driving the array and its sample function	55
$s_1(t),s_2(t)$	the quadrature components of a received sample	176
$s'(t)$	a noise-free signal in the DF system model	16
T_f	the end of the observation interval	16
Tr{ }	the trace of the indicated matrix	121
tr{ }	the trace of the indicated operator	95
t,t_j	the time variable and its jth sample	16
U	a Hilbert-Schmidt operator	132
$u^1(t),u_j^1,\underline{u}^1$	the pointing angle control, its jth sample and a vector of such samples	16

LIST OF SYMBOLS (Concluded)

Symbol	Description	First Reference
$u(t), u_j, \underline{u}$	another representation for $u^1(t)$, u_j^1, $\underline{u}^1$	36
$u^2(t), u_j^2, \underline{u}^2$	the beam width control, its jth sample and a vector of such samples	16
$\hat{\underline{u}}^1, \tilde{\underline{u}}^1, \underline{u}^1_1$	the optimal and two suboptimal forms of $\underline{u}^1$	22
$\underline{u}_j$	a vector of the controls $u_1, u_2, \ldots, u_j$	36
$\underline{u}$	a unit vector drawn from the origin toward the signal source	55
X, X^*	the operator $R_i^{1/2} R_o^{-1/2}$ and its adjoint	117
$[Y_1], [Y_2]$	sufficient statistics which allow the evaluation of ℓ_i	153
$[Y^\alpha(t)], [Y^i(t)], [y(t)]$	the process driving the array when α is the true direction of arrival, the process driving the array when α_i is the true direction of arrival, and a sample function of these processes	60
$y(t), y_j, \underline{y}$	a noisy signal in the DF system model, its jth sample, and a vector of such samples	18
$y'(t)$	a noisy signal in the DF system model	16
$\underline{y}_j$	a vector of the samples $y_1, y_2, \ldots, y_j$	36
$[Z^i], [Z']$	special random processes defined on Ω	103
$\{\ \}^T$	the transpose of the quantity within the brackets	24
$\{\ \}^{-1}$	the inverse of the quantity within the brackets	26
"–"	denotes complex conjugate	85
"*"	denotes convolution	151

ABSTRACT

The purpose of this study is to develop techniques for processing signals received by a spacial array of K omnidirectional antennas so as to produce optimal direction-of-arrival estimates. Such estimates are useful for navigational purposes in which case the signals are electromagnetic in nature as well as for seismic investigations where the signals are mechanical vibrations transmitted through the earth. Another possible application is an underwater direction finding system which utilizes an array of hydrophones for its receiving antennas.

Two distinct approaches to the problem are pursued which apply differing mathematical disciplines. The first approach models the array as an element of a phased array direction finding system and attempts to apply "Stochastic Optimal Control Theory" to produce otpimal control laws for directing the pointing angle and specifying the beam width of the array. Optimal filtering of the signals is also considered. The controls and filter which result in "Minimum Error Variance" are considered optimal. The second approach applies "Estimation Theory" and considers the array only as an information gathering device whose signals are to be processes directly without the "prefiltering" present in phased arrays. The optimality criterion of this approach requires the direction-of-arrival estimate to result in a "Minimum Probability of Error" or, what is shown to be equivalent, to have "Maximum a Posteriori Probability."

In the Control Theory approach, the model is investigated under both open-loop and closed-loop operating conditions. The closed-loop analysis applies "Dynamic Programming" techniques and produces an integral equation whose solution represents the optimal controls. Unfortunately, the nonlinearities in the system prevent an explicit solution of this equation. The open-loop analysis assumes the a priori probability density of possible angles of arrival is concentrated in a "narrow" region about the a priori expected direction of arrival. An optimal filter and optimal controls are obtained for this case.

The "Estimation Theory" approach assumes the received signals are samples from a Gaussian random process which are corrupted by additive Gaussian noise. Extensive use is made of the "Theory of Random Processes" and "Functional Analysis." The Radon-Nikodym derivative of a particular Gaussian probability measure with respect to another Gaussian measure plays a key role in the optimal estimation technique derived. Several special cases are considered in which the signal and noise processes are restricted to being "narrow-band." For these cases, an easily implemented technique is obtained for evaluating the required Radon-Nikodym derivative. Finally, a "narrow-band" numerical example is included along with an analysis of the error that results from the use of this estimation technique.

CHAPTER 1

INTRODUCTION

1.1 BRIEF STATEMENT OF THE PROBLEM

The problem to be considered here is concerned with the surveillance of a region of space in which multiple sources of electromagnetic radiation are present. These sources, which emit random signals and are embedded in a noisy environment, operate in an intermittent mode and have the possibility of varying their relative locations within the region of surveillance. The system, which is to perform the surveillance, receives information concerning the source locations in the form of a sequence of time signals of finite duration, each of which is the output of an element of an array of omnidirectional antennas. The solution which we desire, then, specifies the operations to be performed by a signal processing system in order that optimal, in some sense, estimates of the angular locations of the sources relative to the receiving array can be provided. These angular locations will, henceforth, be called the "angles of arrival" or "directions of arrival" of the signals at the site of the receiving array. In the literature, this problem is sometimes referred to as the "Direction Finding Problem." It includes, as a special case, the "Radar Problem" in which the angular location of a source, whose emitted waveform is known to the surveillance system, is required.

A direction finding system that is capable of operating according to the above specifications has an obvious military application for determining an enemy's strategy. For example, activity in certain geographical areas may indicate an elaborate preparation is underway; activity of a single source at sea may indicate the location of a submarine; and the resolution of a number of signal sources, synchronized to a common signal, may indicate a new air-defense system. To demonstrate a nonmilitary application, suppose the antenna array is replaced by an array of seismic sensors. It is then evident that the above signal processing system can be used, equally well, to determine locations of earth disturbances as well as to map the earth's strata for oil exploration.

1.1.1 History of Direction Finding

In the past, a variety of techniques, depending on the exact application, have been utilized for determining the direction of arrival of signals at a particular point in space. Almost all of these techniques, however, pursue a common method of attack which requires antennas with directional properties to be sequentially scanned through the region of surveillance. The direction of maximum or minimum received signal strength is then used as an indication of the true angle of arrival of the signal.

In the period around 1890, Hertz,[1] operating in the 200 Mc range, utilized cylindrical parabolic mirrors to focus and concentrate energy from a transmitter and hence produce an antenna system with directional properties. Directivity was also used by Marconi[2] before 1900 to increase the range of his transmission by the use of copper parabolic mirrors.

When radio communication, because of the ranges possible, progressed rapidly to the hf range around 1900, the mirror techniques were no longer practical and J. Zenneck,[3] at about that time, introduced simple "director" and "reflector" wires to increase the range and to separate interfering signals. A similar approach was pursued by S. G. Brown[4] who connected together two vertical antennas separated by a half wavelength. Stone,[5] in 1902, proposed to physically rotate this array to determine the direction of arrival of the wave.

The next major advance was due to Bellini and Tosi[6] who provided a second array at right angles to the first. Then, by means of quadrature coils placed outside of a single rotating coil, they were able to sequentially sample the signal being picked up by each antenna pair. This technique avoids the necessity for rotating the entire array. The device for sampling the antennas sequentially is called a "goniometer" and has its counterpart in many present day direction finding systems. Such a device is shown schematically in Fig. 1.1 where the moving coil is connected to a receiver.

An additional improvement was provided by Zenneck who placed an omnidirectional element in the center of the array and connected so as to add phasewise to the output of the moving coil of a Bellini-Tosi goniometer. This arrangement produced an antenna pattern which was a rotating cardioid rather than a figure-eight, thus eliminating the 180° ambiguity which existed with the former configuration.

World War I and the development of the vacuum-tube amplifier with its sensitivity stimulated refinement of the above method of direction finding. Of particular interest was the design of the Adcock array and numerous indicators, both mechanical and cathode-ray types. World War II produced the 8-element Adcock array, H collectors, and crossed loops, all utilizing the spinning goniometer technique. Since about 1950, the University of Illinois has been conducting experiments[7] with the Wullenweber facility which utilizes a goniometer and a 120-element circular array. Furthermore, the recent papers of W. J. Lindsay and D. S. Heim[8] and C. E. Lindahl and B. F. Barton[9] which are concerned with goniometer systems testify to the fact that these techniques are still of importance today.

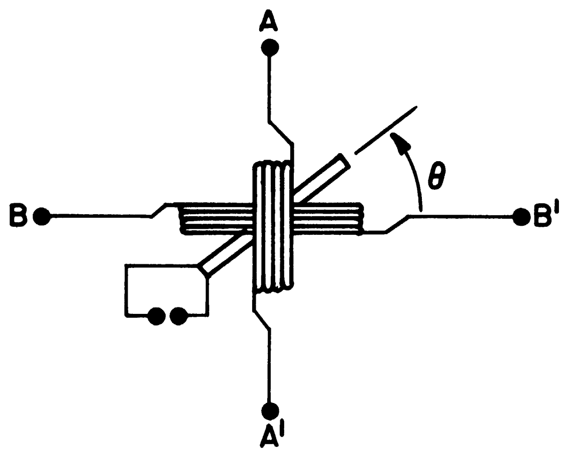

Fig. 1.1. The Bellini-Tosi goniometer for sequentially sampling a pair of the Bellini-Tosi array.

The methods described above employing spinning goniometers are not, however, the only practical techniques for determining the direction of arrival of a wave. World War I also saw the development of several rotating loop direction finding systems which applied the directivity of the loop to determine the direction of minimum received signal intensity. Various other techniques for obtaining directivity employing reflectors and lenses were also utilized during World War II.

Two recent developments have produced the possibility for a tremendous improvement in the art of direction finding. A new interest,[10] with emphasis on arrays of large numbers of elements, has provided the impetus for advances in our knowledge of the principles of arrays and the invention of new arraying techniques. These advances, together with the present availability of compact digital computers with their inherent signal processing capabilities, hold the promise of a vastly increased capacity for classifying and keeping under surveillance a large number of emitters. (These possibilities will be explored in the ensuing chapters of this thesis.)

For the sake of completeness, it is worth mentioning that still other techniques have been devised for solving the direction finding problem. In the situation where a limited number of sources whose signal strength at the site of the surveillance system is strong, time delay and phase comparison techniques are applicable.

1.1.2 Basic Elements of a Direction Finding System

Before examining any direction finding (sometimes abbreviated DF) system for its individual characteristics, it will be useful to consider the elements basic to any system and see how they can affect its performance. A block diagram of a generalized DF system is shown in Fig. 1.2 with a discussion of its individual blocks being presented in the

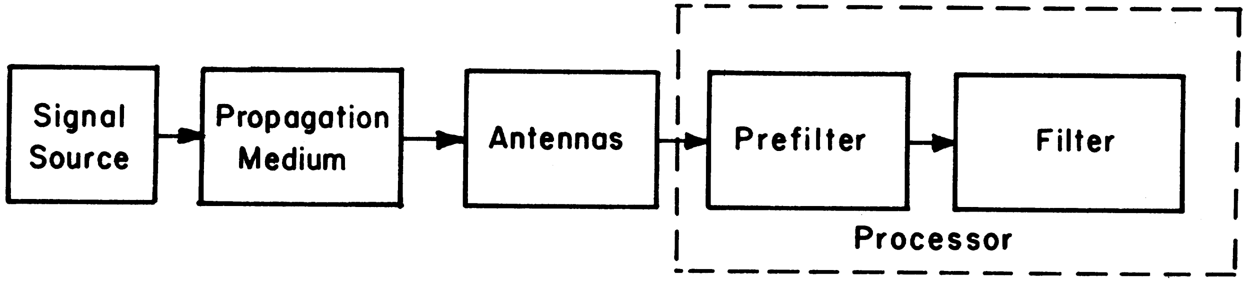

Fig. 1.2. A generalized direction finding system.

following six sections.

1.1.2.1 Signal Source.—The signal source has four properties which affect the performance of a direction finding system. These are its transmission frequency, its power level, its modulation characteristics, and its location relative to that of the DF system's antennas which intercept the emitted signals. The transmission frequency and modulation characteristics influence the choice of the antennas and receiving equipment employed in the system. The source power level together with its location which can be given by specifying the source—receiving antennas separation as well as its angular bearing influence the DF system only as they affect the signal strength of the emitted signal at the site of the receiving antennas.

1.1.2.2 Propagation Medium.—The propagation medium affects a direction finding system by attenuating the signal in varying degrees depending upon the frequency of the carrier, the time of day, the distance between the source and the receiving antennas and many other factors. In addition, it is also possible for the propagation media to introduce various other distortions such as random phase errors which are independent of the source being used. Finally, under certain conditions in long-range direction finding, multipath propagation usually cannot be avoided.

1.1.2.3 Antennas.—All direction finding systems require some device for sensing the electromagnetic waves that have been emitted by the sources. The exact type utilized is a function of the desired antenna pattern together with the carrier frequency and bandwidth of the emitted signals as well as various economic considerations. The type that will be of interest to us employs an array of omnidirectional antennas.

1.1.2.4 Signal Processor.—In most direction finding systems, the processor has two distinct functions to perform as illustrated in Fig.

1.2. These functions are discussed in the two following sections.

1.1.2.5 Prefilter.—The basic purpose of this unit is to combine the signals received by the antennas in a manner such that its output can be interpreted, by the following filter, as a direction of arrival of the signal. In systems employing arrays of omnidirectional antennas, the outputs from the individual antenna elements are usually combined so that the "array-prefilter" pair possesses a highly directional antenna pattern* and, as a result, the prefilter can be thought of as a "beam forming" network. Then, by varying this method of combination with time, the system can be caused to scan the region of surveillance. Systems employing dish antennas also permit an analysis in a manner similar to that of the above system of omnidirectional antennas. In this case, however, the prefilter is restricted by mechanical connections with scanning being performed by a physical rotation of the antenna. In any event, this prefilter is normally preset and independent of past received signals. In the systems to be considered in this presentation, the above restriction will be relaxed with the operations to be performed by the precessor being specified so as to optimize the performance of the total system.

1.1.2.6 Filter.—The filter takes the output of the signal prefilter and provides an estimate of the direction of arrival of the signal according to some established criterion of optimality. As an example of such a criteria, we could require the filter to provide the estimate which has the "maximum a posteriori probability" based on the received signals and any a priori information that might be available.

1.2 APPROACHES TO THE PROBLEM

In this dissertation, two different approaches for determining an

*See Refs. 11 and 12 for a discussion of phased arrays.

optimal estimate of the direction of arrival will be explored. Chapters 2-5 present a solution which attempts to apply the existing theory of "Stochastic Optimal Control" to this problem. In particular, control laws for the optimum beam forming network and the optimum filter are derived. In Chapters 6-9, an alternate approach is pursued in which the separation of the beam forming and filtering operations is absent. Instead, the operations to be performed by the optimal processor are specified by an equation, whose solution represents the optimal estimate. This equation, whose implementation with existing hardware is also considered, is derived by the application of several techniques from the "Theory of Estimation."

1.2.1 Stochastic Optimal Control Theory

Historically, the classical theory of control of deterministic, linear, time-invariant systems was developed during the 1930's and 1940's. Design of control systems according to this theory usually entailed an analysis of its frequency domain characteristics. During the 1940's and 1950's, the theory was extended to linear, time-invariant systems involving stationary random signals by use of Wiener's theory of prediction and filtering.

During the last 10 years, the fundamental direction of research in control theory has turned toward the modern theory of optimal control. This theory requires the determination of a control law which governs the action of a controllable variable so as to optimize (i.e., maximize or minimize) some performance or loss functional. By the use of the state-space techniques, the calculus of variations, dynamic programming, and Pontryagin's maximum principle, a very substantial and satisfactory theory now exists for deterministic systems.

Just as the classical theory was extended to include the effects of random disturbances, attempts have been made to extend modern control theory into this area. Indeed, some successful results have already been obtained, particularly in the case of linear, discrete-time systems.[13] The results which will most interest us are due to the Russian, A.A. Fel'dbaum,[14] who has established a general procedure for handling both linear and nonlinear discrete-time systems. Applying his techniques, solutions can be obtained which, in general, require the use of a digital computer in a manner similar to that required by many solutions employing dynamic programming. As with dynamic programming, however, these solutions demand computation rates which, in many cases, are unrealistic in light of the capabilities of present day computer systems. Other authors, E. P. Maslov,[15] V. P. Zhivoglyadov,[16] and R. L. Stratonovich[17] to name a few, have attempted to apply Fel'dbaum's "dual control" techniques to specific situations with varying degrees of success. Results[18] have also been obtained for the continuous-time optimal control problem with noisy observations, but, in the nonlinear plant case to which our particular problem reduces, the theory is still in a rather preliminary stage of development.

For additional background on the stochastic optimal control problem, see the paper by Wonham,[19] where further references are given.

1.2.2 The Theory of Estimation

"Estimation Theory" was established as a mathematical technique in 1806 with the first publication on "least squares" estimation by Legendre.[20] This concept is generally credited to Gauss,[21] however, since his method published in 1809 was derived from fundamental principles. At that time, it was used mainly by astronomers as a means of reducing observations to obtain the orbital parameters of minor planets and comets.

After the introduction of the least squares, the next major advance was the "method of moments" formulated by K. Pearson[22] around 1900. The main disadvantage of the method of moments is that estimates found with this technique are not the "best" possible from the viewpoint of "efficiency."[24]

Estimation theory was put on a firm foundation by R. A. Fisher with a series of fundamental papers.[23-25] Fisher demonstrated that the method of "maximum likelihood" was usually superior to the method of moments and that estimates derived by this technique could not be improved "essentially."

A renewed interest in this area was stimulated by the rapid development of communication theory. Similar but independent theories were developed by N. Wiener[26] and A. Kolomogorov[27] for separating desired signals from undesired noise signals. Wiener was interested in obtaining optimum estimates of signals by the use of linear, physically realizable electronic filters and assumed the observed signals where sample functions of "stochastic processes." A very complete presentation of the theory of stochastic processes which we shall reference from time-to-time is available in the books by M. Loève,[28] J. L. Doob,[29] and P. R. Halmos.[30]

The application of maximum likelihood techniques to the problem of optimum demodulation of communication signals was first considered by F. W. Lehan and R. J. Parks.[31] Further investigations of this problem were carried on by D. C. Youla[32] and D. Slepian.[33] Youla was concerned with amplitude demodulation while Slepian was interested in the estimation of a finite number of parameters of which the signal was assumed to be a function. J. B. Thomas and E. Wong[34] obtained the a posteriori most probable estimate of the modulation rather than the maximum likelihood estimate. Recently, H. L. Van Trees[35] has attempted to generalize the work of these authors to handle arbitrary modulation as well as diversity communication problems. These notes by Van Trees have stimulated much of the work that is presented in the last four chapters of this dissertation. However, rather than directly apply the techniques of Van Trees, a more fundamental approach has been pursued which extends the work of T. T. Kadota.[36-39] These papers by Kadota are concerned with deciding which of a finite number of Gaussian signals of differing covariance functions, in addition to Gaussian noise, is being received. Finally, there is also a great body of literature concerned with the application of maximum likelihood techniques to the analysis of radar systems with the papers by E. J. Kelley, I. S. Reed, and W. L. Root[40-41] being examples.

CHAPTER 2

SYSTEM MODEL FOR CONTROL THEORY SOLUTION

2.1 INTRODUCTION

In the next four chapters, a solution of the "Direction Finding Problem" by the use of "Stochastic Optimal Control Theory" will be sought. As an initial attempt at a solution, this approach appears to have merit in light of the fact that most existing systems have some control variables which affect the signals received. For example, with a dish-type antenna, the pointing angle affects the received signal strength. In the majority of these systems, however, the scanning modes and filtering techniques are independent of past received signals and often established without any real regard for each other. As a result, we might logically ask the following questions. "Is it possible for the control laws and filtering operations to be coordinated, depending on the past received information, so as to obtain a total system optimization?" "Can we, by the use of optimal control functions and possibly in conjunction with a digital computer, increase the performance of existing systems?" In the following chapters we shall attempt to answer these questions by obtaining an optimal "open-loop" solution and an optimal "closed-loop" solution with emphasis on both the control laws for orienting and shaping the antenna pattern and the filter operations which process the received data. For the closed-loop solution, the control law will be a function of past received signals, past controls, and any priori information that might be available concerning the true angle of arrival of the signals. In the open-loop solution, the control law will be independent of past received signals and past controls but still a function of the a priori information.

2.2 SYSTEM MODEL

The DF system which will be presented in this section for analysis in the next three chapters has been selected because of its simplicity and because it exhibits most of the features of existing direction systems which are of importance when considering their system optimization.

Before introducing this system model, however, let us first consider the direction finding system illustrated in Fig. 2.1. This system will be employed to provide us with some justification for our particular choice of system model. In Fig. 2.1, an incident wave at an angle α relative to some reference direction, $\mathcal{P}$, is impinging upon a group of K omnidirection antennas. This wave has been emitted by a source, sufficiently far removed, so that the wave appears to be a plane wave when viewed at the site of the receiving array. Furthermore, the source waveform is a pure sinusoidal of frequency ω_o while the source location is in a plane which contains all the elements of the antenna array so that the estimation of a single angle is required. (If and when a tractable solution is found for this type of signal waveform and this particular source-array geometry, we may then consider the more general (and difficult) case where the source is emitting a random signal and the estimate of two angles is required.) Recall,* now, a group of K onmidirectional antennas can be connected in such a manner that the group has the characteristics (antenna pattern) of a single directional antenna. By the introduction of proper delays and amplifications of the individual received signals before summing, the pointing angle and, to some extent, the beam width can be controlled. Applying this fact, in Fig. 2.1 the individual received signals are shown as inputs to the devices $B_j, j\varepsilon\{1,2,...,K\}$, which delay and amp-

*See Ref. 11, Vol. 2, Chapter 1.

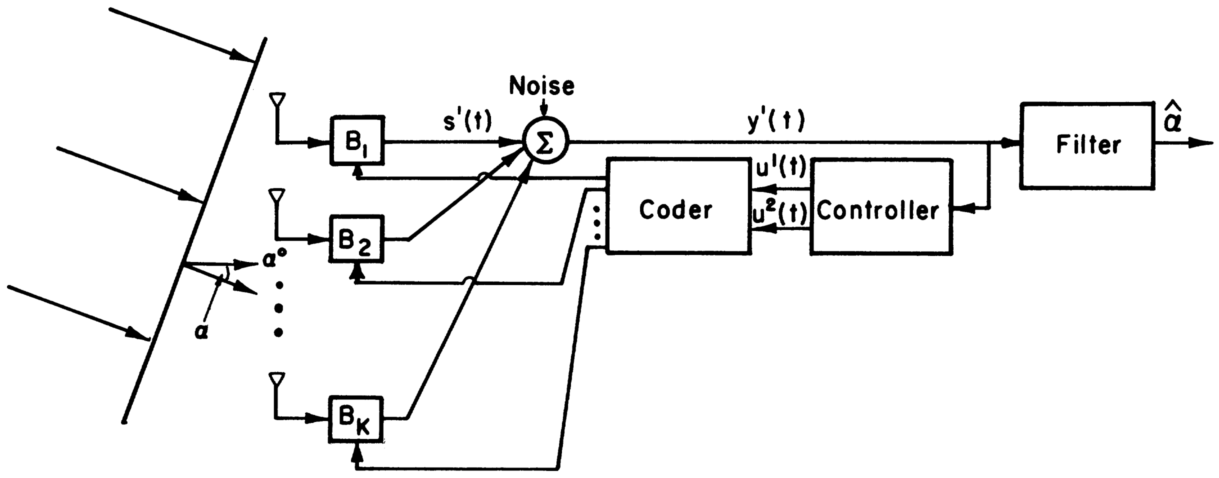

Fig. 2.1. A phased array DF system.

lify as required by the coder. The coder, in turn, specifies the delays and amplifications which are necessary to comply with the controls $u^1(t)$ and $u^2(t)$, the desired pointing angle and beam width, respectively, specified by the controller. Moreover, since we are interested in optimal DF systems, the controller operates so as to optimize the performance of the filter whose output $\hat{\alpha}$, is a "best" estimate of the angle of arrival of the signals based on the data $y'(t)$ received during the finite observation interval $[0,T_f]$. Finally, in this figure, an additive noise source is shown which is assumed to be independent of the controls.

Let us now consider a model which, as we shall see, possesses the essential features of the preceding direction finding system. Since the coder performs known operations, as far as the control-estimation problem is concerned, the model shown in Fig. 2.2 is equivalent to the system of Fig. 2.1. In this figure α, β, and θ are unknown parameters, α representing the unknown direction of arrival, β representing the unknown signal amplitude and θ the unknown phase angle. The signal $s'(t)$ is given by

$$s'(t) = C'(\alpha,\beta,u^1,u^2) \sin (\omega_o t+\theta)$$

where C' is the deterministic function of the variables α, β, u^1, and u^2 which summarizes the effects on the signal $s'(t)$ produced by the unknowns α and β, the antennas, the coder, the summer, and the devices $B_j, j\varepsilon\{1,2,\ldots,K\}$, of Fig. 2.1. (Note, if we eliminate the control variable u^2, this model could equally well serve as a model for a system which employs a dish-type antenna where the pointing angle u^1 is to be controlled and whose beam width is fixed.) The actual function $C'(\alpha,\beta,u^1,u^2)$ is linear in β so that it may be written as β times $C''(\alpha,u^1,u^2)$ where $C''(\alpha,u^1,u^2)$, as a function of the variables α, u^1, and u^2, is most easily obtained by experimental

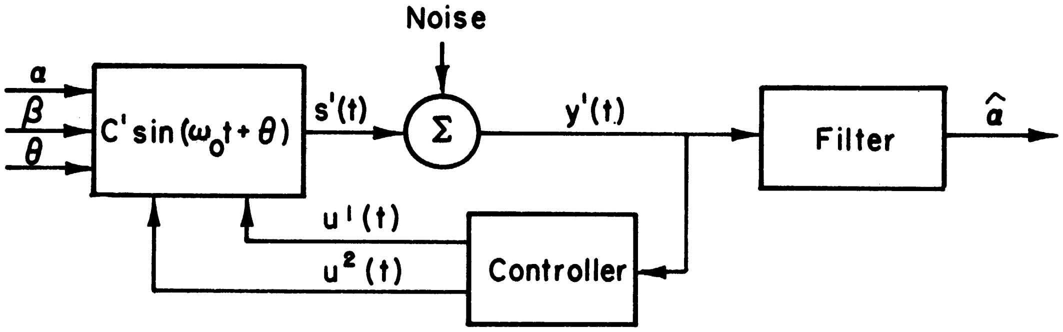

Fig. 2.2. Simplification of the DF system illustrated in Fig. 2.1.

techniques. However, an approximation for the function $C''(\alpha,u^1,u^2)$ (usually referred to as the amplitude "antenna pattern") which retains its essential properties is given by

$$C(\alpha,u^1,u^2) = \exp\{-u^2(\alpha-u^1)^2\} .$$

As with $C''(\alpha,u^1,u^2)$, this function has a maximum when u^1 equals α and a beam width dependent on u^2.

We are now in a position to present the system model which will be the subject of the ensuing analysis. This model is illustrated in Fig. 2.3 where it can be seen to have a structure similar to that given in Fig. 2.2. In this model, however, additional simplifications have been introduced by substituting βC for C' and assuming θ is known so $y'(t)$ can be synchronously detected to eliminate the sin $(\omega_o t+\theta)$ term which contains no information about the angle α. (A phase-locked loop can be employed to provide the variable θ.) The signal $y(t)$ is then given by

$$y(t) = \beta C(\alpha,u^1(t),u^2(t)) + \eta(t)$$

where $\eta(t)$ represents the noise in the system and is assumed to be independent of the controls. In addition, since we are interested in possibly utilizing a digital computer which demands discrete-time data and since such problems are usually easier to solve, the observation interval $[0,T_f]$ has been divided into J subintervals with the signal $y(t)$ being sampled after each subinterval to obtain the sequence $\{y_j\},j\varepsilon\{1,2,\ldots,J\}$. The controls $u^1(t)$ and $u^2(t)$ can then also be restricted to a discrete form with u_j^1 and u_j^2 being the respective controls applied during the jth subinterval. For convenience, moreover, we shall assume that the noise component of $y(t)$ at each sample time t_j, η_j, is an independent random

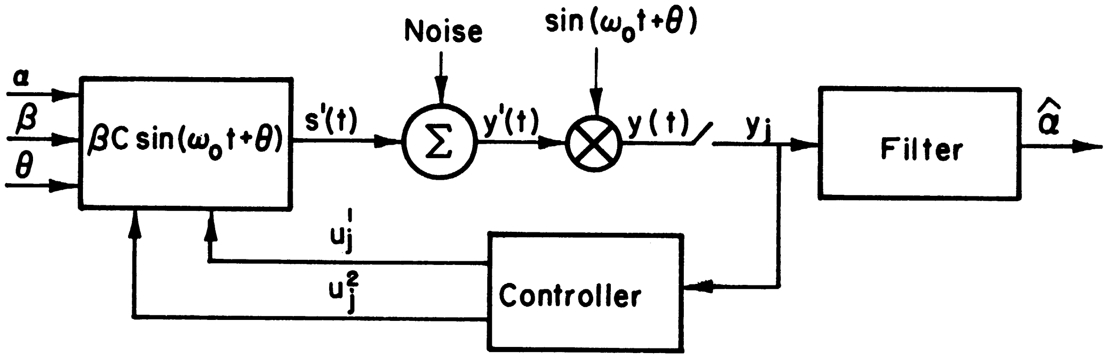

Fig. 2.3. System model to be analyzed in Chapters 3-5.

variable of zero mean and variance σ_η^2. Finally, in many problems it is reasonable to assume that the direction finding system has available some a priori information concerning the true values of the unknown parameters. For this model, we shall exhibit this information in the form of a known probability distribution of these parameters.

It is now possible to be more specific as to what we shall mean by a closed-loop and an open-loop solution for the direction finding problem using control theory techniques. A solution which specifies the operations to be performed by the controller and the filter for the model of Fig. 2.3 will be called a closed-loop solution while a solution for this model when the controller is required to operate without knowledge of the past samples of y will be called an open-loop solution.

To summarize, in hopes of obtaining a tractable solution, we have developed a very simplified model of a direction finding system while, at the same time, retaining many of the features important to the control-estimation problem. The significant point to remember is, if no solution can be found for this model, there appears to be little hope for solving the more realistic problem in this control-estimation framework.

CHAPTER 3

THE OPEN-LOOP SOLUTION

3.1 INTRODUCTION

Although we are interested primarily in the closed-loop solution, we shall begin with an investigation of the open-loop system from which insight (and experience in obtaining a solution) can be gained. The results of this chapter are presented in Section 3.3. A discussion of these results, however, will be deferred until Chapter 5 at which time the closed-loop results will also be discussed.

3.2 ASSUMPTIONS

The solution presented in Section 3.3 is optimal when the following assumptions together with the system model of Fig. 2.3 represent a valid description of the environment of the direction finding system.

A3.2.1

The control u^2 is fixed which implies the beam width does not change during the observation interval $[0,T_f]$. Also, the controls $u_j^1, j\varepsilon\{1,2,...,J\}$, are chosen a priori since we are considering the open-loop system.

A3.2.2

The unknowns α and β are independent random variables having means of m_α and m_β, respectively, and variances σ_α^2 and σ_β^2, respectively. In addition, the probability density of α is an even function about m_α.

A3.2.3

The probability density function of α, $p(\alpha)$, is "narrow" compared with the beam width so that, over the range of α where $p(\alpha)$ is different from zero, the function $C(\alpha,u_j^1,u^2)$ can be written as

$$C(\alpha,u_j^1,u^2) = p_j + q_j(\alpha-m_\alpha) + r_j(\alpha-m_\alpha)^2 + \varepsilon(\alpha,u_j^1,u^2) \quad (3.1)$$

where

$$p_j = \exp\{-u^2(m_\alpha-u_j^1)^2\} , \quad (3.2)$$

$$q_j = -2u^2(m_\alpha-u_j^1)\exp\{-u^2(m_\alpha-u_j^1)^2\} , \quad (3.3)$$

$$r_j = 2(u^2)^2(m_\alpha-u_j^1)^2\exp\{-u^2(m_\alpha-u_j^1)^2\} - u^2\exp\{-u^2(m_\alpha-u_j^1)^2\} , \quad (3.4)$$

and the absolute value of $\varepsilon(\alpha,u_j^1,u^2)$ is less than the real positive constant ε_o. (The first three terms on the right of Eq. (3.1) are, of course, the first three terms of the power series expansion of C about the point m_α when C is considered to be a function of α only. As a result, for any $\varepsilon_o > 0$, there is a width for the density $p(\alpha)$, below which, Eq. (3.1) is valid.)

A3.2.4

The class of filters from which an optimal filter is sought operates on the received signals to produce an estimate α_e according to the equation

$$\alpha_e = b_o + \sum_{j=1}^{J} b_j y_j \quad (3.5)$$

where y_j is the value of y at the jth sample time, b_o represents a bias of the sum $\sum_{j=1}^{J} b_j y_j$ and $\{b_j\}$ is a sequence of weights applied to the respective samples by the filter.

A3.2.4

The optimal controls $\hat{u}_j^1, j\varepsilon\{1,2,\ldots,J\}$, and the "weights" $\hat{b}_j, j\varepsilon\{1,2,\ldots,J\}$, together with the bias b_o which characterizes the optimal filter minimize the expected value of the squared error, i.e., they minimize $E\{(\alpha_e-\alpha)^2\}$.

According to Assumption A3.2.3, the solution that will be obtained in this chapter corresponds to the limiting case in which a great deal of information is available concerning the true value of α and only minor refinements are required. The restriction of the optimal filter to the class specified in Assumption A3.2.4 has been made because of the ease of implementation of such filters.

3.3 RESULTS

The following results were obtained when the assumptions of Section 3.2 were applied to the system model illustrated in Fig. 2.3. These results together with those derived by the closed-loop analysis of the next chapter will be discussed in Chapter 5.

R3.3.1

Suppose J, the total number of samples upon which our estimate is to be based, is an even number while

$$\underline{\tilde{u}}^1 = m_\alpha \underline{c}_o + (1/\sqrt{2u^2})\ \underline{c}\ , \tag{3.6}$$

$$\underline{\tilde{b}} = \frac{(1/\sqrt{J})q_{max}\sigma_\alpha^2 m_\beta}{\sigma_\alpha^2(\sigma_\beta+m_\beta^2)q_{max}^2+\sigma_\eta^2}\ \underline{c} \tag{3.7}$$

and

$$\tilde{b}_o = m_\alpha - m_\beta \sum_{j=1}^{J} \tilde{b}_j \tilde{p}_j \tag{3.8}$$

where $\underline{c}_o$ denotes the J-dimensional column vector whose elements are plus one, $\underline{c}$ denotes the J-dimensional column vector for which J/2 arbitrary elements are plus one and its other J/2 elements are minus one, $q_{max} = \sqrt{2Ju^2}$ $\exp\{-1/2\}$, $\tilde{b}_j$ is the jth element of $\underline{\tilde{b}}$, and $\tilde{p}_j$ is the value of p_j when the jth element of $\underline{\tilde{u}}^1$ is the control $\tilde{u}_j^1$. Then, as the beam width becomes large compared with the width of the probability density function $p(\alpha)$, the use of $\tilde{b}_o$ as the filter bias and the jth elements of $\underline{\tilde{u}}^1$ and $\underline{\tilde{b}}$ as the jth control and jth filter weight, respectively, results in an expected value of the squared error which approaches arbitrarily close to the expected value of the squared error resulting from the use of the optimal values of these quantities.

R.3.3.2

If the controls, the filter weights, and the bias specified in R3.3.1 are applied, the expected value of the squared error becomes

$$E\{(\alpha_e-\alpha)^2\} = \sigma_\alpha^2 \left\{ \frac{\sigma_\alpha^2\sigma_\beta^2 + \frac{\sigma_\eta^2}{q_{max}^2}}{\sigma_\alpha^2(\sigma_\beta^2+m_\beta^2) + \frac{\sigma_\eta^2}{q_{max}^2}} \right\} + \varepsilon' \qquad (3.9)$$

where ε' is a term which goes to zero as the beam width becomes large compared with the width of $p(\alpha)$.

It will be shown, the controls specified in R3.3.1 direct the antenna beam so that, for equal time intervals, each of the two angles at which the antenna pattern (the function $C(\alpha,u^1,u^2)$) has maximum slope as a function of α are oriented in the a priori expected direction m_α.

3.4 SOLUTION

Let us begin by defining $\underline{y}$ and $\underline{b}$ to be the J-dimensional column vectors whose jth components are given by y_j and b_j, respectively. Then, using a superscript T to denote the transpose of a vector, we find that Eq. (3.5) can be rewritten as

$$\alpha_e = b_o + \underline{b}^T\underline{y} \qquad (3.10)$$

where $\underline{b}^T\underline{y}$ is understood to represent the matrix product of $\underline{b}^T$ and $\underline{y}$. Our problem, then, according to the optimality criterion of Assumption A3.2.5, reduces to the determination of b_o, $\underline{b}$ and the control vector $\underline{u}^1$, $\underline{u}^1$ being the J-dimensional column vector whose jth element is u_j^1, that minimize the expected value of the squared error which results from the use of Eq. (3.10) to estimate α.

Forming $E\{(\alpha_e-\alpha)^2\}$, we find

$$E\{(\alpha_e-\alpha)^2 = E\{(\underline{b}^T\underline{y}-\alpha)^2\} - [E\{(\underline{b}^T\underline{y}-\alpha)\}]^2 + [b_o-E\{(\underline{b}^T\underline{y}-\alpha)\}]^2. \quad (3.11)$$

Then, since b_o appears only in the last term of Eq. (3.11), it is easily seen that

$$b_o = \hat{b}_o = E\{(\underline{b}^T\underline{y}-\alpha)\} \quad (3.12)$$

minimizes $E\{(\alpha_e-\alpha)^2\}$ with respect to b_o and, therefore, $\hat{b}_o$ as defined in Eq. (3.12) is the optimal bias. Moreover, when $b_o = \hat{b}_o$, the expected value of the squared error reduces to

$$E\{(\alpha_e-\alpha)^2\} = E\{(\underline{b}^T\underline{y}-\alpha)^2\} - [E\{(\underline{b}^T\underline{y}-\alpha)\}]^2 . \quad (3.13)$$

In order to obtain the optimal "weights" $\{b_j\}$, let us first simplify Eq. (3.13) by squaring the indicated quantities and noting that only the random vector $\underline{y}$ and the random variable α are affected by the expectation. Then, applying Assumption A3.2.2, Eq. (3.13) becomes

$$E\{(\alpha_e-\alpha)^2\} = \sigma_\alpha^2 + \underline{b}^T D\underline{b} - 2\underline{b}^T\underline{d} \quad (3.14)$$

where

$$D = E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^T\} \quad (3.15)$$

which is the covariance function of the random vector $\underline{y}$ and

$$\underline{d} = E\{(\alpha-m_\alpha)\ \underline{y}\} . \quad (3.16)$$

We now have the expected value of the squared error in a form, as given in Eq. (3.14), which will allow us to obtain the optimal "weights."

Taking the gradient of Eq. (3.14) with respect to $\underline{b}$ and setting it equal to zero, we find the vector of optimal "weights" $\underline{b}$ must satisfy

$$2D\hat{\underline{b}} - 2\underline{d} = 0$$

or equivalently,

$$\hat{\underline{b}} = D^{-1}\underline{d} \tag{3.17}$$

where D^{-1} denotes the inverse of the matrix D which we shall assume exists, at least for the time being. The jth component of $\underline{b}$ is then the weight applied to the jth sample by the optimal filter. Finally, substituting $\hat{\underline{b}}$ for $\underline{b}$ in Eq. (3.14), $E\{(\alpha_e-\alpha)^2\}$ becomes

$$E\{(\alpha_e-\alpha)^2\} = \sigma_\alpha^2 - \underline{d}^T D^{-1}\underline{d} \tag{3.18}$$

which now must be minimized with respect to the control vector $\underline{u}^1$.

Applying Assumptions A3.2.2 and A3.2.3 and denoting the J-dimensional column vectors whose jth elements are p_j, q_j, and r_j, by $\underline{p}$, $\underline{q}$, and $\underline{r}$, respectively, the matrix D becomes

$$D = \sigma_\beta^2\underline{p}\underline{p}^T + (\sigma_\beta^2+m_\beta^2)\sigma_\alpha^2\underline{q}\underline{q}^T + [(\sigma_\beta^2+m_\beta^2)\gamma_\alpha^4 - m_\beta^2(\sigma_\alpha^2)^2]\underline{r}\underline{r}^T$$

$$+\sigma_\alpha^2\sigma_\beta^2 \ [\underline{p}\underline{r}^T + \underline{r}\underline{p}^T] + \sigma_n^2 I + D_e(\underline{p},\underline{q},\underline{r}) \tag{3.19}$$

where $\gamma_\alpha^4 = E\{(\alpha-m_\alpha)^4\}$, I is the identity matrix, and $D_e(\underline{p},\underline{q},\underline{r})$ is a matrix whose elements go to zero as ε_0 goes to zero and, hence, as the beam width becomes large compared with the width of the density function $p(\alpha)$. (Note, to obtain D, we have used the fact that $E\{(\alpha-m_\alpha)^3\} = 0$ which follows from Assumption A3.2.2 where $p(\alpha)$ was assumed to be an even function about m_α.) The vector $\underline{d}$ on the other hand, reduces to

$$d = m_\beta \sigma_\alpha^2 \underline{q} + \underline{d}_\varepsilon \tag{3.20}$$

where $\underline{d}_\varepsilon$ is a J-dimensional column vector whose elements also go to zero with ε_o. It can now be seen, even if D_e and $\underline{d}_e$ are neglected in Eqs. (3.19) and (3.20), the minimization of Eq. (3.18) with respect to the control vector $\underline{u}^1$ is a problem which is not readily solvable since the functional relationships between $\underline{p}$, $\underline{q}$, $\underline{r}$, and the control vector $\underline{u}^1$ are non-trivial. It appears that only a computer solution can produce the desired minimizing control vector $\underline{u}^1$. Thus, we find that the method of solution which calls for the successive minimization of $E\{(\alpha_e-\alpha)^2\}$ with respect to b_o, $\underline{b}$ and $\underline{u}^1$, although being relatively straight forward as far as the b_o and $\underline{b}$ minimizations are concerned, becomes analytically unmanageable when the $\underline{u}^1$ minimization is considered.

Let us now back up to Eq. (3.13) and take a slightly different approach. Rather than derive the optimal vector $\underline{b}$ at this point, let us write Eq. (3.13) in the equivalent form

$$E\{(\alpha_e-\alpha)^2\} = E\{[(\underline{b}^T\underline{y}-\alpha) - E\{(\underline{b}^T\underline{y}-\alpha)\}]^2\} \tag{3.21}$$

and apply Assumptions A3.2.2 and A3.2.3 which, after some rearrangement of terms, gives us

$$\begin{aligned} E\{(\alpha_e-\alpha)^2\} = {} & \sigma_\alpha^2((\sigma_\beta^2+m_\beta^2)^{1/2}\underline{b}^T\underline{q} - m_\beta(\sigma_\beta^2+m_\beta^2)^{-1/2})^2 + \sigma_\eta^2\underline{b}^T\underline{b} \\ & + E\{[(\beta-m_\beta)\underline{b}^T\underline{p} + (\beta(\alpha-m_\alpha)^2)-m_\beta\sigma_\alpha^2)\underline{b}^T\underline{r}]^2\} + \frac{\sigma_\alpha^2\sigma_\beta^2}{(\sigma_\beta^2+m_\beta^2)} + \varepsilon_1 \end{aligned} \tag{3.22}$$

where ε_1 is a term which goes to zero with ε_o. Then, since the fourth term of Eq. (3.22) is independent of $\underline{b}$ and $\underline{u}^1$, the use of the J-dimensional column vector of filter "weights" $\tilde{\underline{b}}$ and the J-dimensional column vector of controls $\underline{\tilde{u}}^1$ which minimize the sum of the first <u>three</u> terms of Eq. (3.22) result in a loss, $E\{(\alpha_e-\alpha)^2\}$, which approaches arbitrarily close to the loss resulting from the use of the optimal "weights" and optimal controls as ε_1 goes to zero. But this implies, the performance of a system, as measured by $E\{(\alpha_e-\alpha)^2\}$, which employs $\underline{\tilde{b}}$ and $\underline{\tilde{u}}^1$ instead of $\underline{\hat{b}}$ $\underline{\hat{u}}^1$ approaches the optimal performance as the beam width becomes large compared with the width of $p(\alpha)$.

To obtain $\underline{\tilde{b}}$ and $\underline{\tilde{u}}^1$, note that the vector of "weights" $\underline{\tilde{b}}_1$ and the vector of controls $\underline{\tilde{u}}_1^1$ which minimize the sum of the first <u>two</u> terms of Eq. (3.22) must be such that $\underline{\tilde{b}}_1$ and $\underline{\tilde{q}}_1$ (the vector $\underline{q}$ when $\underline{\tilde{u}}_1^1$ is the control vector) are colinear while $\underline{\tilde{q}}_1$ must be of maximal length, i.e., $\underline{\tilde{q}}_1^T\underline{\tilde{q}}_1$ must achieve the maximum value of $\underline{q}^T\underline{q}$. If $\underline{\tilde{b}}_1$ and $\underline{\tilde{q}}_1$ did not satisfy these conditions, we could always find another pair, call them $\underline{\tilde{b}}_2$ and $\underline{\tilde{q}}_2$ for which $\underline{\tilde{b}}_2^T\underline{\tilde{q}}_2 = \underline{\tilde{b}}_1^T\underline{\tilde{q}}_1$ and $\underline{\tilde{b}}_2^T\underline{\tilde{b}}_2 < \underline{\tilde{b}}_1^T\underline{\tilde{b}}_1$ and, as a result, the sum of first two terms of Eq. (3.22) could always be decreased. Moreover, if $\underline{\tilde{b}}_1$ and $\underline{\tilde{u}}_1^1$ are such that $\underline{\tilde{b}}_1^T\underline{\tilde{p}}_1$ and $\underline{\tilde{b}}_1^T\underline{\tilde{r}}_1$ are equal to zero, $\underline{\tilde{b}}_1$ and $\underline{\tilde{u}}_1^1$ will also minimize the sum of the first three terms of Eq. (3.22) which implies these vectors are suitable choices for $\underline{\tilde{b}}$ and $\underline{\tilde{u}}^1$, respectively. (The vectors $\underline{\tilde{p}}_1$ and $\underline{\tilde{r}}_1$ are vectors $\underline{p}$ and $\underline{r}$ when $\underline{\tilde{u}}_1^1$ is the control vector.)

Looking now at Fig. 3.1 where q_j is plotted, we see that $\underline{q}^T\underline{q}$ is maximum when $u_j^1, j\varepsilon\{1,2,\ldots,J\}$, satisfy

$$(m_\alpha - u_j^1) = \pm 1/\sqrt{2u^2} \qquad (3.23)$$

or, equivalently,

$$u_j^1 = m_\alpha \mp 1/\sqrt{2u^2} .$$

As a result $\underline{\tilde{q}}_1$, must be of the form

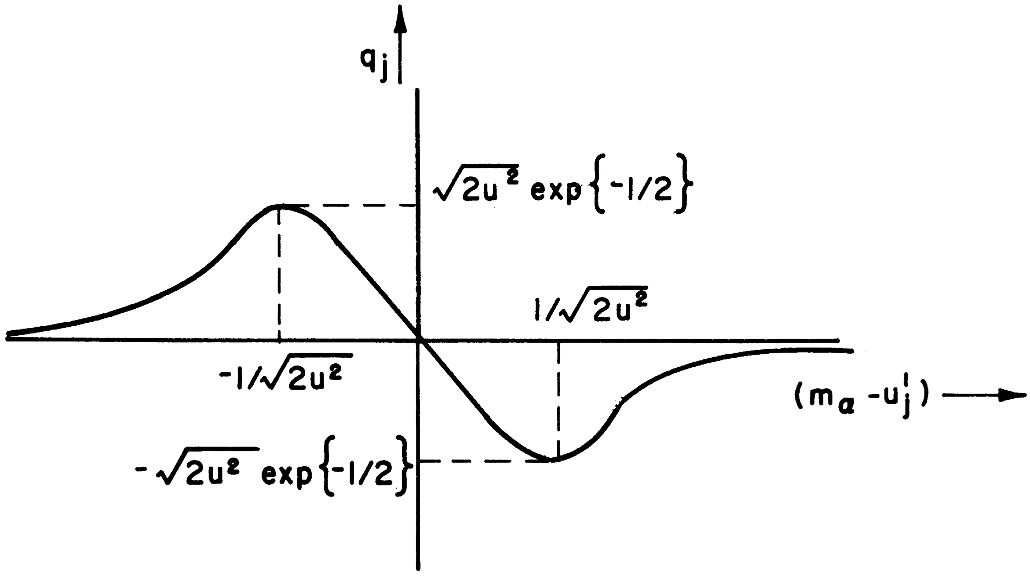

Fig. 3.1. Plot of q_j vs. $(m_\alpha - u_j^1)$.

$$\tilde{\underline{q}}_1 = \sqrt{2u^2}\exp\{-1/2\}\underline{c}' \tag{3.24}$$

where $\underline{c}'$ is a J-dimensional column vector whose elements are either plus or minus one. Furthermore, if $\tilde{\underline{q}}_1$ has this form,

$$\tilde{\underline{q}}_1{}^T\tilde{\underline{r}}_1 = 0 \tag{3.25}$$

and

$$\tilde{\underline{q}}_1{}^T\tilde{\underline{p}}_1 = \sqrt{2u^2}\exp\{-1\}\sum_{j=1}^{J} c'_j \tag{3.26}$$

where c'_j is the jth element of $\underline{c}'$. Then, if J is an even number and half of the c'_j's are plus one while the other half are minus one, $\tilde{\underline{q}}_1{}^T\tilde{\underline{p}}_1$ is also zero. But this together with Eq. (3.25) implies $\tilde{\underline{b}}_1{}^T\tilde{\underline{p}}_1$ and $\tilde{\underline{b}}_1{}^T\tilde{\underline{r}}_1$ are zero since $\tilde{\underline{b}}_1$ and $\tilde{\underline{q}}_1$ are collinear. Thus, if $\underline{c}'=\underline{c}$, the control vector and the filter "weights" vector which minimize the sum of the first two terms in Eq. (3.18) also minimize the sum of the first three terms in Eq. (3.18) so that $\tilde{\underline{b}} = \tilde{\underline{b}}_1$ and $\tilde{\underline{u}}^1$ is given by

$$\tilde{\underline{u}}^1 = m_\alpha\underline{c}_o + (1/\sqrt{2u^2})\,\underline{c}\ . \tag{3.27}$$

(The vectors $\underline{c}_o$ and $\underline{c}$ are defined in Section 3.3.)

Let us now determine the vector $\tilde{\underline{b}}_1$. Since $\tilde{\underline{b}}_1$ and $\tilde{\underline{q}}_1$ are collinear, we can write

$$\tilde{\underline{b}}_1 = a\tilde{\underline{q}}_1 \tag{3.28}$$

where a is a real number. Then, the first two terms of Eq. (3.22) can be written as

$$\sigma_\alpha^2(a(\sigma_\beta^2+m_\beta^2)^{1/2}\tilde{\underline{q}}_1{}^T\tilde{\underline{q}}_1 - m_\beta(\sigma_\beta^2+m_\beta^2)^{-1/2})^2 + a^2\sigma_\eta^2\tilde{\underline{q}}_1{}^T\tilde{\underline{q}}_1 \tag{3.29}$$

which is easily shown to be minimized when

$$a = \frac{\sigma_\alpha^2 m_\beta}{\sigma_\alpha^2(\sigma_\beta^2+m_\beta^2)q_{max}^2+\sigma_\eta^2} \tag{3.30}$$

and $q_{max}^2 = \underline{q}_1^T \underline{q}_1 = 2Ju^2\exp\{-1\}$. Finally, substituting Eq. (3.30) and Eq. (3.24) with $\underline{c}' = \underline{c}$ into Eq. (3.28), we find

$$\tilde{\underline{b}}_1 = \frac{(1/\sqrt{J})q_{max}^2\sigma_\alpha^2 m_\beta}{\sigma_\alpha^2(\sigma_\beta^2+m_\beta^2)q_{max}^2+\sigma_\eta^2}\,\underline{c} \tag{3.31}$$

which, together with Eq. (3.27), reduces Eq. (3.22) to

$$E\{(\alpha_e-\alpha)^2\} = \sigma_\alpha^2\left\{\frac{\sigma_\alpha^2\sigma_\beta^2+\dfrac{\sigma_\eta^2}{q_{max}^2}}{\sigma_\alpha^2(\sigma_\beta^2+m_\beta^2)+\dfrac{\sigma_\eta^2}{q_{max}^2}}\right\}+\varepsilon_1 . \tag{3.32}$$

Finally, let us now investigate a suboptimal bias that is a function of known quantities. In particular, if we return to Eq. (3.11) and substitute

$$\tilde{b}_o = m_\alpha - \tilde{\underline{b}}_1^T\tilde{\underline{p}}_1 \tag{3.33}$$

for b_o instead of $\hat{b}_o$, $E\{(\alpha_e-\alpha)^2\}$ will increase by ε_2 which also goes to zero with ε_o. Then, if we let $\varepsilon' = \varepsilon_1+\varepsilon_2$, the remainder of the results presented in Section 3.3 follow.

Several attempts were made to provide, under less stringent conditions than above, the control vector, the filter "weight" vector, and the filter bias whose application results in a loss which approaches that obtained by the use of the optimal forms of these vectors. In particular, additional terms were added to the power series expansion in Assumption A3.2.3 with a search similar to that pursued above being conducted for approximations to the optimal control vector, the optimal filter "weights" vector, and the optimal bias. A solution was not found, however, and it appears that only a computer solution of the type discussed in the paragraph containing Eq. (3.20) will produce a solution. Moreover, since we are really interested in the closed-loop solution, a computer solution was not attempted and the investigation of the open-loop system was terminated at this point. The solution presented in this chapter does, however, bring out some interesting points which are discussed in Chapter 5.

CHAPTER 4

THE CLOSED-LOOP SOLUTION

4.1 INTRODUCTION

In this chapter, optimal filter and control functions for the closed-loop system illustrated in Fig. 2.3 are obtained. This is accomplished by making use of techniques which were developed by A. A. Fel'dbaum[14] and which resemble the dynamic programming approach of R. Bellman.[42]

4.2 ASSUMPTIONS

The system model discussed in Chapter 2 and illustrated in Fig. 2.3 together with the following assumptions specify the environment in which the solution presented in this chapter is optimal.

A4.2.1

The unknowns α and β are random variables with a joint probability density function given by $p(\alpha,\beta)$.

A4.2.2

The control u^2 is fixed during the interval $[0,T_f]$, i.e., we are again considering the case where the beam width does not vary.

A4.2.3

A control function (a function which specifies the operation of the controller) and a filter function (a function which specifies the operation of the filter) will be considered optimal when they jointly result in a system which possesses a minimum expected value of the squared error, i.e., if a system employes these functions, the expected value of the error will be minimized.

4.3 RESULTS

The following results were obtained when the assumptions of Section 4.2 were applied to the system model illustrated in Fig. 2.3. As stated in the previous chapter, these results together with those obtained in Chapter 3 will be discussed in Chapter 5.

R4.3.1

The optimal filter operates on the received samples $\{y_i\}$ according to the equation

$$\hat{\alpha} = \frac{\int_{\Omega(\alpha,\beta)} \alpha p(\alpha,\beta) \prod_{j=1}^{J} p(y_j/u_j^1,\alpha,\beta)d\alpha d\beta}{\int_{\Omega(\alpha,\beta)} p(\alpha,\beta) \prod_{j=1}^{J} p(y_j,u_j^1,\alpha,\beta)d\alpha d\beta} \tag{4.1}$$

i.e., the estimate $\hat{\alpha}$ which is provided by an optimal filter satisfies Eq. (4.1). In this equation, $\Omega(\alpha,\beta)$ denotes the region of the variation of the variables α and β over which they must be integrated while the function $p(y_j/u_j^1,\alpha,\beta)$ is given by

$$p(y_j/u_j^1,\alpha,\beta) = p_o(y_j-\beta C(\alpha,u_j^1,u_j^2)) \tag{4.2}$$

when $p_o(\eta_j)$ is the probability density function of the noise random variable η_j. (The function $\beta C(\alpha,u_j^1,u^2)$ was discussed in Chapter 2 and represents the value y_j would achieve in the absense of noise when α,β,u_j^1, and u^2 are the angle of arrival, the signal level, the pointing angle, and the beam width, respectively.) If, for example, the variables $\{\eta_j\}$ are Gaussian, $p(y_j/u_j^1,\alpha,\beta)$ takes the form

$$p(y_j/u_j^1,\alpha,\beta) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left\{-\frac{1}{2\sigma_\eta^2} [y_j-\beta C(\alpha,u_j^1,u^2)]^2\right\}$$

where σ_η^2 is the variance of η_j.

R4.3.2

The optimal control $\hat{u}_j$ applied during the jth subinterval of the observation interval $[0,T_f]$ minimizes the function

$$G_j(u_j^1) = \int_{\Omega(y_j)} \left\{ \min_{u_{j+1}^1} \int_{\Omega(y_{j+1})} \cdots \left\{ \min_{u_J^1} \int_{\Omega(y_J)} \left[\int_{\Omega(\alpha,\beta)} (\alpha-\hat{\alpha})^2 \right.\right.\right.$$

$$\left.\left.\left. (x)\ p(\alpha,\beta) \prod_{j=1}^{J} p(y_j/u_j^1,\alpha,\beta)\,d\alpha d\beta \right] dy_J \right\} \ldots dy_{j+1} \right\} dy_j \qquad (4.3)$$

which is, in turn, a succession of minimizations with respect to the future controls and integrations with respect to future samples $y_i, i\varepsilon\{j,j+1,\ldots,J\}$, as well as the variables α and β over their respective regions of variation. (The variable $\hat{\alpha}$ in this equation has the functional form illustrated by Eq. (4.1.).)

We shall find in the next chapter, the evaluation of Eq. (4.1) and the minimization of Eq. (4.3) with respect to u_j^1 requires the use of numerical integration which, in turn, necessitates the need for a digital computer.

4.4 SOLUTION

Let us begin by denoting with L, the loss function which equals the expected value of the squared error. We can then write

$$L = \int_{\Omega(\alpha,\alpha_e)} (\alpha-\alpha_e)^2 p(\alpha,\alpha_e)\,d\Omega(\alpha,\alpha_e) \qquad (4.4)$$

where α_e represents the output of the filter, $\Omega(\alpha,\alpha_e)$ is the region of variation of the variables α and α_e, $d\Omega(\alpha,\alpha_e)$ is an infinitely small element of $\Omega(\alpha,\alpha_e)$ which will also be written as $d\alpha d\alpha_e$, and $p(\alpha,\alpha_e)$ is the joint probability density of α and α_e. (Throughout this chapter, we shall use $p(x_1,x_2,\ldots,x_n)$ to denote the joint probability density of the arbitrary variables $x_1,x_2,\ldots,x_n$ listed in parentheses.) Therefore, to satisfy the optimality criterion of Assumption A4.2.3, a filter function and control function must jointly minimize this loss function.

To simplify the notion in the following discussion, we shall now discard the superscript on the control u^1. No ambiguity will result since u^2 is assumed fixed and will not enter the derivation. Then, applying the fact* that the integral of $p(\alpha,\alpha_e,\beta,y_{\bar{J}},u_{\bar{J}})$ over the region of variation of $\beta,y_{\bar{J}}$ and $u_{\bar{J}}$ is equal to $p(\alpha,\alpha_e)$, i.e.,

$$\int_{\Omega(\beta,y_{\bar{J}},u_{\bar{J}})} p(\alpha,\alpha_e,\beta,y_{\bar{J}},u_{\bar{J}})\,d\beta\,dy_{\bar{J}}\,du_{\bar{J}} = p(\alpha,\alpha_e) , \qquad (4.5)$$

we can write the loss function L of Eq. (4.4) in the alternate form

$$L = \int_{\Omega(\alpha,\alpha_e,\beta,y_{\bar{J}},u_{\bar{J}})} (\alpha-\alpha_e)^2 p(\alpha,\alpha_e,\beta,y_{\bar{J}},u_{\bar{J}})\,d\alpha\,d\alpha_e\,d\beta\,dy_{\bar{J}}\,du_{\bar{J}} \qquad (4.6)$$

where $y_{\bar{J}} = (y_1,y_2,\ldots,y_J)$, $u_{\bar{J}} = (u_1,u_2,\ldots,u_J)$, $dy_{\bar{J}} = dy_1 dy_2 \ldots dy_J$ and $du_{\bar{J}} = du_1 du_2 \ldots du_J$. (Remember y_j is the jth sample of the received signal y.) Moreover, if we make use of the relation*

$$p(x_1/x_2) = \frac{p(x_1,x_2)}{p(x_2)}$$

which defines the conditional probability density of x_1 given x_2, we can write

*See Ref. 43, pp. 29-30.

$$p(\alpha,\alpha_e,\beta,y_{\bar{J}},u_{\bar{J}}) = p(\alpha,\beta)p(\alpha_e/y_{\bar{J}},u_{\bar{J}},\alpha,\beta)p(y_{\bar{J}},u_{\bar{J}}/\alpha,\beta) \tag{4.7}$$

and

$$p(y_{\bar{J}},u_{\bar{J}}/\alpha,\beta) = p(u_J/y_{\overline{J-1}},u_{\overline{J-1}},\alpha,\beta)p(y_J/y_{\overline{J-1}},u_{\bar{J}},\alpha,\beta)p(y_{\overline{J-1}},u_{\overline{J-1}}/\alpha,\beta) \ . \tag{4.8}$$

In this last equation, the second term on the right can also be written as

$$p(y_J/y_{\overline{J-1}},u_{\bar{J}},\alpha,\beta) = p(y_J/u_J,\alpha,\beta) \tag{4.9}$$

since the noise random variables $\{\eta_j\}$ are assumed to be independent and since the function βC in our model of Fig. 2.3 has no memory which implies the probability of y_J given $y_{\overline{J-1}}$, $u_{\bar{J}}$, α, and β is just the probability of y_J given u_J, α, and β alone.

Let us now assume the filter function exhibits a random strategy, i.e., the filter generates an estimate α_e based on the data $y_{\bar{J}}$ and $u_{\bar{J}}$ according to some probability density function which we shall denote by $\Gamma(\alpha_e/y_{\bar{J}},u_{\bar{J}})$. (By allowing a random strategy for the filter function rather than requiring it to be a deterministic function, we have increased the number of elements in the class of possible filters from which an optimal element is sought. We shall find, however, the optimal element in this class exhibits a deterministic strategy rather than a random strategy as might be expect.) It is then obvious that the conditional probability density $p(\alpha_e/y_{\bar{J}},u_{\bar{J}},\alpha,\beta)$ on the right of Eq. (4.7) can be written as

$$p(\alpha_e/y_{\bar{J}},u_{\bar{J}},\alpha,\beta) = p(\alpha_e/y_{\bar{J}},u_{\bar{J}}) = \Gamma(\alpha_e/y_{\bar{J}},u_{\bar{J}}) \tag{4.10}$$

since the output of the filter depends only on the values of $y_{\overline{J}}$ and $u_{\overline{J}}$. If we also assume the control function for the Jth subinterval exhibits a random strategy, denote it $\Gamma_J(u_J/y_{\overline{J-1}},u_{\overline{J-1}})$, we find the first term on the right of Eq. (4.8) can similarly be written as

$$p(u_J/y_{\overline{J-1}},u_{\overline{J-1}},\alpha,\beta) = p(u_J/y_{\overline{J-1}},u_{\overline{J-1}}) = \Gamma_J(u_J/y_{\overline{J-1}},u_{\overline{J-1}}) \tag{4.11}$$

since the control during the Jth subinterval depends only on $y_{\overline{J-1}}$, and $u_{\overline{J-1}}$. Then, substituting (4.9) and (4.11) into (4.8) and proceedings by induction, we obtain

$$p(y_{\overline{J}},u_{\overline{J}}/\alpha,\beta) = \prod_{j=1}^{J} \Gamma_j p(y_j/u_j,\alpha,\beta) \tag{4.12}$$

where $\Gamma_j = p(u_j/y_{\overline{j-1}},u_{\overline{j-1}})$ is the control function for the jth subinterval and $\Gamma_1 = p(u_1)$.

If we now defined the function $G(\alpha_e)$ by

$$G(\alpha_e) = \int_{\Omega(\alpha,\beta)} (\alpha-\alpha_e)^2 p(\alpha,\beta) \prod_{j=1}^{J} p(y_j/u_j,\alpha,\beta)\,d\alpha d\beta \tag{4.13}$$

and substitute (4.7), (4.10), (4.12), and (4.13) into (4.6), the loss function becomes

$$L = \int_{\Omega(y_{\overline{J}},u_{\overline{J}})} \left\{ \int_{\Omega(\alpha_e)} G(\alpha_e)\Gamma d\alpha_e \right\} \prod_{j=1}^{J} \Gamma_j dy_{\overline{J}} du_{\overline{J}} . \tag{4.14}$$

Moreover, since the filter makes its estimate after the Jth subinterval, for purposes of determining the optimal filter function we can assume the Γ_j's are known. As a result, it can be seen from Eq. (4.14) that L is minimized if

$$\int_{\Omega(\alpha_e)} G(\alpha_e)\Gamma d\alpha_e \tag{4.15}$$

achieves a minimum value for each possible combination of the vectors $y_{\overline{J}}$ and $u_{\overline{J}}$. It can also be seen that (4.15) achieves this minimum value if

$$\Gamma = \delta(\alpha_e - \hat{\alpha}) \tag{4.16}$$

where δ is the Dirac delta function and $\hat{\alpha}$ minimizes $G(\alpha_e)$ for any set of vectors $y_{\overline{J}}$ and $u_{\overline{J}}$. Therefore, minimizing $G(\alpha_e)$ with respect to α_e by taking the partial derivative of $G(\alpha_e)$ with respect to α_e and setting it equal to zero, we obtain

$$\hat{\alpha} = \frac{\int_{\Omega(\alpha,\beta)} \alpha p(\alpha,\beta) \prod_{j=1}^{J} p(y_j/u_j,\alpha,\beta)\, d\alpha d\beta}{\int_{\Omega(\alpha,\beta)} p(\alpha,\beta) \prod_{j=1}^{J} p(y_j/u_j,\alpha,\beta)\, d\alpha d\beta} \tag{4.17}$$

which, together with (4.16), represents the optimal filter function. (The second derivative of $G(\alpha_e)$ with respect to α_e is easily found to be positive for all α_e so that $\hat{\alpha}$ as defined in Eq. (4.17) does, indeed, minimize $G(\alpha_e)$.) Note that this filter is optimal regardless of the control mode, so, when no restriction to a particular class of filters is made, it is also an optimal open-loop filter.

Turning now to the problem of obtaining the optimal control function, let us define the function $G_J(u_J)$ by

$$G_J(u_J) = \int_{\Omega(y_J)} G(\hat{\alpha})\, dy_J \,. \tag{4.18}$$

Then, substituting (4.18) into (4.14), the loss function L becomes

$$L = \int_{\Omega(y_{\overline{J-1}},u_{\overline{J-1}})} \left\{ \int_{\Omega(u_J)} G_J(u_J)\Gamma_J du_J \right\} \prod_{j=1}^{J-1} \Gamma_j dy_{\overline{J-1}} du_{\overline{J-1}} . \tag{4.19}$$

It can now be seen, as in the previous paragraph, L is minimized when the term within the bracket is a minimum for any combination of the vectors $y_{\overline{J-1}}$ and $u_{\overline{J-1}}$. But this term achieves its minimum value when

$$\Gamma_J = \delta(u_J-\hat{u}_J) \tag{4.20}$$

and $\hat{u}_J$ minimizes $G_J(u_J)$. Thus, Γ_J as given in Eq. (4.20) is the optimal control function during the Jth subinterval.

If the above procedure is repeated, the remaining optimal control functions can be found to have the similar form

$$\Gamma_j = \delta(u_j-\hat{u}_j) \tag{4.21}$$

where $\hat{u}_j$ minimizes the quantity $G_j(u_j)$ defined by

$$G_j(u_j) = \int_{\Omega(y_j)} \left\{ \min_{u_{j+1}} \int_{\Omega(y_{j+1})} \cdots \left\{ \min_{u_J} \int_{\Omega(y_J)} \left[\int_{(\alpha,\beta)} (\alpha-\hat{\alpha})^2 p(\alpha,\beta) \right.\right.\right.$$
$$\left.\left.\left. (x) \prod_{j=1}^{J} p(y_j/u_j,\alpha,\beta)\, d\alpha d\beta \right] dy_J \right\} \cdots dy_{j+1} \right\} dy_j . \tag{4.22}$$

The problems associated with the implementation of this equation in a real situation will be discussed in the next chapter.

This completes the derivation of the optimal filter function, (4.16) together with (4.17), and the optimal control functions, (4.21) together with (4.22), which jointly minimize the loss function L. By reinstituting the proper superscript on the control variable u_j, these optimal functions can be seen to coincide with those presented in Section 4.3. The expression for $p(y_j/u_j^1,\alpha,\beta)$ given in Eq. (4.2) is obtained by replacing the noise variable η_j by $[y_j-\beta C(\alpha,u_j^1,u^2)]$ in the probability density function for η_j. This follows from the fact that the probability of y_j being in a particular set A given u_j^1, α, and β is just the probability that η_j is in the set specified by the collection of point $y_j-\beta C(\alpha,u_j^1,u^2)$ as y_j ranges over the set A. (Remember, u^2 is also known due to Assumption A3.2.1.)

CHAPTER 5

DISCUSSIONS OF CONTROL THEORY SOLUTIONS

5.1 DISCUSSION

The goal of the three previous chapters has been to determine if "Stochastic Optimal Control Theory" can be applied to the direction finding problem. In many respects, this approach seems the most natural since most existing direction finders operate by controlling the pointing angle of a directional receiving antenna. This control mode is normally independent of past received signals or, at least, it is chosen without any real gard to what is optimal. Therefore, if "Stochastic Optimal Control Theory" can be applied, possibly in conjunction with a digital computer, many existing systems could be easily converted to take advantage of these techniques.

Rather than investigate a complicated system as to its optimal performance, a much simplified model was constructed which still retained the essential features of a typical direction finding systems. By using such a model, it was hoped that techniques could be established for solving the more general problem. Certainly, if no reasonable solution can be found for this simple model, it is doubtful that the more complicated models could be optimized to any real extent.

With the above considerations in mind, an investigation of the model under open-loop operation was initiated. This investigation was concerned with determining optimal filter and control functions which depend only on a priori information, i.e., filter and control functions which are not allowed to be altered during the observation interval as more information is received. These constraints are typical of those usually imposed on existing direction finding systems. In addition, because of their ease of implementation, the class of possible filters from which an optimal element was sought was restricted to those whose outputs are a linear combination of the received signals plus, possibly, a known bias level.

The filter "weights," the filter bias, and the controls presented in Section 3.3 were shown to result in a system whose performance approached arbitrarily close to that of an optimal open-loop system as the beam width became large compared with the width of the probability density function $p(\alpha)$. (The system performance was measured by the expected value of the squared error which, for the above filter and controls, has the form illustrated in Eq. (3.9).) This particular relationship between the beam width and the width of $p(\alpha)$ corresponds to the limiting case in which a great deal of a priori information is available concerning the true direction of arrival and only minor refinements are required.

Observing now the functional form of the filter "weights," the filter bias, and the controls specified in Section 3.3, we find that they exhibit the following properties:

P5.1
The filter "weights" are dependent on the applied controls.

P5.2
The controls are such that the antenna is always directed so the angle of maximum slope of its pattern is oriented in the a priori expected direction m_α.

P5.3
The controls are such that each of the two angles for which the antenna pattern slope is maximum are utilized for equal time intervals.

To understand this behavior, let us substitute Eq. (3.8) into Eq. (3.5) and assume the filter "weights" vector is $\tilde{\underline{b}}$ as defined in Eq. (3.7). We can then write

$$\alpha_e = m_\alpha + \sum_{j=1}^{J} \tilde{b}_j(y_j - \bar{y}_j) \qquad (5.1)$$

where $\bar{y}_j = m_\beta \tilde{p}_j$ which approaches arbitrarily close to the expected value of y_j as the beam width becomes large compared with the width of $p(\alpha)$. It can now be seen, the estimate α_e is the sum of the a priori expected value of α plus a weighted sum of the deviation of the samples $\{y_j\}$ from the values $\{\bar{y}_j\}$.

Let us now consider the jth sample and approximate the antenna pattern as a function of α in the vicinity of m_α by the first three terms of its power series expansion assuming the control is $m_\alpha +/\sqrt{2u^2}$. (Note for this control, the antenna pattern slope is maximum in the direction $\alpha = m_\alpha$ which implies the third term of the above power series expansion is zero.) This approximation to the pattern times m_β is then illustrated in Fig. 5.1 where the deviation of the sample y_j from $\bar{y}_j$ is plotted against α for the case where $\sigma_\beta^2 = 0$ and $\sigma_\eta^2 = 0$, i.e., for the case where β is known a priori and the sample y_j is noise free. (Remember, $y_j = \beta C(\alpha, u_j^1, u^2) + \eta_j$ in the general case so, when $\eta_j = 0$ and $\beta = m_\beta$, the sample y_j becomes $m_\beta C(\alpha, u_j^1, u^2)$.) Suppose, now, Δ_α is the deviation of the true angle of arrival and Δ_y is the corresponding deviation of the sample y_j from $\bar{y}_j$. Then, for this situation, it is obvious from Fig. 5.1 that Δ_α should be calculated by dividing Δ_y by the slope of $m_\beta C(\alpha, u_j^1, u^2)$ at m_α. Observing Eq. (3.7), we find this is exactly the operation performed by the filter specified by Eqs. (3.7) and (3.8) when σ_β^2 and σ_η^2 are zero. The quantity $m_\beta q_{max}/\sqrt{J}$ is the slope of $m_\beta C(\alpha, u_j^1, u^2)$ at $\alpha = m_\alpha$ when the control u_j^1 is $m_\alpha + 1/\sqrt{2u^2}$. Moreover, since the antenna pattern slope at $\alpha = m_\alpha$ depends on the control, it is clear that the filter should also depend on the controls as observed in Property P5.1.

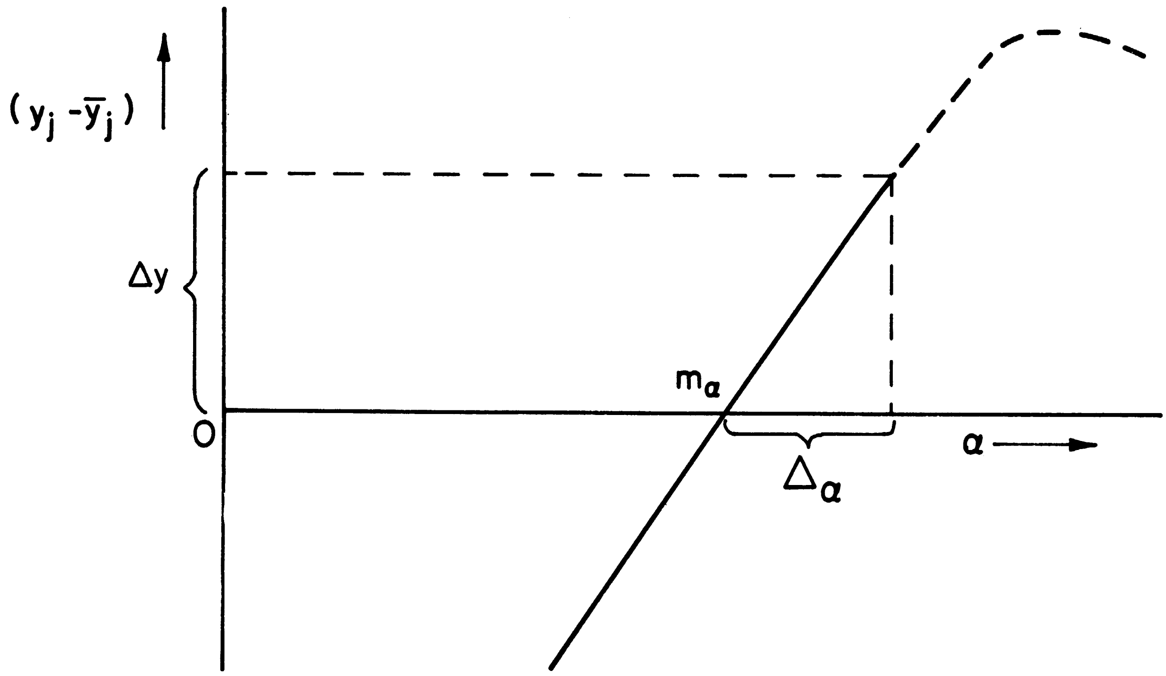

Fig. 5.1. A plot of the deviation of the sample y_j from $\bar{y}_j$ as a function of α in the vicinity of m_α when $u_j^1 = m_\alpha + 1/\sqrt{2u_j^2}$, $\sigma_\beta^2 = 0$ and $\sigma_\eta^2 = 0$.

If we now let y_j be a noisy sample so its value is displaced by an amount Δ_η from the value Δ_y illustrated in Fig. 5.1, dividing the deviation $\Delta_y+\Delta_\eta$ by the slope of $m_\beta C(\alpha,u_j^1,u^2)$ will result in the error Δ_α^η shown in Fig. 5.2. Moreover, it can be seen that Δ_α^η will be least when the slope is maximum which accounts for the second property listed above. Note, however, in the noisy case, the filter weights as specified in Eq. (3.7) are dependent on the variance of the noise as well as the slope of $m_\beta C(\alpha,u_j^1,u^2)$.

To understand Property P5.3, let us again assume y_j is a noise free sample but, now, allow β to be different from m_β. For this case, a plot of $m_\beta C(\alpha,u_j^1,u^2)$ in the vicinity of m_α when u_j^1 is equal to $m_\alpha+1/\sqrt{2u^2}$ differs from a similar plot of $m_\beta C(\alpha,u_j^1,u^2)$ as shown in Fig. 5.3 where the respective curves are labeled "Curve a" and Curve b." "Curve a" is displaced from "Curve b" by an amount Δ_β which depends on $\beta-m_\beta$ and possesses a correspondingly greater slope. Also in this figure, "Curve c" is drawn parallel to "Curve b" but intersecting "Curve a" and the line $\alpha=m_\alpha$ at a common point. Now, if Δ_α is again the deviation of the angle from m_α, dividing the corresponding sample deviation Δ_y by the slope of "Curve b" results in the estimation error $\Delta_{\alpha_1}^\beta$. (We are forced to divide by the slope of "Curve b" instead of the slope of "Curve a" because β is unknown.) This error, $\Delta_{\alpha_1}^\beta$, is equal to the sum of $\Delta_{\alpha_1}^{\beta 1}$ and $\Delta_{\alpha_1}^{\beta 2}$ as shown in Fig. 5.3. Let us now look at Fig. 5.4 where the previous figure is redrawn but with u_j^1 equal to $m_\alpha-1/\sqrt{2u^2}$. For this case, the error obtained by dividing the deviation Δ_y by the slope of "Curve b" is $\Delta_{\alpha_2}^\beta$ which equals $\Delta_{\alpha_2}^{\beta 1}$ minus $\Delta_{\alpha_2}^{\beta 2}$. Furthermore, the errors $\Delta_{\alpha_1}^{\beta 1}$ and $\Delta_{\alpha_2}^{\beta 1}$ are equal as are $\Delta_{\alpha_1}^{\beta 2}$ and $\Delta_{\alpha_2}^{\beta 2}$. It can now be seen, the use of the controls $m_\alpha+1/\sqrt{2u^2}$ and $m_\alpha-1/\sqrt{2u^2}$ for equal time intervals as stated in Property P5.3 eliminates the errors $\Delta_{\alpha_1}^{\beta 2}$ and $\Delta_{\alpha_2}^{\beta 2}$ since they cancel each

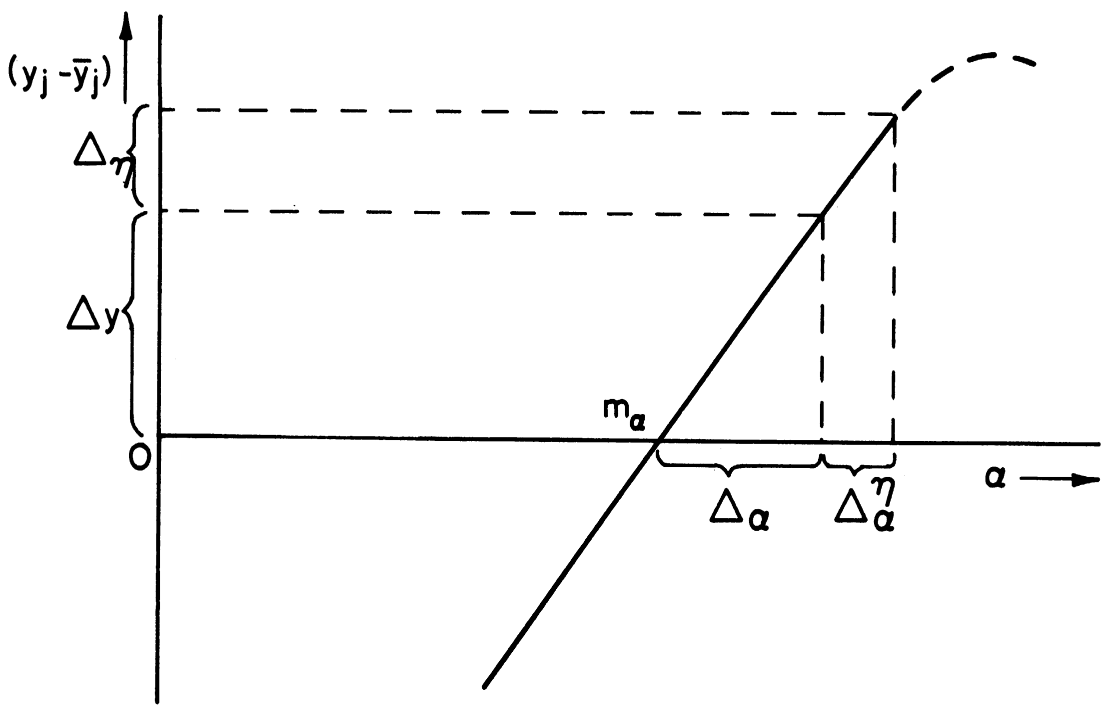

Fig. 5.2. A plot of the deviation of the sample y_j from $\bar{y}_j$ as a function of α in the vicinity of m_α when $u_j^1 = m_\alpha + 1/\sqrt{2u_j^2}$, $\sigma_\beta^2 = 0$ and Δ_η represents the noise in the sample.

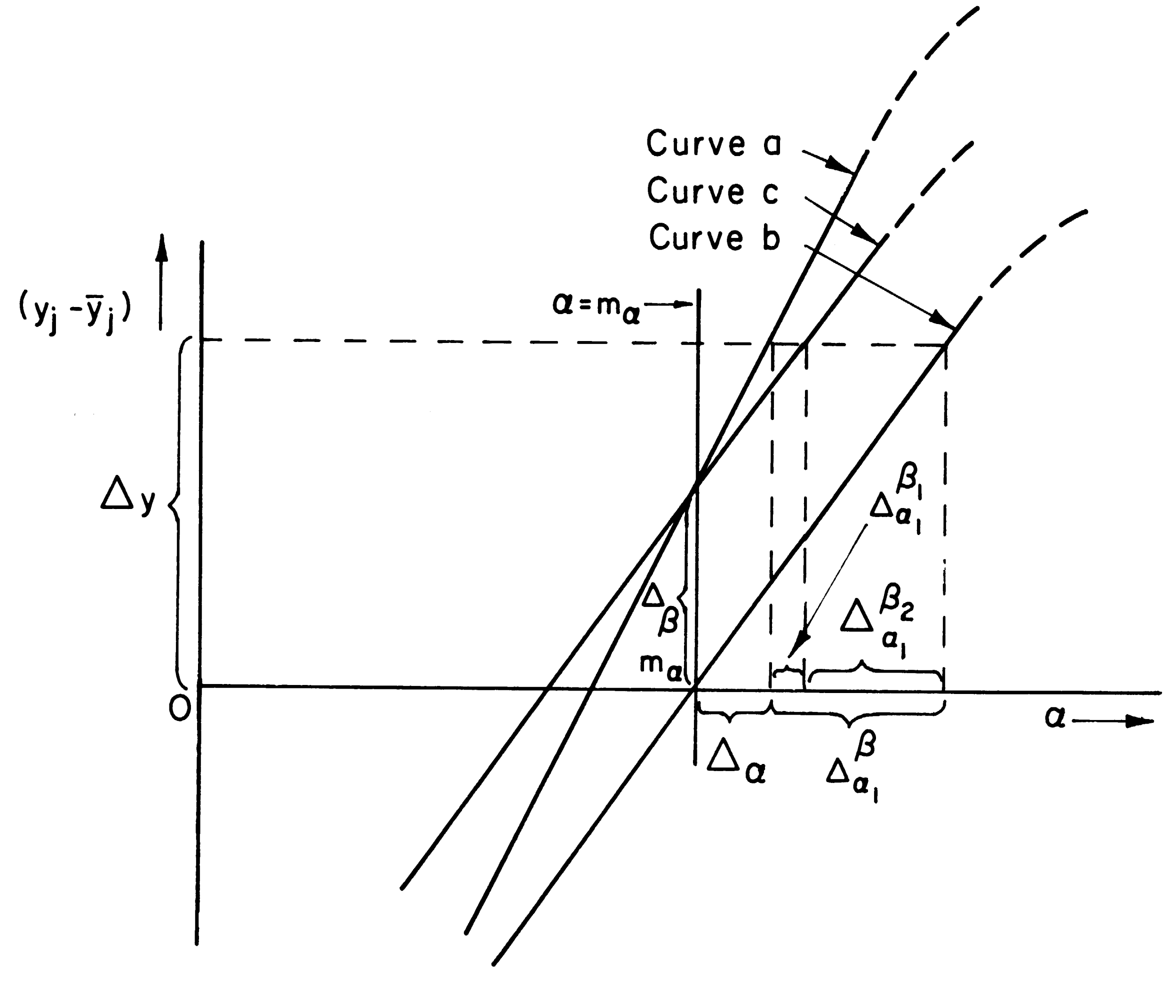

Fig. 5.3. Plot of the deviation of the sample y_j from $\bar{y}_j$ as a function of α in the vicinity of m_α when $u_j^1 = m_\alpha + 1/\sqrt{2u_j^2}$, $\sigma_\beta^2 \neq 0$ and $\sigma_\eta^2 = 0$.

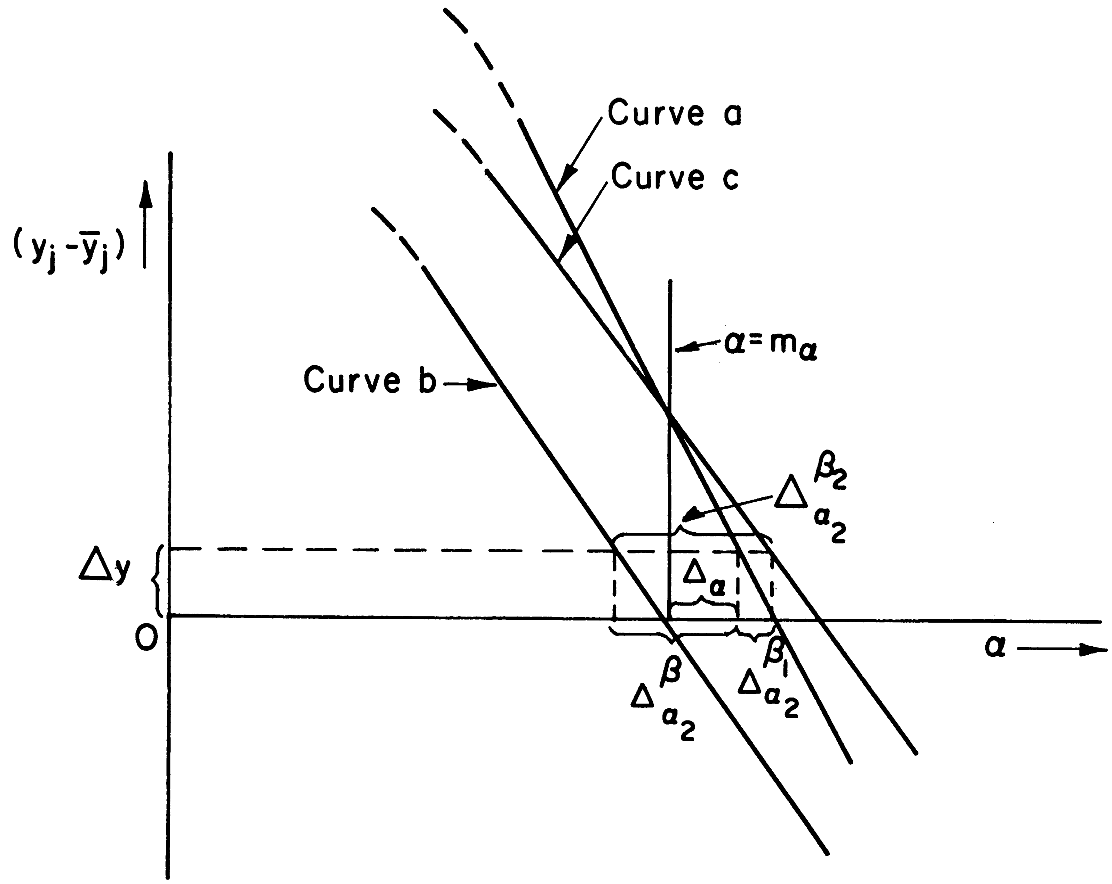

Fig. 5.4. Plot of the deviation of the sample y_j from $\bar{y}_j$ as a function of α in the vicinity of m_α when $u_j^1 = m_\alpha \pm 1/\sqrt{2u_j^2}$, $\sigma_\beta^2 \neq 0$ and $\sigma_\eta^2 = 0$.

other when the estimates of Δ_α are averaged. The errors $\Delta_{\alpha_1}^{\beta 1}$ and $\Delta_{\alpha_2}^{\beta 1}$, on the other hand, cannot be eliminated since they add when the average is taken.

One additional property could be added to the above list which simplifies the implementation of the filter specified by Eqs. (3.7) and (3.8). It is easily shown that $\sum_{j=1}^{J} \hat{b}_j' \hat{p}_j$ is zero so Eq. (5.1) reduces to

$$\alpha_e = m_\alpha + \sum_{j=1}^{J} \tilde{b}_j y_j \qquad (5.2)$$

and, as a result, it is unnecessary to subtract $\bar{y}_j$ from y_j for each sample as suggested by Eq. (5.1). This follows from the fact that the slopes of $\beta C(\alpha, u_j^1, u^2)$ at m_α when $u_j^1 = m_\alpha + 1/\sqrt{2u^2}$ and $u_j^1 = m_\alpha - 1/\sqrt{2u^2}$ are the negative of each other while $\bar{y}_j$ remains fixed for all samples.

Finally, to obtain an indication of how the error variance is affected when Eqs. (3.6)-(3.8) are utilized, let us write Eq. (3.9) in the equivalent form

$$E\{(\alpha_e - \alpha)^2\} = \sigma_\alpha^2 \left\{ \frac{C_1 + C_2}{1 + C_1 + C_2} \right\} + \varepsilon' \qquad (5.3)$$

where $C_1 = \sigma_\beta^2 / m_\beta^2$ and $C_2 = \sigma_\eta^2 / (\sigma_\alpha^2 m_\beta^2 q_{max}^2)$. A plot of the function $(C_1+C_2)/(1+C_1+C_2)$ against C_2 for various values of C_1 is then illustrated in Fig. 5.5, which, together with Eq. (5.3), can be used to evaluate the performance of the estimation technique presented in Chapter 3. (Remember, the term ε' goes to zero as the beam width becomes large compared with the width of $p(\alpha)$.) The constant q_{max}^2 is equal to J times the square of the antenna pattern slope at m_α.

The question can now be asked, "Are the above stated properties independent of the particular antenna pattern utilized or do they depend in a special way on the assumed form that was employed?" Returning to the

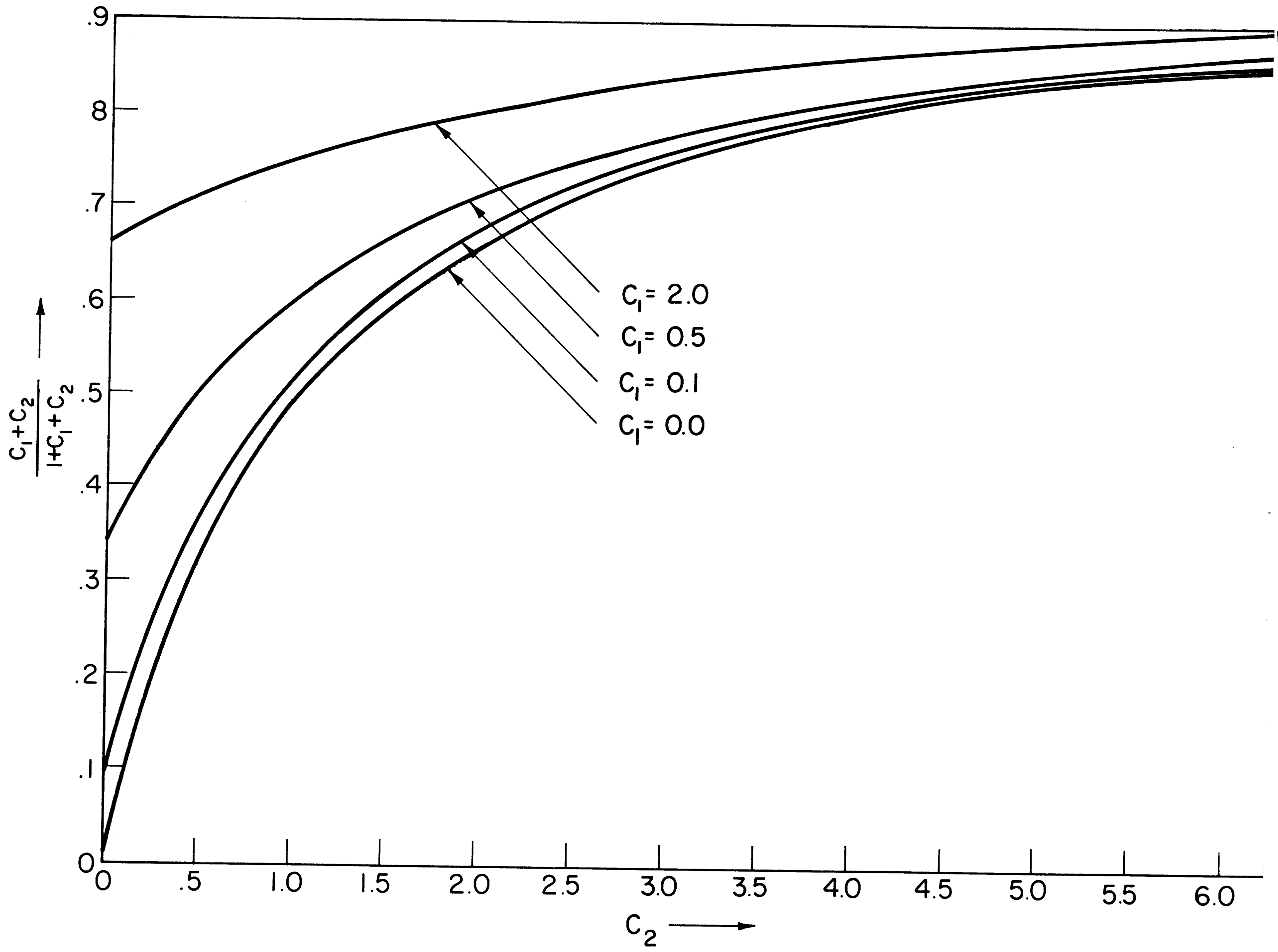

Fig. 5.5. A plot of $(C_1+C_2)/(1+C_1+C_2)$ vs. C_2 for various values of C_1.

deviation of Chapter 3, we find the same solution will be obtained regardless of the antenna pattern as long as it is symmetrical about its pointing angle. If the pattern is not symmetrical, it is not clear what form the optimal control should take although the resulting error can be calculated easily from Eq. (3.22).

Turning now to the closed-loop results presented in Section 4.3, we see immediately the optimal filter is a nonlinear function of the samples $\{y_j\}$. In addition, we see the present optimal control depends not only on past samples $\{y_j\}$ but also on future controls. This is typical of many solutions obtained by the use of dynamic programming techniques. If we now take a closer look at Eq. (4.3), a major difficulty is directly noticed. Rather than provide an explicit expression for the jth optimal control, Eq. (4.3) represents an identity that the optimal control u_j^1 must satisfy. In order to implement this result in a realistic problem, a method for solving this equation must be available. Several attempts were made to perform the indication integrations analytically when various probability densities were substituted for p_o and various antenna patterns were assumed present. In one instance, the integration was attempted when p_o had the form of a Gaussian probability density and $C(\alpha,u_j^1,u^2)$ had the form $\exp\{-u^2(\alpha-u_j^1)^2\}$. These attempts did not meet with any success, however, so the possibility of performing the integrations numerically with a digital computer was explored. Although useful for implementing the optimal filter of Eq. (4.1), numerical integration was found to be impractical approach to a solution of Eq. (4.3) since the variables $\{y_j\}$ range from $-\infty$ to $+\infty$. There are simply too many arithmetical operations to be performed, even with a digital computer, when a reasonable number of samples J is considered. Finally attempts were made to approximate the optimal controls by assuming the probability density $p(\alpha,\beta)$ was a "narrow" function about m_α and m_β as done in Chapter 3. Some preliminary results were obtained but, because of their marginal applicability, have not been included in this presentation.

In conclusion, it was found that "Stochastic Optimal Control Theory," at least that applied by the author, did not give a satisfactory answer to the question of how an optimal direction finding system should operate. As a result, it was decided to alter the above approach and attempt to apply, instead, "Estimation Theory" to this problem. This altered approach, utilizing "Estimation Theory," is the subject of the remaining chapters.

CHAPTER 6

STATEMENT OF THE ESTIMATION PROBLEM

6.1 INTRODUCTION

In the previous chapters, a direction finding system was considered which employed a beam forming network. This network delayed and combined the outputs from the various elements of an antenna array in a manner specified by a controller so as to produce a system which had the characteristics of a single antenna of variable directivity and whose direction of maximum gain could be varied. These directional properties of the system were then exploited for the purpose of determining the direction of arrival of the impinging signal with a "control function" and "filter function" being derived which resulted in optimal performance of the system. In the remaining chapters, a direction finding system will be considered in which the beam forming network is absent and the outputs of the antenna elements are processed directly. This method of attack results in what appears to be a more general solution in the sense that the system is, initially, less constrained. This statement must be qualified, however, since a different criterion of optimality is applied to the model. In any event, this technique results in a system which lends itself to a more complete mathematical analysis and which can be implemented by hardware that is available today.

6.2 STATEMENT OF THE PROBLEM

The objective of the remaining chapters, as in the previous chapters, will be to derive a technique for determining a "best* estimate" of the

*The optimality criterion is discussed in Assumption A6.3.3 of Section 6.3.

direction of arrival of a random signal which has been emitted by a point source. This estimate is to be based on the time signals observed at the outputs of K antennas which form a receiving antenna array. In addition to the signal originating from the point source, we shall assume the received signals include additive noise terms. We shall also assume the source is far enough removed that, at the site of the receiving array, the random signal takes the form of a plane wave. The geometry of a portion of the array relative to the plane wave is illustrated in Fig. 6.1 while the system which is to provide the estimate is illustrated in Fig. 6.2.

In Fig. 6.1, a coordinate system has been drawn with an arbitrary element of the array positioned at its origin. A portion of the wavefront assumed to be impinging upon the array together with a unit vector $\underline{u}$ originating at the origin and perpendicular to this wavefront have been indicated. The angles between $\underline{u}$ and the x and y axes are denoted by α^1 and α^2, respectively, and jointly represent the angle of arrival of the wave. Also in this figure, the vector $\underline{\ell}_k$ has been drawn from the origin to the site of the kth element of the array. We shall now assume, without loss of generality, the vectors $\underline{u}$ and $\underline{\ell}_k$ are column vectors whose components are the projections of these vectors on the coordinate axes. It can now be seen, if s(t) is the time signal received at the origin due exclusively to the point sources, the corresponding signal $[s(t)]_k$ received by the kth element of the array is given by

$$[s(t)]_k = s\left(t + \frac{\underline{\ell}_k^T \underline{u}}{c}\right) \tag{6.1}$$

where c is the velocity of propogation of the wave, $\underline{\ell}_k^T$ is the transpose of $\underline{\ell}_k$, and $\underline{\ell}_k^T\underline{u}$ is the matrix product of $\underline{\ell}_k^T$ and $\underline{u}$. (In the following pre-

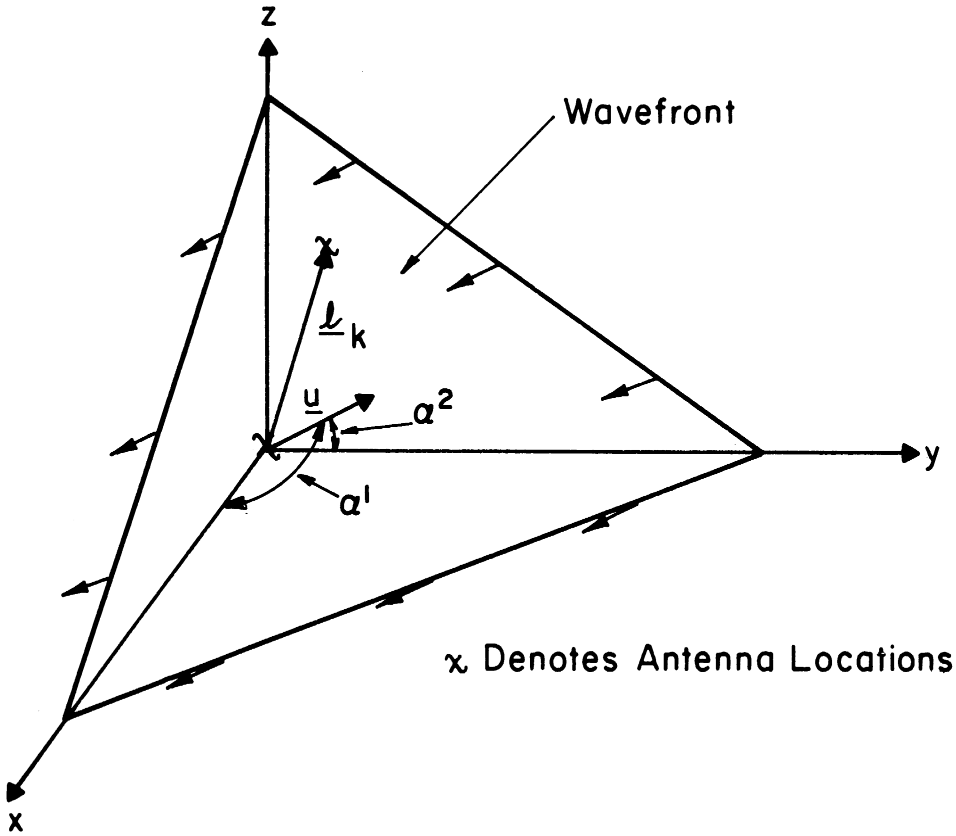

Fig. 6.1. Geometry of the receiving array relative to the direction of arrival of the signal.

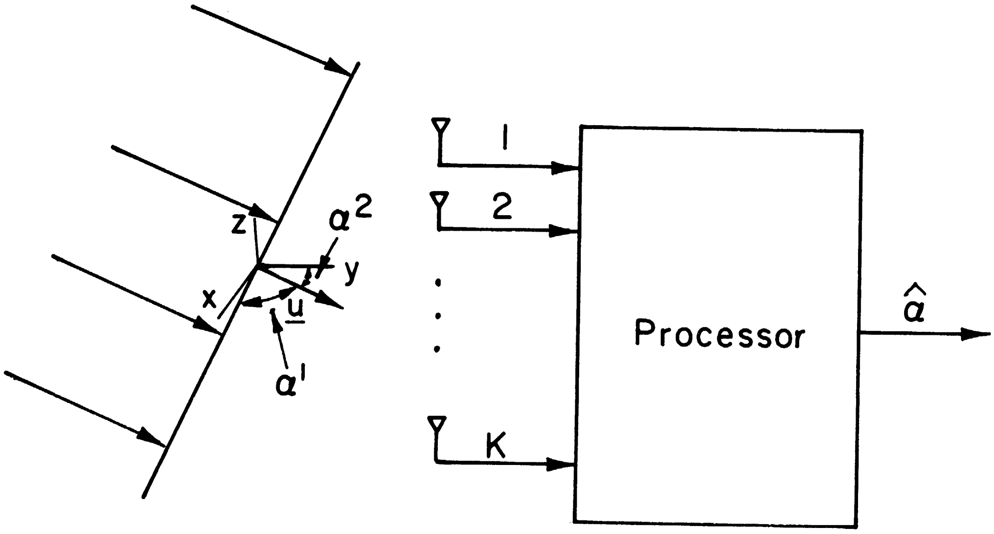

Fig. 6.2. Processing system to be used to provide an estimate of the direction of arrival of the signal.

sentation, brackets will be used to distinguish K-dimensional column vectors whose individual elements are either time functions or random processes. Moreover, the KxK matrix covariance functions corresponding to these random processes will also be characterized with brackets while individual elements of these matrices as well as the above K-dimensional vectors will be denoted by subscripts. Thus, we shall use $[s(t)]$ to represent the column vector of signals received by the antenna array due to the point source and $[s(t)]_k$, as in Eq. (6.1), to denote the kth element of $[s(t)]$ which is the signal received by the kth array element due to the point source.)

In Fig. 6.2, the outputs from the various antenna elements are shown as inputs to a processor whose output is represented by $(\hat{\alpha}^1,\hat{\alpha}^2) = \hat{\alpha}$ which is a "best estimate" of $(\alpha^1,\alpha^2) = \alpha$.

The problem now reduces to the determination of the operations to be performed on the input data by the processor so as to produce an optimal estimate. We will say these operations have been specified when an equation has been provided whose solution is readily obtainable and represents the optimal estimate. In addition to providing this equation, consideration will be given to the problem of obtaining an operational method for providing a solution to the equation by the use of existing hardware.

6.3 ASSUMPTIONS (General Case)

In order to obtain a solution to this problem it will be necessary to be more specific concerning the type of noise that will be present, the type of signals that might be emitted by a typical point source and the criterion of optimality that must be satisfied. We shall begin by specifying the most general conditions under which we are able to obtain a solution. In the next section, we shall present conditions which are more restrictive but which, at the same time, allow us to obtain a more tractable solution. The general solution, which is derived in Chapter 7, is optimal when the following assumptions are valid.

A6.3.1

The time signal $[y(t)]_k$ received by the kth antenna of the array during the observation interval $[0,T_f]$ is the sum of noise and a signal which is due to the point source alone whose angular location is desired. Each of these terms represent the kth component of a vector sample function from a zero mean K-dimensional, vector Gaussian process. (See Section 7.2.1 for a discussion of random processes.) If $r_S(t,v)$ is the covariance function of the signal (point source) process driving the antenna located at the origin in Fig. 6.1 and if α is the true direction of arrival, the matrix covariance function of the signal process $[S^{\alpha}(t)]$ driving the entire array is given by $[r^{(\alpha)}(t,v)]$ and its jkth element takes the form

$$[r^{(\alpha)}(t,v)]_{jk} \;=\; r_S(t+\tau_j,v+\tau_k) \tag{6.2}$$

where

$$\tau_n \;=\; \frac{\underline{\ell}_n^T\underline{u}(\alpha)}{c} \quad ,$$

$n\varepsilon\{j,k\}$, and c is the velocity of propagation of the wave (see Eq. (6.1)). Furthermore, the elements of this matrix are continuous and bounded on $[0,T_f]x[0,T_f]$ where $[0,T_f]$ represents the observation interval. The noise process $[N(t)]$ driving the array is independent of the signal process and the elements of its covariance function $[r^0(t,v)]$ are bounded and continuous on $[0,T_f]x[0,T_f]$ also. In addition, the noise process has the property of "separability" (see Ref. 29, p. 52) and there exist constants $C > 0$ and $\delta > 0$ such that

$$\sum_{k\epsilon K} \{[r^o(t,t)]_{kk} + [r^o(v,v)]_{kk} - [r^o(v,t)]_{kk} - [r^o(t,v)]_{kk}\} < C|t-v|^\delta$$

when $t,v\epsilon[0,T_f]$. (These last two conditions will be used to guarantee the continuity of "nearly all" sample functions of the noise process $[N(t)]$. Continuity of the sample functions is a reasonable requirement since they are the observed outputs of physical devices which possess finite response times.) Thus, the total process $[Y^\alpha(t)]$ driving the array is given by

$$[Y^\alpha(t)] = [S^\alpha(t)] + [N^\alpha(t)]$$

and is Gaussian, of zero mean, and possesses a covariance function of the form

$$[r^\alpha(t,v)] = [r^{(\alpha)}(t,v)] + [r^o(t,v)] \, . \tag{6.3}$$

Finally, the covariance function of the noise process $[N(t)]$ satisfies

$$\int_0^{T_f}\int_0^{T_f} [f(t)]^T[r^o(t,v)][f(v)]dt\ dv > 0 \tag{6.4}$$

for every K-dimensional, column vector valued function $[f(t)]$ which satisfies

$$0 < \int_0^{T_f} [f(t)]^T[f(t)]dt < \infty$$

(The superscript T on the vector $[f(t)]$ denotes the transpose.)

A6.3.2

The signal processor has available certain a priori information concerning the possible values of the angle of arrival. In particular, this information takes the form of a known probability distribution function for the possible angles of arrival.

A6.3.3

The "criterion of optimality" to be satisfied is that of "Minimum Probability of Error (MPE)." If the direction of arrival has one of M possible values, this criterion requires the space of possible received signals Ω to be partitioned into M disjoint subsets $(\Lambda_1, \Lambda_2, \ldots, \Lambda_M)$ which have the property

$$P_e = \sum_{i=1}^{M} \Pr(\alpha_i) \Pr_i(\Omega - \Lambda_i) \qquad (6.5)$$

is minimum among the set of possible partitions. In Eq. (6.5), $\Pr(\alpha_i)$ represents the a priori probability that α_i is the true direction of arrival and $\Pr_i(\Omega - \Lambda_i)$ is the probability of the received signal not being an element of Λ_i when the true direction of arrival is α_i. (Another term can be added to the sum in Eq. (6.5) if the possibility of noise alone being received is present. See Section 7.3 for additional disccusion.) Thus, if a decision scheme is used which specifies α_i as the true direction of arrival when the received signal is contained in Λ_i, it can be seen that P_e is truly the probability of error. A similar expression is obtained in the more realistic case where $M \to \infty$. (In Appendix V, this optimality criterion is shown to result in a partition that is "equivalent" to that obtained by the use of the "Maximum A Posteriori Probability (MAP)" criterion. This latter criterion places a particular received signal in the subset Λ_i if α_i has the greatest probability of being the true angle of arrival based on the received signal and the a priori information.)

Assumption A6.3.1 implies (1) the dimensions of the array are small compared with the distance separating the source and the array so that the wavefront of the source signal is essentially a plane wave when viewed at the site of the array, (2) the individual elements of the array are separated by a distance which is sufficient to produce negligible interaction of elements (this requires* the interelement spacing to be somwhat greater than the largest dimension of the elements), and (3) the antenna elements have onmidirectional receiving patterns over the region of surveillance. In an attempt to justify the assumption of known covariance functions, let us remember that, in the case of stationary sources, knowledge of the covariance functions is equivalent to the knowledge of the power spectrums which is a reasonable assumption in many direction finding problems. The assumption that the signal and noise processes are Gaussian is typical of the restrictions normally placed on communications problems. As always, the main justification for this assumption lies in the fact that it simplifies the calculations and results in reasonable solutions to the problem. Finally, we shall find the property presented in Eq. (6.4) easily verified in the cases which will be of particular interest to us, namely, those in which the noise process driving each array element is stationary and independent of all others.

Assumption A6.3.2 is particularly appropriate in the case where continuous surveillance of a mobile source is not possible, e.g., either the source does not emit continuously or the direction finding system cannot devote its efforts to this task continuously. Between each period of surveillance, the source is allowed to change its position with the result that, before each period, there exists an uncertainty as to

*See Ref. 10, p. 122.

its true position which is reflected in the given probability distributions.

Many different optimality criterions could have been used instead of that specified in Assumption A6.3.3. Intuitively, the probability of error appears to be a reasonable measure of performance of the system. The preponderant reason for choosing this criterion, however, is the fact that the use of it results in a solution that can be easily implemented with existing hardware.

6.4 ASSUMPTIONS (Special Cases)

By placing further restrictions on the functional form of the noise and signal covariance functions, we are able to produce several special cases which are often reasonable descriptions of the environment of a direction finding system. Moreover, the solutions obtained for these cases represent a tremendous simplification over the general case solution obtained in Chapter 7 as far as implementation is concerned. In each of these special cases which are analyzed in Chapter 8, the covariance function of the noise process will have a different assumed form. (See Section 6.4.1 and 6.4.2 for characterizations of these covariance functions.) Before specifying these forms, however, let us introduce the following assumption which, also, will be employed in attaining each of the special case solutions.

<u>A6.4.1</u>

The noise and signal processes are stationary. Furthermore, each of these processes has the "band-limited" property that the Fourier transform of the jkth element of its covariance function, $[F(\omega)]_{jk}$, is representable in the form illustrated by Fig. 6.3 where "most" of the trans-

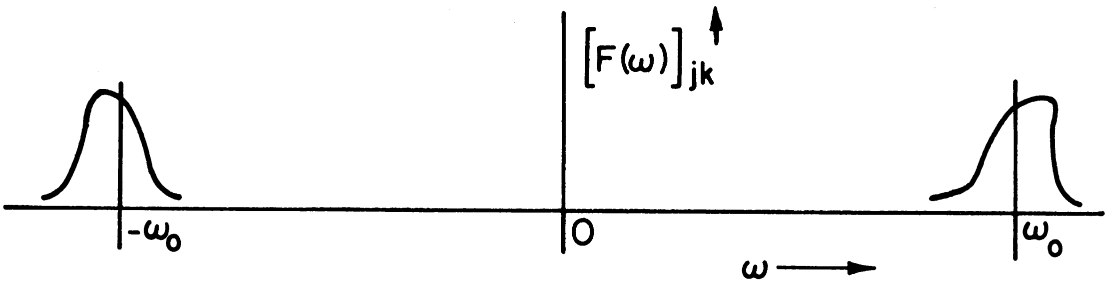

Fig. 6.3. Typical Fourier transform of $[r^{\alpha}(\tau)]_{jk}$ which represents the cross-covariance function of the processes driving the jth and kth antenna of the receiving array.

form is concentrated near a frequency ω_0 and where the reciprocal of the difference between the lowest and highest frequency (of significant energy) is large compared with the propagation time of signals across the array. Also, there exists a symmetric kernel $[h_\alpha^T(t,s)]$ which is square integrable on TxT and which satisfies

$$\int_0^{T_f} \int_0^{T_f} [r^o(u-t)][h_\alpha^T(t,s)][r^\alpha(s-v)]dtds = [r^\alpha(u-v)] - [r^o(u-v)] . \tag{6.6}$$

Finally, T_f is large compared with the reciprocal of the difference between the lowest and highest frequency (of significant energy) of $[F(\omega)]_{jk}$ so that $[h_\alpha^T(t,s)]$, for $t,s\varepsilon[0,T_f]$, can be "approximated" by the kernel $[h_\alpha^\infty(t-s)]$ which satisfies

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [r^o(u-t)][h_\alpha^\infty(t-s)][r^\alpha(s-v)]dtds = [r^\alpha(u-v)] - [r^o(u-v)] . \tag{6.7}$$

(The sense in which $[h_\alpha^T(t,s)]$ is to be approximated by $[h_\alpha^\infty(t-s)]$ will be discussed in Chapter 8.)

The solution of Eq. (6.7) is easily obtained by the use of Fourier transforms and conditions on $[r^o(t,s)]$ and $[r^\alpha(t,s)]$ are easily found in order that this solution is symmetric and square integrable on TxT. Moreover, it seems reasonable that the solution of Eq. (6.6) should "approach" the solution of Eq. (6.7) as T_f becomes large. We shall find, however, the existence of a bounded symmetric solution to Eq. (6.6) and the fact that it is "approximated" by $[h_\alpha^\infty(t-s)]$ is very difficult to verify when specific covariance functions are considered.

The following two sections characterize the forms that the noise covariance functions will possess in the special cases analyzed.

6.4.1 Independent, Identically Distributed Noise

This special case is a reasonable description of the case where the noise is due primarily to the thermal noise introduced by the receiving elements of the antenna array.

<u>A6.4.1.1</u>

The noise processes driving each of the receiving antenna elements are independent and their covariance functions are identical and denoted by $r_N(\tau)$.

6.4.2 Independent, Identically Distributed (Except for an Amplitude Factor) Noise

This case is similar to that of Section 6.4.1 except now we are allowing the noise level to vary between antenna elements.

<u>A6.4.2.1</u>

The noise processes driving each of the receiving antenna elements are independent and their covariance functions are identical except for an amplitude factor with the kth process covariance function denoted by $a_k' r_N(\tau)$.

CHAPTER 7

GENERAL SOLUTION OF THE ESTIMATION PROBLEM

7.1 INTRODUCTION

The method of solution that will be employed in this chapter, will rely heavily on the theories of probability, stochastic processes and functional analysis. Such books as Probability Theory by M. Loève,[28] Stochastic Process by J. Doob,[29] Measure Theory by P. Halmos,[30] and Functional Analysis by A. Taylor[44] are excellent references for this presentation. However, an attempt will be made to make this treatment self-contained in the sense that all necessary definitions will be included with theorems being quoted if they are needed. In general, the proofs of the theorems will be omitted if they are readily available in the literature.

In this chapter, a technique will be developed for detecting a signal and determining its angle of arrival at the site of a receiving antenna array. The application of this technique will result in the specification of either (1) "no signal is present" or (2) "an estimate of the true angle of arrival of the unknown signal is α_i" where α_i is contained in the set of possible angles of arrival. This specification will be optimal in the sense that it minimizes the probability of error. (Remember, the optimality criterion chosen in Section 6.3 was "Minimum Probability of Error.") Moreover, it specifies the hypothesis of "Maximum a Posteriori Probabiliy," i.e., it specifies the hypothesis that is "most probable" based on the signals received by the array and any priori information that is available concerning the true angle of arrival and the random processes present. The form of these signals and the nature of the a priori information is discussed in Section 6.3.

The approach pursued in this chapter is an extension of that used by T. T. Kadota,[36-39] T. S. Pitcher,[45] and B. H. Bharucha[46] to solve the problem of distinguishing between possible transmitted signals of differing covariance functions by an observation of a single received signal. In the direction finding problem we are interested in distinguishing between possible directions of arrival by the observation of a vector of received signals which is a sample function of a vector valued process whose covariance function varies in a known manner with the direction of arrival of the signals.

7.2 MATHEMATICAL PRELIMINARIES

Concepts and theorems that will be useful in the remainder of this chapter will be briefly reviewed in the next two sections.

7.2.1 Probability Theory and Stochastic (Random) Processes

Elementary probability theory suffices to handle problems of the type considered in the four preceding chapters of this thesis in which only a finite number of random variables are present. However, when continuous parameter processes are present, questions concerning existence and uniqueness become important and necessitate a more fundamental approach in which very precise definitions are required. As a result, we will begin by stating these definitions and proceeding from this point.

In order to consider random processes, we must first recall what is meant by a probability space and a random variable.

Definition 7.1: A probability (measure) space $(\Omega, \mathcal{B}, P)$ is the triple of the sure event Ω, the (nonempty) σ-field $\mathcal{B}$ of events (or measurable sets) and the probability (measure) P defined on $\mathcal{B}$. (The symbols $\mathcal{B}$ and P with various subscripts will be reserved for the above quantities while ω will be used to denote the individual elements of Ω.)

The terms measure,* probability and probability measure will be used interchangably throughout this treatment since a probability can be defined as a measure for which $P(\Omega) = 1$. A special type of probability measure that will be of particular interest is called a <u>complete</u> probability measure. This measure has the property that subsets of sets of measure zero are measurable, i.e., if a set A is contained in a set B and if $P(B) = 0$, then $A \epsilon \mathcal{B}$ and $P(A) = 0$. Actually, this is a very reasonable property for a probability space to possess since, intuitively, we would expect a subset of a set of probability measure zero to have zero probability also. It can be shown** that a given probability measure can always be completed in a unique way by a slight enlargement of the σ-field $\mathcal{B}$. This enlargement, $\bar{\mathcal{B}}$, consists of all sets of the form $B \cup N$ wher $B \epsilon \mathcal{B}$ and N is a subset of a set in $\mathcal{B}$ of measure zero. (The σ-field $\bar{\mathcal{B}}$ is also given by sets of the form $B \Delta N$ where B and N are as above and $B \Delta N$ denotes the symmetric difference of B and N which equals $(B-N) \cup (N-B)$.) The measure on $\bar{\mathcal{B}}$, $\bar{P}$, is then defined by

$$\bar{P}(B \cup N) = \bar{P}(B \Delta N) = P(B) \tag{7.1}$$

<u>Definition 7.2</u>: A <u>real random variable</u> X on $(\Omega, \mathcal{B}, P)$ is a mapping X of Ω into the real line $\Omega_R = (-\infty, \infty)$ which is <u>measurable</u> relative to $\mathcal{B}$ and the σ-field $\mathcal{B}_R$ of Borel sets*** in Ω_R, i.e., for**** every $B \epsilon \mathcal{B}_R$, we have $\{\omega; X(\omega) \epsilon B\} \epsilon \mathcal{B}$. (The concept of an <u>expectation</u> of a random vari-

*See Ref. 30, p. 30, for the definition of a measure.
**See Ref. 30, p. 55.
***See Ref. 30, p. 62, for the definition of Borel sets.
****It is shown in Ref. 30, p. 79 that this definition of measurability of a random variable is equivalent to requiring, for all real a, $\{\omega : X(\omega) \leq a\} \epsilon \mathcal{B}$.

able X, E{X}, will also prove to be very useful and is defined as the integral of the random variable X over the probability space Ω, i.e., $E\{X\} = \int_\Omega X\, dP$.)

A random variable as defined above can be seen to induce a probability measure on the Borel sets of the real line. In particular, there exists a distribution function

$$F(\lambda) = P\{\omega; X(\omega) \leq \lambda\} \qquad (7.2)$$

which is defined for all real λ. This distribution function then defines a probability measure P_R

$$P_R(B) = \int_B dF(\lambda) \qquad (7.3)$$

for all B where B is a Borel set and the integral is of the Lebesgue-Stieltjes type. Actually, we have performed a measure preserving transformation from $(\Omega, \mathcal{B}, P)$ to $(\Omega_R, \mathcal{B}_R, P_R)$ as discussed by Doob.* The utility of this transformation becomes apparent if we consider, for example, the problem of determining the expectation of a random variable of the form $\Phi(X)$ where Φ is a Baire** function of the real random variable X. It can be shown that

$$\int_\Omega \Phi[X(\omega)]dP = \int_{-\infty}^{\infty} \Phi(\xi)dF(\xi) \qquad (7.4)$$

and, as a result, it is unnecessary to revert to the original probability space to calculate the expectation. Moreover,*** any problem involving random variables defined on $(\Omega, \mathcal{B}, P)$ which are measurable with respect to $\mathcal{B}(X)$, (the smallest σ-field on which X is measurable), can be ex-

*See Ref. 29, p. 617.

**Any function defined on the real line which is measurable is called a Baire function.

***See Ref. 29, p. 621.

pressed as the corresponding problem in terms of random variables defined on the induced probability space $(\Omega_R, \mathcal{B}_R, P_R)$. (The σ-field $\mathcal{B}$ should not be confused with the σ-field $\mathcal{B}(\cdot) \subset \mathcal{B}$ which is dependent on the quantity contained within the parentheses.)

In practice, the inverse of the above operation arises. That is, what properties must a given function $F(\lambda)$ possess in order that there exist a probability space and a random variable with $F(\lambda)$ as its distribution function? Loève* shows that requiring $F(\lambda)$ to be monotone nondecreasing, continuous on the right, and

$$\lim_{\lambda \to -\infty} F(\lambda) = 0, \qquad \lim_{\lambda \to +\infty} F(\lambda) = 1 \tag{7.5}$$

is sufficient for the existence of such a probability space and random variable. If, in addition, $F(\lambda)$ can be written as the Lebesgue integral

$$F(\lambda) = \int_{-\infty}^{\lambda} p(X)\,dX , \tag{7.6}$$

we say that $p(X)$ is the <u>density function</u> corresponding to $F(\lambda)$. We will be mainly interested in the case where X is a <u>Gaussian</u> random variable. In this case, $p(X)$ exists and takes the form

$$p(X) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{1}{2\sigma_X^2}(X-M_X)^2\right\} \tag{7.7}$$

where M_X and σ_X^2 are the <u>mean</u> and <u>variance</u>, respectively, of the random variable X, i..e, $M_X = E\{X\}$ and $\sigma_X^2 = E\{X-M_X\}^2$.

Since we will be interested in families of random variables, we must define what is known as a multivariate distribution function. In particular if $X_1,\ldots,X_n$ are real random variables, the function defined by

$$F(\lambda_1,\ldots,\lambda_n) = P\{\omega; X_j(\omega) \le \lambda_j, \quad j=1,2,\ldots,n\} \tag{7.8}$$

---

*See Ref. 28, p. 167.

is called their multivariate distribution function. Conditions that a function must satisfy in order that there exist a probability space and random variables with such a multivariate distribution are given by Loève.* These conditions require $F(\lambda_1,\ldots,\lambda_n)$ to be nondecreasing and continuous on the right in each of variables $\{\lambda_j\}$ with $F(\lambda_1,\ldots,\lambda_n) = 0$ if any $\lambda_j = -\infty$ and $F(\lambda_1,\ldots,\lambda_n) = 1$ when each $\lambda_j = +\infty$. In the case where the family of random variables has more than a finite number of elements, Kolmogorov** has proved that, if for every finite t set $t_1,\ldots,t_n$, the multivariate distribution function $F_{t_1,\ldots,t_n}$ of $X_{t_1},\ldots,X_{t_n}$ is prescribed and if

$$(1)\ F_{t_1,\ldots,t_n}(\lambda_1,\ldots,\lambda_n) = F_{t_{\pi_1},\ldots,t_{\pi_n}}(\lambda_{\pi_1},\ldots,\lambda_{\pi_n}) \qquad (7.9)$$

where $\pi_1,\ldots,\pi_n$ is a permutation of $1,\ldots,n$, and

$$(2)\ F_{t_1,\ldots,t_m}(\lambda_1,\ldots,\lambda_m) = \lim_{\substack{\lambda_j\to\infty \\ j=m+1,\ldots,n}} F_{t_1,\ldots,t_n}(\lambda_1,\ldots,\lambda_n) \qquad (7.10)$$

for $m < n$, then there exists a probability space and a family of random variables with the above finite multivariate distribution functions.

When a distribution function of n random variables is available, it can also be shown to induce a probability measure P_R^n on the n-dimensional Euclidean space

$$\Omega_R^n = \prod_{j=1}^{n} \Omega_{R_j}$$

by

*See Ref. 28, p. 169.
**See Ref. 29, p. 10.

$$P_R^n(B) = \int_B \cdots \int dF(\lambda_1,\ldots,\lambda_n) \tag{7.11}$$

where $\Omega_{R_j} = \Omega_R$ for all j and B is any Borel set in Ω_R^n. The extension of this result to a nonfinite family of random variables can be accomplished and is discussed by Halmos.* As in the single variable case, any problem involving random variables measurable with respect to $\mathcal{B}(X_t;t\epsilon T)$, the minimum σ-field on which X_t is measurable when t is contained in the set T, can be expressed as the corresponding problem in terms of random variables defined on the induced probability space.

The concept of a measure (probability) as discussed above, allows us now to define several types of convergence which will be useful in the following development.

Definition 7.3: A sequence of random variables $\{X_n\}$ is said to converge almost everywhere (a.e.) to a random variable X, with respect to a measure P, if there exists a set N for which P(N) = 0 and $X_n(\omega)\to X(\omega)$ when $\omega \notin N$.

Definition 7.4: A sequence of random variables $\{X_n\}$ is said to converge in the mean to a random variable X, with respect to a measure P, (denoted $\underset{n\to\infty}{\text{l.i.m.}} X_n = X$) if, for every $\varepsilon > 0$, there exists a real number n_o such that, for $n \geq n_o$,

$$\int_\Omega |X_n - X|^2 dP < \varepsilon \; .$$

Definition 7.5: A sequence of random variables $\{X_n\}$ is said to converge in measure (probability) to a random variable X, with respect to a measure P, if, for every $\epsilon_1, \epsilon_2 > 0$, there exists a real number n_o such that, for $n \geq n_o$,

*See Ref. 30, p. 154.

$$P\{\omega: |X_n - X| > \epsilon_1\} < \varepsilon_2. \qquad (7.12)$$

Definition 7.6: A sequence of random variables $\{X_n\}$ is said to converge in distribution to a random variable X if the distribution functions $\{F_n\}$ of $\{X_n\}$ converge to the distribution function F of X at every continuity point of the latter.

Any standard text* on measure theory proves that convergence almost everywhere implies convergence in measure and in distribution. Likewise, convergence in the mean implies convergence in measure and in distribution. And finally, convergence in measure implies convergence in distribution.

We now list several theorems which will be necessary to the following treatment. The proof of these theorems can be found in the references noted in each theorem.

Theorem 7.1: If the sequence of random variables $\{X_n\}$ converge in the mean to X, then some subsequence $\{X_{n_k}\}$ converges almost everywhere** to X.

Theorem 7.2: If $\{X_n\}$ is a sequence of random variables which converges in measure to two different random variables, X and Y, then X = Y almost everywhere.***

Theorem 7.3: If $\{X_n\}$ is a mean fundamental sequence of integrable random variables, i.e., if for every $\varepsilon > 0$ there exists a real number n_o such that

$$\int_\Omega |X_n - X_m|^2 dP < \varepsilon \qquad (7.13)$$

when $m, n \geq n_o$, then there exists a square integrable random variable X

*See Ref. 30, pp. 92 and 103 and Ref. 47, p. 77.
**See Ref. 28, p. 119.
***See Ref. 30, p. 92.

such that*

$$\underset{n\to\infty}{\text{l.i.m.}}\ X_n = X .$$

Theorem 7.4: A random variable is integrable with respect to a measure P if and only if its absolute value is integrable with respect to that measure.**

Theorem 7.5: If $\{X_n\}$ is an increasing sequence of nonnegative random variables which are integrable with respect to a measure P, then***

$$\int_\Omega \lim_{n\to\infty} X_n dP = \lim_{n\to\infty} \int_\Omega X_n dP \tag{7.14}$$

This theorem specifies conditions under which the order of the limit and integration operations may be interchanged.

Theorem 7.6: (Schwartz inequality). If X_1 and X_2 are square integrable functions defined on the finite measure**** space $(\Omega, \mathcal{B}, P)$, then*****

$$\int_\Omega |X_1 X_2| dP \leq \left(\int_\Omega X_1^2 dP\right)^{1/2} \left(\int_\Omega X_2^2 dP\right)^{1/2} . \tag{7.15}$$

Schwartz's inequality gives us an upper bound on the expectation of the product of two random variables in terms of the second moments of the individual random variables.

Theorem 7.7: (Tchebycheff's inequality). If X is a random variable such that $E\{X\} = 0$ and $E\{X^2\} = \sigma_X^2 \leq \infty$, then, for every****** positive ε,

$$P\{\omega; |X| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} \sigma_X^2 \tag{7.16}$$

*See Ref. 49, p. 146.
**See Ref. 30, p. 113.
***See Ref. 30, p. 112.
****A measure space $(\Omega, \mathcal{B}, P)$ is finite if $P(\Omega) < \infty$.
*****See Ref. 30, p. 175.
******See Ref. 30, p. 200.

We are now in a positiion to give the definition of a random process.

Definition 7.7: A random process is any family of random variables $\{X_t; t\varepsilon T\}$ where T is some index set. It may also be denoted with $\{X(t,\omega); \omega\varepsilon\Omega, t\varepsilon T\}$ to emphasize the fact that two variables (t,ω) are now involved rather than one as in the case of a random variable.

The index set T is normally associated with the time of observation with the process being either a discrete parameter of continuous parameter process depending on whether the set T is an integer set or an interval of the real line. The function of $t\varepsilon T$ obtained by fixing $\omega\varepsilon\Omega$ and letting t vary is called a sample function or trajectory of the process.

In this presentation, we will be concerned chiefly with Gaussian random processes which are defined by specifying the form of the multivariate distribution functions. In particular, if, for every finite subset $t_1,\ldots,t_n$ of T,

$$F(\lambda_{t_1},\ldots,\lambda_{t_n}) = \int_{-\infty}^{\lambda t_1} \cdots \int_{-\infty}^{\lambda t_n} \frac{1}{(2\pi)^{n/2}\sqrt{|V_n|}} \exp\left\{-\frac{1}{2}(X-M_X)^T V_n^{-1}(X-M_X)\right\} dX \tag{7.17}$$

where

$$X = (X_{t_1},\ldots,X_{t_n})^T ,$$

$$M_X = E\{(X_{t_1},\ldots,X_{t_n})^T\}$$

and

$$V_n = E\{(X-M_X)(X-M_X)^T\}$$

(note that $(X-M_X)^T$ denotes the transpose of $(X-M_X)$ while V_n^{-1} and $|V_n|$ denote the inverse and determinant, respectively, of the matrix V_n), then the process will be said to be a Gaussian process. It is easily shown that such a system of distribution functions satisfy the conditions of Loève and Kolmogorov discussed above, and, as a result, there exists a probability space and a family of random variables with these multivariate distribution functions. Note also that knowledge of the first two moments* of a Gaussian process is sufficient to characterize the process. We shall, henceforth, refer to the first moment of a random process $\{X_t;t\epsilon T\}$, $\{E\{X_t\};t\epsilon T\}$, as its <u>mean value function</u> and to the joint second moment $\{E\{(X_t-E\{X_t\})(X_v-E\{X_v\})^T;t,v\epsilon T\}$, as its <u>covariance function</u>.

The specification of the multivariate distribution functions of a random process may not, however, provide sufficient information to answer all questions that might be of interest. For example, we may want to consider the ω-function defined by

$$\int_T X(t,\omega)dt$$

where $\{X_t;t\epsilon T\}$ is a random process and T is the interval $[0,T_f]$. This operation is not necessarily defined, however, since X need not be a function measurable with respect to t. Therefore in some cases, it is necessary to restrict the class processes considered. This will be done with the help of the following definition.

<u>Definition 7.8</u>: The random process $X = \{X_t;t\epsilon T\}$ is <u>measurable</u> on the interval T if $X_t(\omega) = X(t,\omega)$ defines a function measurable on the product measure space** $(\Omega xT,\mathcal{B}x\mathcal{B}_T,Pxt)$ where $\mathcal{B}_T = \mathcal{B}_R \cap T$ represents the Borel sets on the interval $[0,T_f]$ and Pxt is the corresponding product measure.

*See Ref. 43, p. 49, for a definition of moments of a random variable.
**See Ref. 30, p. 137, for a discussion of product measure spaces.

Thus, if a process is measurable, it is reasonable to consider the integral of the process with respect to t while ω is held fixed since such a function is then measurable.*

Although not every random process on the interval T is measurable, in many cases they are "nearly" measurable. To make this statement more precise, we will need the following definition and theorem.

Definition 7.9: Two random processes X and Y defined on the same probability space $(\Omega, \mathcal{B}, P)$ and the same interval T are said to be equivalent if $X_t = Y_t$ a.e.(P) for every $t \in T$.

Theorem 7.8: (Borel-Cantelli lemma). For any sequence of events $\{A_n\}$ from the σ-field $\mathcal{B}$ associated** with an arbitrary probability space $(\Omega, \mathcal{B}, P)$, one has***

$$\sum_n P(A_n) < \infty \Longrightarrow \overline{\lim_{n\to\infty}} A_n = \phi \text{ a.e.}(P) \tag{7.18}$$

where ϕ denotes the empty set.

We can now present the following theorem which specifies conditions on a given random process that guarantee the existence of an equivalent process which is measurable.

Theorem 7.9: Let $X = \{X_t; t \in T\}$ be a random process on the probability space $(\Omega, \mathcal{B}, P_1)$ for which the second moment $E_1\{X_t X_{t'}\}$ is continuous when $t, t' \in T = [0, T_f]$. Then, there exists a random process X' defined on $(\Omega, \mathcal{B}, P_1)$ which is measurable and equivalent to X. In addition, if X is also a random process on the probability space $(\Omega, \mathcal{B}, P_2)$ with $E_2\{X_t X_{t'}\}$ continuous for $t, t' \in T$, there exists a process X" which is measurable and

---

*See Ref. 30, p. 142.

**See Ref. 50, p. 128.

***$\overline{\lim_{n\to\infty}} A_n = \bigcap_n \bigcup_{m \geq n} A_m$.

equivalent to X regardless of which of the spaces, $(\Omega,\mathcal{B},P_1)$ or $(\Omega,\mathcal{B},P_2)$, is the underlying probability space.*

Proof: Applying Theorem 7.6, we can write

$$\int_\Omega |X_t-X_{t'}|dP_1 \le \left\{\int_\Omega (X_t-X_{t'})^2 dP_1\right\}^{1/2} = \{E_1\{(X_t)^2\} + E_1\{(X_{t'})^2\}-2E\{X_tX_{t'}\}\}^{1/2} \tag{7.19}$$

But, since $E_1\{X_tX_{t'}\}$ is continuous for $t,t'\varepsilon T$, for every $\varepsilon > 0$, there exists a δ such that, when $|t-t'| < \delta$, the expression on the right of Eq. (7.19) is less than ε^2. Then, if we let $B_\varepsilon = \{\omega;|X_t-X_{t'}| > \varepsilon\}$, we see that

$$\int_\Omega |X_t-X_{t'}|dP_1 > \epsilon P_1(B_\epsilon)$$

which, together with Eq. (7.19), implies

$$P_1\{\omega;|X_t-X_{t'}| > \varepsilon\} < \varepsilon \tag{7.20}$$

when $|t-t'| < \varepsilon$, i.e., $P_1\{\omega;|X_t-X_{t'}| > \epsilon\}$ goes to zero as $t' \to t$.

Now, for every integer $n > 0$, let us choose a finite increasing sequence from T, say $0 < t_1^n < \dots < t_{k_n}^n < T$, such that $P_1\{\omega;|X_u-X_v|>n^{-1}\} <2^{-n}$ if u and v belong to the same interval $[t_{i-1}^n,t_i^n]$. (Without loss of generality, we may suppose that $\{t_i^n\}_i \subset \{t_i^{n+1}\}_i$ for every n and that the countable set T' consisting of all the t_i^n is dense in T.) Then, let us define a sequence $\{X^n;n>0\}$ of mappings of $\Omega x T$ into the real line by

$$X^n(t,\omega) = X(t_i^n,\omega)$$

if $t_i^n \le t < t_{i+1}^n$. These mappings are measurable on $(\Omega x T,\mathcal{B} x \mathcal{B}_T)$. Moreover, the inequality

$$\sum_n P_1\{\omega;|X_t-X_t^n| > n^{-1}\} < \sum_n 2^{-n} < \infty$$

*This theorem is an extension of a theorem presented in Ref. 50, p. 91.

holds for every $t \epsilon T$ and, due to Theorem 7.8,

$$\overline{\lim_{n\to\infty}} \{\omega; |X_t - X_t^n| > n^{-1}\} = \phi \quad \text{a.e.}(P) .$$

But this implies that the jointly measurable function

$$\overline{\lim_{n\to\infty}} \; X^n(t,\omega)$$

equals $X(t,\omega)$ a.e. (P_1) for every $t\epsilon T$.

To prove the second part of this theorem, we need only choose the sequences $\{t_i^n\}_i$ such that the relation $P_2\{\omega; |X_u - X_v| > n^{-1}\} < 2^{-n}$ is also satisfied. The remainder of the proof then follows the construction presented above.

Q.E.D.

We can now state the following theorems which illustrate several properties of measurable processes.

Theorem 7.10: (Fubini). If X is a nonnegative measurable random process on $\Omega x T$, then

$$\int_{\Omega x T} X d(Pxt) = \int_{\Omega}\int_{T} X dt dP = \int_{T}\int_{\Omega} X dP dt \qquad (7.21)$$

where P, t and Pxt are the measures on Ω,T and $\Omega x T$, respectively. Also, if A is a measurable subset of $\Omega x T$ of measure zero, then almost every section has measure zero.* (The set $A_t = \{\omega; (\omega,t)\epsilon A\}$ is called the section of A determined by t.)

Theorem 7.11: (Fubini). If X is an integrable process on $\Omega x T$ then (7.21) is valid.**

*See Ref. 30, p. 147.
**See Ref. 30, p. 148.

The above two theorems will be used to justify the interchange of the orders of integration with respect to time and with respect to the probability measure of the sample space.

Theorem 7.12: Let $\{X_t; t\varepsilon T\}$ be a zero mean measurable Gaussian process defined on the probability space $(\Omega, \mathcal{B}, P)$ and let T be a finite closed interval of the real line. If $E\{X_t X_s\}$ is continuous for t, $s\varepsilon T$, and $\{g_j\}$ is any sequence of continuous real valued functions on T, then the sequence of random variables $\{\theta_j\}$ defined by

$$\theta_j(\omega) = \int_T X(t,\omega) g_j(t)\,dt \qquad (7.22)$$

is jointly Gaussian.*

Families of random variables which have the property of being "independent" will also play a key role in this presentation. As a result, we state the following definition.

Definition 7.10: A family $\{X_j; j\varepsilon J\}$ of random variables where J denotes an arbitrary index set is said to be independent if

$$P[\bigcap_{J'} \{\omega; X_j < a_j\}] = \prod_{J'} P\{\omega; X_j < a_j\} \qquad (7.23)$$

for every finite subset J' of J and every choice of real $a_j, j\varepsilon J'$. Note that by this definition, the distribution function of a finite subset of an independent family of random variables is the product of the distribution functions of the individual random variables of the finite subset. This implies, in the case of an independent family of Gaussian random variables which eventually enters our discussion, the density function** of any finite subset of the family if the product of the individual density functions of the finite subset.

*This Theorem is a trivial extension of the discussion presented in Ref. 43, p. 155.
**See Ref. 28, p.227.

The concept of a "conditional expectation" of a random variable relative to a σ-field $\mathcal{B}'$ will also be useful. The following defintion will serve to classify this operation.

Definition 7.11: If X is a random variable on $(\Omega, \mathcal{B}, P)$, the conditional expectation $E^{\mathcal{B}'}X$ of X with respect to $\mathcal{B}' \subset \mathcal{B}$ is a random variable on $(\Omega, \mathcal{B}', P_{\mathcal{B}'})$ such that

$$\int_B X dP = \int_B E^{\mathcal{B}'} X dP_{\mathcal{B}'} \qquad (B \varepsilon \mathcal{B}) \qquad (7.24)$$

where $P_{\mathcal{B}'}$ is the restriction of P to $\mathcal{B}'$. ($E^{\mathcal{B}'}X$ is unique* a.e.$(P_{\mathcal{B}'})$.)

This definition allows us to write the following theorem.

Theorem 7.13: Let X be a random variable of finite mean on $(\Omega, \mathcal{B}, P)$ and let $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \cdots \subset \mathcal{B}$ be σ-fields of measurable sets. Let $\mathcal{B}'$ be the minimum σ-field containing $\mathcal{B}_1, \mathcal{B}_2, \ldots$, i.e., let $\mathcal{B}' = \mathcal{B}(\mathcal{B}_1, \mathcal{B}_2, \ldots)$, then**

$$\lim_{n\to\infty} E^{\mathcal{B}_n} X = E^{\mathcal{B}'} X \qquad \text{a.e.}(P') \qquad (7.25)$$

where P' is the restriction of P to $\mathcal{B}'$.

We shall conclude this section with two additional definitions which are concerned with the mutual relationships that exist when two different measures are defined on the same σ-field and three theorems which are consequences of these definitions.

Definition 7.12: If $(\Omega, \mathcal{B}, P_1)$ and $(\Omega, \mathcal{B}, P_2)$ are two probability spaces,we say that P_2 is absolutely continuous with respect to P_1, in symbols $P_2 << P_1$, if $P_2(B) = 0$ for every measurable set B for which

*See Ref. 50, p. 121.
**See Ref. 29, p. 331.

$P_1(B) = 0$. If $P_1 << P_2$ and $P_2 << P_1$, then P_1 and P_2 are said to be equivalent, in symbols $P_1 \equiv P_2$.

Definition 7.13: If $(\Omega, \mathcal{B}, P_1)$ and $(\Omega, \mathcal{B}, P_2)$ are two probability spaces, we say that P_1 and P_2 are mutually singular, or more simply, that P_1 and P_2 are singular, in symbols $P_1 \perp P_2$, if there exists two disjoint measurable sets A and B whose union is Ω and

$$P_1(A) = P_2(B) = 0 .$$

Theorem 7.14: If $(\Omega, \mathcal{B}, P_1)$ is a probability space and if P_2 is a probability measure that is absolutely continuous with respect to P_1, then there exists a finite-valued measurable function

$$\frac{dP_2}{dP_1}$$

on Ω such that

$$P_2(B) = \int_B \frac{dP_2}{dP_1} dP_1 \tag{7.26}$$

for every measurable set B. The function dP_2/dP_1 is unique in the sense that if, also,

$$P_2(E) = \int_B f' dP_1$$

for all $B \varepsilon \mathcal{B}$, then

$$f' = \frac{dP_2}{dP_1} \quad \text{a.e.}(P_1).$$

The function dP_2/dP_1 is called the Radon-Nikodym derivative of the measure P_2 with respect* to the measure P_1.

*See Ref. 30, p. 129.

The above Radon-Nikodym derivative will prove to be an indispensable tool in the solution of the estimation problem.

Theorem 7.15: (Kakutani). Let $\{P_{1n}\}$ and $\{P_{2n}\}$ be a sequences of probability measures where, for all n, P_{1n} and P_{2n} are defined on a σ-field $\mathcal{B}_n$ of sets from a space Ω_n and $P_{1n} \equiv P_{2n}$. Then the infinite direct product measures $P_1' = \prod_{n=1}^{\infty} P_{1n}$ and $P_2' = \prod_{n=1}^{\infty} P_{2n}$ are either equivalent, $P_1' \equiv P_2'$, or mutually singular, $P_1' \perp P_2'$, according as the infinite product

$$\prod_{n=1}^{\infty} \int_{\Omega_n} \left(\frac{dP_{1n}}{dP_{2n}}\right)^{1/2} dP_{2n} \tag{7.27}$$

is greater than zero or equal to zero.*

This theorem will be applied to the case where $\{\theta_j\}$ is a sequence of random variables such that θ_j is defined on the two probability spaces, $(\Omega_j, \mathcal{B}_j, P_{oj})$ and $(\Omega_j, \mathcal{B}_j, P_{ij})$. Then, due to the above theorem, the product measures $P_o^* = \prod_{j=1}^{\infty} P_{oj}$, and $P_i^* = \prod_{j=1}^{\infty} P_{ij}$ defined on $(\prod_{j=1}^{\infty} \Omega_j, \prod_{j=1}^{\infty} \mathcal{B}_j)$ are either equivalent or mutual singular.

7.2.2 Functional Analysis

Functional analysis is the study of functions in which the individual elements of a class are considered to be points of an appropriate infinite-dimensional space. In this way, many theories of these functions can be derived as simple extensions of theories that are true for finite-dimensional Euclidean spaces.

The function spaces of interest to us can be characterized by the following definitions.

Definition 7.14: A complex (real) linear** space W is called a

*See Ref. 51, p. 295.
**See Ref. 44, p. 9 for the definition of a complex (real) linear space.

normed linear space if there exists a real-valued function on W, whose value at $w \epsilon W$ we denote by $||w||$, with the properties:

(1) $||w_1+w_2|| \leq ||w_1|| + ||w_2||$

(2) $||aw|| = |a| \, ||w||$

(3) $||w|| \geq 0$

(4) $||w|| \neq 0$ if $w \neq 0$

when w, w_1, $w_2 \epsilon W$ and a is an arbitrary complex (real) scalar.

Definition 7.15: A complex (real) linear space W is called a complex (real) inner product space if there exists a complex (real)-valued function on WxW, whose value we denote by (w_1,w_2) when w_1, $w_2 \epsilon W$, with the properties:

(1) $(w_1+w_2,w_3) = (w_1,w_3) + (w_2,w_3)$

(2) $(w_1,w_2) = (\overline{w_2,w_1})$ (the bar denoting complex conjugate)

(3) $(aw_1,w_2) = a(w_1,w_2)$

(4) $(w,w) \geq 0$ and $(w,w) \neq 0$ if $w \neq 0$

when w, w_1, w_2, $w_3 \epsilon W$ and a is an arbitrary complex (real) scalar.

Note that if we define $||w|| = (w,w)^{1/2}$, an inner product space is also a normed linear space.

One class of functions with which we will be concerned is denoted by $L_2[T]$ and includes all real-valued functions which are square integrable on the interval T, i.e., all functions f for which the Lebesgue integral

$$\int_T f^2 dt, \tag{7.28}$$

where t represents the Lebesgue measure on the interval T, is finite. For this class of functions, we can define an inner product by

$$(f,g) = \int_T fg dt \tag{7.29}$$

and it can be shown* that $L_2[T]$, with this inner product, satisfies the axioms of a real inner product space. It is also a normed linear space if we define

$$||f||^2 = \int_T f^2 dt. \tag{7.30}$$

Furthermore, this space** is complete which means that all Cauchy*** sequences in $L_2[T]$ converge in norm to an element in $L_2[T]$, i.e., if $\{f_n\}$ is a Cauchy sequence in $L_2[T]$, then there exists an element $f \varepsilon L_2[T]$ such that

$$\lim_{n \to \infty} ||f-f_n|| = 0.$$

A complete complex (real) inner product space is also called a complex (real) Hilbert space. Henceforth, we shall refer to a complex (real) Hilbert space as a Hilbert space since the theorems that will be required are true in both cases. In addition, $L_2[T]$ is separable which implies that there exists a sequence of functions $\{g_i\}$ such that

$$\begin{aligned} (g_j, g_k) &= 1 \text{ if } j=k \\ &= 0 \text{ if } j \neq k \end{aligned} \tag{7.31}$$

and such that, for any $g \varepsilon L_2[T]$,

$$\lim_{n \to \infty} \int_T |g - \sum_{j=1}^{n} (g, g_j) g_j|^2 dt = 0.$$

The sequence $\{g_n\}$ is known as an orthonormal basis for the Hilbert space. (If a sequence of functions satisfies Eq. (7.31) only, it is called an orthonormal sequence.)

*See Ref. 44, p. 108.

**See Ref. 44, p. 377.

***If $\{f_n\}$ is a Cauchy sequence, for every $\varepsilon > 0$, there exists a number n_o such that, for $n, m \geq n_o$, $||f_n - f_m|| < \varepsilon$.

If we now consider the set of functions

$$\overline{L}_2[T] = \prod_{k=1}^{K} L_2^k[T] = L_2^1[T]x \ldots xL_2^K[T] \qquad (7.32)$$

where $L_2^k[T] = L_2[T]$ for all k so that a point in $\overline{L}_2[T]$ is a vector valued function with each component of the vector being an $L_2[T]$ function, we find that $\overline{L}_2[T]$ is also a separable Hilbert space when the inner product is defined by

$$([f],[g]) = \sum_{k=1}^{K} \int_T [f]_k[g]_k \, dt. \qquad (7.33)$$

In this equation, $[f]$, $[g] \varepsilon \overline{L}_2[T]$ and $[f]_k$ and $[g]_k$ are the kth components of the vectors [f] and [g], respectively. The space $\overline{L}_2[T]$ will be the space of interest in the remainder of this chapter.

The next topic to be reviewed is concerned with transformations defined on the space $\overline{L}_2[T]$. As an example, consider the integral operator R defined on $\overline{L}_2[T]$ by

$$R([f]) = \int_T [r(t,v)] \, [f(v)]dv \qquad (7.34)$$

where $[r(t,v)]$ is a KxK matrix of functions of two variables and $[r(t,v)][f(v)]$ is the matrix product of $[r(t,v)]$ and $[f(v)]$. The properties of such a transformation when the kernel $[r(t,v)]$ is of a particular type will be discussed later. In the meantime, we shall classify the transformation to be used in the next section.

<u>Definition 7.16</u>: If W_1 and W_2 are linear spaces, then an operator R on W_1 into W_2 is called linear if the following two conditions are satisfied:

(1) $R(w_{11}+w_{12}) = R(w_{11}) + R(w_{12})$

(2) $R(aw_1) = aR(w_1)$ (7.35)

where a is an arbitrary scalar and w_1, w_{11}, $w_{12} \varepsilon W_1$.

Definition 7.17: If R is a linear operator, then the inverse of R, denoted R^{-1}, is such that

$$R(R^{-1}(w)) = w$$

and

$$R^{-1}(R(w)) = w \tag{7.36}$$

for all w. (Note that the inverse does not always exist.)

Definition 7.18: If W_1 and W_2 are normed linear spaces, a linear operator R on W_1 into W_2 is bounded if

$$\sup_{||w_1|| \leq 1} ||R(w_1)|| < \infty.$$

We define the norm of R, denoted $||R||$, by

$$||R|| = \sup_{||w_1|| \leq 1} ||R(w_1)|| \tag{7.37}$$

if it exists. (It can be shown* that the norm of R is also given by $\sup_{w_1 \neq 0} \frac{||R(w_1)||}{||w_1||}$.)

Definition 7.19: If W_1 and W_2 are normed linear spaces and R is a bounded linear operator on W_1 into W_2, it is said to be completely continuous if it has the following property: If $\{w_{1n}\}$ is any bounded sequence of elements from W_1, i.e., $||w_{1n}|| \leq M$ for some finite M and all n, then, the sequence $\{R(w_{1n})\}$ contains a convergent subsequence.

Definition 7.20: If W is an inner product space, an operator R on W into W is said to be positive if

*See Ref. 44, p. 86.

$$(R(x),w) \geq 0 \tag{7.38}$$

for every $w \epsilon W$. It is <u>positive definite</u> if

$$(R(w),w) > 0 \tag{7.39}$$

for every nonzero $w \epsilon W$.

<u>Definition 7.21</u>: If W_1 and W_2 are inner product spaces and R is a linear operator on W_1 into W_2, then the adjoint R^* is defined on W_2 into W_1 by

$$(R(w_1),w_2) = (w_1,R^*(w_2)) \tag{7.40}$$

for $w_1 \epsilon W_1$ and $w_2 \epsilon W_2$. If $W_1=W_2$ and $R^* = R$, we say that R is <u>self-adjoint</u>.

<u>Definition 7.22</u>: Let R be a positive self-adjoint operator on W into W. A square root of R, denoted $R^{1/2}$, satisfies

$$R^{1/2}(R^{1/2}(w)) = R(w) \tag{7.41}$$

for every $w \epsilon W$. (Every positive self-adjoint bounded operator possesses a unique positive self-adjoint bounded square root.*)

Having made the above definitions, we will now state several theorems which will be useful in the following development.

<u>Theorem 7.16</u>: If R is a bounded linear operator defined on a dense subset of a Hilbert space W into W, then there exists a unique bounded linear extension R' of R to all of W, i.e., R' is bounded and

$$R'(w) = R(w) \tag{7.42}$$

if $w \epsilon \mathcal{D}_R$, the domain of R.

<u>Proof</u>: If $w \epsilon W$, there exists a sequence $\{w_n\}$ such that $w_n \epsilon \mathcal{D}_R$ for each n and

*See Ref. 52, p. 62.

$$||w_n - w|| \to 0$$

since $\mathcal{D}_R$ is dense in W. Define

$$R'(w) = \lim_{n\to\infty} R(w_n) = z_1$$

which exists because R is bounded and W is complete. Then

$$\frac{||R'(w)||}{||w||} = \lim_{n\to\infty} \frac{||R(w_n)||}{||w_n||} \leq ||R||$$

since the norm is a continuous* function. Now assume that there exists another bounded extension R" such that

$$R''(w) = z_2.$$

Then

$$||z_1 - z_2|| = ||R'(w) - R''(w)|| = \lim_{m,n\to\infty} ||R(w_n) - R(w_m)|| = 0$$

and we see that R' is unique.

Q.E.D.

Theorem 7.17: If R is a bounded self-adjoint linear operator defined on a Hilbert space W into W with its null space** consisting of the zero element only, then the inverse R^{-1} is densely defined.***

Theorem 7.18: Let $R_{jk}, j,k\varepsilon\{1,2,\ldots,K\}$, be a finite set of completely continuous operators, each defined on the Hilbert space W into W. Then, the operator R defined on the product space $W^K = \prod_{k=1}^{K} W_k, W_k = W$ for all k, into W^K by

$$w_1 = R(w_2) \tag{7.43}$$

*See Ref. 44, p. 84.

**The null space of an operator R defined on W consists of all $w\varepsilon W$ such that $R(w) = 0$.

***See Ref. 44, p. 226.

where $w_1 = (w_{11}, w_{12}, \ldots, w_{1K}) \varepsilon W^K$, $w_2 = (w_{21}, w_{22}, \ldots, w_{2K}) \varepsilon W_K$, and

$$w_{1i} = \sum_{k=1}^{K} R_{ik}(w_{2k}) \qquad (7.44)$$

is completely continuous.*

Theorem 7.19: If R is the operator defined on $L_2[T]$ into $L_2[T]$ by

$$R(f) = \int_T r(t,v) f(v) dv, \qquad (7.45)$$

then the condition

$$\int_T \int_T |r(t,v)|^2 dt\, dv < \infty \qquad (7.46)$$

is sufficient to ensure that R is completely continuous.**

Thus, if the transformation defined by Eq. (7.34) has a kernel which satisfies

$$\sum_{j,k=1}^{K} \int_T \int_T |[r(t,v)]_{jk}|^2 dt\, dv < \infty \qquad (7.47)$$

where $[r(t,v)]_{jk}$ is the jkth element of $[r(t,v)]$, it is completely continuous by the above two theorems.

The next two theorems are known as the Spectral Theorems for completely continuous self-adjoint operators and bounded self-adjoint operators, respectively. In order to write these theorems, we must recall that a closed linear subspace of a Hilbert space is a linear subspace in which all Cauchy sequences converge to a limit which is in the subspace. Also, if W_j is a closed linear subspace of a Hilbert space W, every element of W can be written as the unique sum of two elements, one in W_j and one in the orthogonal complement of W_j which is the set of all $w \varepsilon W$ which satisfy $(w_j, w) = 0$ for all $w_j \varepsilon W_j$.

*See Ref. 53, p. 316.
**See Ref. 53, p. 320.

Theorem 7.20: (Completely continuous self-adjoint case). Let R be a completely continuous self-adjoint transformation defined on a Hilbert space W. Then, the null space (a closed linear subspace) of R is orthogonal to its range, i.e., if w_1 is contained in the null space of R and w_2 in its range, then $(w_1,w_2) = 0$. Furthermore, there exist a sequence of distinct real numbers $\mu_1,\mu_2,\ldots$, and an associated sequence of closed linear subspaces $W_1,W_2,\ldots$, each having nonzero dimensionality with the following properties:

(1) On W_j, $R = \mu_j I$, i.e., the vectors of W_j are the eigenvectors of R corresponding to the eigenvalues μ_j.

(2) If $j \neq k$, then $W_j \perp W_k$, i.e., $(w_j,w_k) = 0$ for any $w_j \varepsilon W_j$ and $w_k \varepsilon W_k$.

(3) The closed linear subspaces $\{W_j\}$ span the range of R in the sense that any w in the range of R can be written as the sum

$$w = w_1 + w_2 + \ldots \tag{7.48}$$

where $w_j \varepsilon W_j$ for each j, and with the series converging in norm to w.

(4) The subspaces $\{W_j\}$ each have finite dimensionality.

(5) The sequence $\{\mu_j\}$ converges to zero.

The numbers $\{\mu_j\}$ and the subspaces $\{W_j\}$ are uniquely determined by R.*

Note that the above theorem implies the existence of a sequence of functions which are eigenfunctions of R and which form an orthonormal basis for the range of R.

Theorem 7.21: (Bounded self-adjoint case). If R is a bounded self-adjoint linear transformation on a Hilbert space W and if

$$m(R) = \inf_{||w||=1} (R(w),w) \tag{7.49}$$

*See Ref. 54, p. 115 and Ref. 44, p. 336.

and

$$M(R) = \sup_{||w||=1} (R(w),w) , \qquad (7.50)$$

then there exists a family of projections $\{E_\mu\}$ defined for each real μ having the following properties:

(1) $E_{\mu_1}E_{\mu_2} = E_{\mu_2}E_{\mu_1} = E_{\mu_1}$ if $\mu_1 \leq \mu_2$,

(2) $\lim_{\mu_2 \to \mu_1+} E_{\mu_2}(w) = E_{\mu_1}(w)$,

(3) $E_\mu = 0$ if $\mu < m(R)$, $E_\mu = I$ if $M(R) \leq \mu$,

(4) $E_\mu R = RE_\mu$. (7.51)

For each w_1 and w_2 in W, $(E_\mu(w_1),w_2)$ is of bounded variation as a function of μ, and

$$(R(w_1),w_2) = \int_{\beta^1}^{\beta^2} \mu d(E_\mu(w_1),w_2). \qquad (7.52)$$

The integral is an ordinary Stieltjes integral over any interval $[\beta^1,\beta^2]$ such that $\beta^1 < m(R)$, $M(R) \leq \beta^2$. Also, the formula

$$R = \int_{\beta^1}^{\beta^2} \mu dE_\mu \qquad (7.53)$$

holds, the integral on the right being defined as the limit of Riemann-Stieltjes sums with convergence in the norm of W.*

<u>Corollary 7.1</u>: If R is a bounded self-adjoint operator on the Hilbert space W whose spectrum is <u>discrete</u>, i.e., if there exists a sequence of real numbers $\{\mu_j\}$ for which the family of projections $\{E_\mu\}$ discussed in the previous theorem has the property, for every j,

$$E_\mu(w) = E_{\mu_j}(w)$$

*See Ref. 44, p. 345.

when $w \epsilon W$ and $\mu_j \leq \mu < \mu_{j+1}$, and if

$$\lim_{j \to \infty} \mu_j = 0 \qquad (7.54)$$

with the range of $E_\mu - E_{\mu-}$ being finite for all μ, then R is completely continuous.*

The Spectral Theorem for completely continuous operators can be applied to prove the following theorem.

Theorem 7.22: (Mercer's Theorem). Let the kernel $[r(t,s)]$ of the operator R defined in Eq. (7.34) be the covariance function of a K-dimensional vector valued random process with each element of $[r(t,s)]$ being continuous and bounded on TxT. Then, there exist an orthonormal sequence of functions $\{[g_n]\}$, $[g_n] \epsilon \overline{L}_2[T]$ for each n, and a sequence of real numbers $\{\mu_n\}$ that satisfy

$$R([g_n]) = \mu_n [g_n] \qquad (7.55)$$

for all n and such that

$$[r(t,s)] = \sum_{n=1}^{\infty} \mu_n [g_n(t)] \cdot [g_n(s)]^T \qquad (7.56)$$

where the series converges uniformly for all t, $s \epsilon T$.**

This theorem gives us a "decomposition" of the kernel $[r(t,s)]$ which separates its variation with t and s into functions which are dependent on a single variable.

One additional type of operator will be of interest in the following treatment, the so-called Hilbert-Schmidt operator. The integral operator R defined on $\overline{L}_2[T]$ by (7.34) and such that

*See Ref. 55, p. 234.
**See Ref. 56, p. 215.

$$\sum_{j,k=1}^{K} \int_T \int_T |[r(t,s)]_{jk}|^2 dtds < \infty \tag{7.57}$$

is an example of such an operator (see Lemma 7.1). To characterize these operations, we make the following definitions.

Definition 7.23: If W is a separable Hilbert space, the trace of an operator R on this space is defined by

$$\text{tr}(R) = \sum_{j=1}^{\infty} (R(w_j),w_j) \tag{7.58}$$

where $\{w_j\}$ is an orthonormal basis for the Hilbert space.

Definition 7.24: The Hilbert-Schmidt norm N of an operator R on a Hilbert space W is defined by

$$N(R) = (\text{tr}(R^*R))^{1/2} = \left(\sum_{j=1}^{\infty} ||R(w_j)||^2\right)^{1/2} \tag{7.59}$$

where $\{w_j\}$ is an orthonormal basis for the space. A Hilbert-Schmidt operator, then, is any operator R for which $N(R) < \infty$.

The following theorems for the Hilbert-Schmidt operators will be necessary for the solution of the estimation problem.

Theorem 7.23: The Hilbert-Schmidt norm has the following properties:

(1) N(R) is independent of the basis used in its definition.

(2) $|\text{tr}(R_1R_2)| \leq N(R_1)N(R_2)$,

(3) $N(R_1R_2) \leq ||R_1||N(R_2)$,

$N(R_1R_2) \leq ||R_2||N(R_1)$,

(4) $||R|| \leq N(R)$, (7.60)

when R, R_1, R_2 are Hilbert-Schmidt operators.*

Theorem 7.24: Every Hilbert-Schmidt operator is also a completely continuous operator.**

*See Ref. 49, p. 1012.
**See Ref. 49, p. 1010.

It will be shown later* that the operator defined by Eq. (7.34) and satisfying (7.57) has a Hilbert-Schmidt norm equal to (7.57) and, as a result, is a Hilbert-Schmidt operator.

7.3 SOLUTION

Having stated the above mathematical preliminaries, we are now in a position to apply these ideas to the solution of the direction finding problem as discussed in Chapter 6. Let us begin by defining an observation space to be the space $\Omega = \sum_{k=1}^{K} R_T^k$ where, for each k, R_T^k is the space of all real-valued functions on the observation interval $T = [0,T_f]$ and K is the number of receiving elements in the antenna array. Due to its frequency of occurrence in the remaining chapters, we are denote the interval $[0,T_f]$ with the symbol T. Then, an <u>observation</u>, a point in the observation space, takes the form of the vector valued function $[y(t)] = (y_1(t),y_2(t),\ldots,y(t_K))^T$, $t\varepsilon T$, with $y_k(t)\varepsilon R_T^k$ and representing a possible signal received by the kth antenna of the array.

Initially, instead of allowing the angle of arrival to have any value within a specified sector, let us consider the case in which this unknown angle is limited to a finite number of specific values, $\{\alpha_1,\alpha_2,\ldots,\alpha_M\}$. In this case, a <u>solution</u> takes form of a disjoint partition $(\Lambda_0,\Lambda_1,\ldots,\Lambda_M)$ of observation space Ω with the understanding that if $y\varepsilon\Lambda_i, i\varepsilon\{1,2,\ldots,M\}$, we will make the decision that the true angle of arrival is given by α_i and if $y\varepsilon\Lambda_0$, we will make the decision noise alone is being received. An optimal solution is then defined to be the disjoint partition of Ω which satisfies the "MPE" optimality criterion discussed on Section 6.3. (It will be shown later that this solution is also optimal with respect to the "MAP" optimality criterion mentioned in Sec-

*See Lemma 7.1.

tion 6.3.) This criterion requires the minimization of the probability of error P_e, which, in this case, is defined by

$$P_e = \sum_{i=1}^{M} Pr(\alpha_i)Pr_i(\Omega-\Lambda_i) + Pr(\alpha_o)Pr_o(\Omega-\Lambda_o) \tag{7.61}$$

where $Pr(\alpha_o)$ is the a priori probability of receiving noise alone, $Pr_o(\Omega-\Lambda_o)$ is the probability of an incorrect decision being made when noise alone is present, and, for each $i\epsilon\{1,2,\ldots,M\}$, $Pr(\alpha_i)$ is the a priori probability that α_i is the true angle of arrival while $Pr_i(\Omega-\Lambda_i)$ is the probability of an incorrect decision being made when α_i is the true value of the parameter. Thus, the optimal partition of Ω is that partition of the observation space which results in a minimum probability of error.

At this point, due to their repeated occurrence in the following discussion, let us agree to denote the sets $\{1,2,\ldots,M\}$ and $\{o,1,2,\ldots,M\}$ with the shorthand notation M and M_o, respectively. The fact that we are using the symbol M to represent both the integer M and the set of integers $\{1,2,\ldots,M\}$ will not result in confusion since it will be obvious from the particular application to which quantity we are referring.

There is one obvious difficulty with the above formulation, however, and it is concerned with the interpretation of the probability $Pr_i(\Omega-\Lambda_i)$. For example, does this probability exist for all partitions of Ω and when it does exist, how do we evaluate it? To answer these questions, let us recall that $[y(t)]$, $t\epsilon T$, is a sample function of one of the vector valued random processes $[Y^i] = \{[Y^i_t];t\epsilon T\}$, $i\epsilon M_o$, as discussed in Section 6.3. (The process $[Y^i]$ should actually be denoted as $[Y^{\alpha_i}]$ according to the notation of Section 6.3 but, for convenience, we have replaced the superscript α_i by i.) Then, as a result of the discussion of Section 7.2.1, we know that there exists a probability space $(\tilde{\Omega}_i, \bar{\bar{\mathcal{B}}}_i, \bar{\bar{P}}_i)$ on which $[Y^i]$ is defined. Note that the probability measure $\bar{\bar{P}}_i$ and the σ-field $\bar{\bar{\mathcal{B}}}_i$ are dependent on the true angle of arrival since the covariance function which determines this measure and σ-field is dependent on α_i. We then see that, by $Pr_i(\Omega-\Lambda_i)$, one really means $\bar{\bar{P}}_i\{\omega;[Y^i(\omega)]\varepsilon\Omega-\Lambda_i\}$. It is now obvious that all possible partitions of Ω are not necessarily possible since all sets of the form $\{\omega;[Y^i(\omega)]\varepsilon\Omega-\Lambda_i\}$ need not be measurable. Therefore, let us now determine for which subsets of Ω, the probabilities of their inverse images under $[Y^i]$ are defined, i.e., let us determine the subsets of Ω which are possible candidates to be members of the optimal partition of Ω.

The knowledge that a particular process, say $[Y^i]$, is Gaussian with a specified covariance and mean value function provides the probabilities of the sets in the minimum σ-field with respect to which each random variable $[Y^i_t]$ is measurable when $t\varepsilon T$. Let us denote this minimum σ-field $\tilde{\mathcal{B}}_i$ and the measure defined on it by $\tilde{P}_i$. According to the discussion of Section 7.2.1, let us also make the standard construction which completes this measure so the σ-field of sets of known probabilities can be enlarged slightly, by completion, to $\bar{\tilde{\mathcal{B}}}_i$ with the associated measure being $\bar{\tilde{P}}_i$. Note that if the inverse image of a subset of Ω under $[Y^i]^{-1}$ is contained in $\bar{\tilde{\mathcal{B}}}_i$, it need not be in $\bar{\tilde{\mathcal{B}}}_j$ under $[Y^j]^{-1}$ when $i \neq j$. This detail will be considered in the next paragraph where we shall construct probability spaces which contain all the information that is necessary for obtaining an optimal partition of Ω and whose sample points are elements of Ω.

Let us now define $\mathcal{B}$ to be the σ-field of subsets of Ω generated by the class of sets $\mathcal{D}$ of the form

$$D = \{[f]\epsilon\Omega;\ [f(t_1)] \leq \underline{a}_1,\ [f(t_2)] \leq \underline{a}_2, \ldots, [f(t_n)] \leq \underline{a}_n\} \quad (7.62)$$

where n is any arbitrary finite positive integer, $t_j \epsilon T$ for $j\epsilon\{1,2,\ldots,n\}$, and $\underline{a}_j$, $j\epsilon\{1,2,\ldots,n\}$, is any real vector in K-dimensional Euclidean space. In Appendix I, it is shown that $[Y^i]^{-1}(\mathcal{B}) = \tilde{\mathcal{B}}_i$ where i is any $i\epsilon M_o$. Then, if we define a measure P_i such that

$$P_i(B) = \tilde{P}_i(\tilde{B}_i) \quad (7.63)$$

when $B\epsilon\mathcal{B}$ and $\tilde{B}_i = [Y^i]^{-1}(B)$, we see that $(\Omega, \mathcal{B}, P_i)$ is a probability space with a measure defined on $\mathcal{B}$ that is consistent with the measure on $\tilde{\mathcal{B}}_i$. Furthermore, if we complete P_i to $\overline{P}_i$ on $\overline{\mathcal{B}}_i$, we find $(\Omega, \overline{\mathcal{B}}_i, \overline{P}_i)$ possesses the desired property that every subset of Ω, whose inverse image under $[Y^i]$ is measurable with respect to $\overline{\tilde{\mathcal{B}}}_i$, is an element of $\overline{\mathcal{B}}_i$. This follows from the fact that $[Y^i]^{-1}(B_1 \cup B_2) = [Y^i]^{-1}(B_1) \cup [Y^i]^{-1}(B_2)$ for B_1, $B_2 \epsilon \mathcal{B}$. The measures of the corresponding elements of these two σ-fields are easily seen to be equal. In a similar manner, the probability spaces $(\Omega, \overline{\mathcal{B}}_i, \overline{P}_i)$, $i\epsilon M_o$, can be constructed.

To summarize the above discussion, the probability spaces $(\Omega, \overline{\mathcal{B}}_i, \overline{P}_i)$, $i\epsilon M_o$, can be constructed so as to have the property that any subset of Ω whose inverse image under $[Y^i]$ is contained in $\overline{\tilde{\mathcal{B}}}_i$ is an element of $\overline{\mathcal{B}}_i$ and the probabilities of the subset and its preimage are equal. Also, any element in $\overline{\mathcal{B}}_i$ has a preimage in $\overline{\tilde{\mathcal{B}}}_i$ of equal measure. As a result, the sets that make up the optimal partition of Ω must be elements of the appropriate σ-fields $\overline{\mathcal{B}}_i$, $i\epsilon M_o$.

It will be shown later that, in the case of most practical interest, the above measures have the property $P_i \equiv P_o$, $i\varepsilon M$, i.e., these measures are equivalent. For this case, the following theorem with $\bar{\mathcal{B}} = \bar{\mathcal{B}}_o$ gives us a clue to the construction of an optimal partition of the space Ω. Note that, if $P_i \equiv P_o$, then $\bar{P}_i << \bar{P}_o$, $\bar{P}_o << \bar{P}_i$ and $\bar{\mathcal{B}}_i = \bar{\mathcal{B}}_o$.

Theorem 7.24:* Let $(\Omega, \bar{\mathcal{B}}, \bar{P}_i)$, $i\varepsilon M_o$, be probability spaces such that $\bar{P}_i$, $i\varepsilon M_o$, is absolutely continuous with respect to $\bar{P}_o$, i.e.,

$$\bar{P}_i << \bar{P}_o, \; i\varepsilon M_o \; . \tag{7.64}$$

Let a_i, $i\varepsilon M_o$, be such that $0 \leq a_i \leq 1$ and $\sum_{i\varepsilon M_o} a_i = 1$ and let $(\Lambda_o, \Lambda_1, \ldots, \Lambda_M)$ be a partition of Ω, i.e., $\bigcup_{i\varepsilon M_o} \Lambda_i = \Omega$, $\Lambda_i \varepsilon \bar{\mathcal{B}}$ for each $i\varepsilon M_o$ and $\Lambda_i \cap \Lambda_j = \phi$ if $i \neq j$. Let $d\bar{P}_i/d\bar{P}_o$, $i\varepsilon M_o$, be the Radon-Nikodym derivatives of $\bar{P}_i$ with respect to $\bar{P}_o$, and let $(\Lambda_o^*, \Lambda_1^*, \ldots, \Lambda_M^*)$ be the partition of Ω whose elements are defined by

$$\Lambda_i^* = \{\omega; a_i \frac{d\bar{P}_i}{d\bar{P}_o}(\omega) > a_j \frac{d\bar{P}_j}{d\bar{P}_o}(\omega), j < i\}$$

$$\bigcap \{\omega; a_i \frac{d\bar{P}_i}{d\bar{P}_o}(\omega) \geq a_j \frac{d\bar{P}_j}{d\bar{P}_o}(\omega), j \geq i\} \tag{7.65}$$

for all $i\varepsilon M_o$. (Note, in writing Eq. (7.65), we have treated the subscript o as the number 0 so that $j < i$ and $j \geq i$ are meaningful.) Then, for any partition $(\Lambda_o, \Lambda_1, \ldots, \Lambda_M)$ of Ω, we have

$$\sum_{i\varepsilon M_o} a_i \bar{P}_i(\Omega - \Lambda_i^*) \leq \sum_{i\varepsilon M_o} a_i \bar{P}_i(\Omega - \Lambda_i) \; . \tag{7.66}$$

Proof: Since the Radon-Nikodym derivatives are measurable functions, it can be seen that the sets Λ_i^*, $i\varepsilon M_o$, are measurable. Moreover, we have

*The remaining theorems and lemmas of this chapter are extensions to the vector case of results originally obtained by Kadota,[36-39] Pitcher,[45] Bharucha,[46] and Root.[51]

$$\sum_{i\varepsilon M_o} a_i\overline{P}_i(\Lambda_i^*) - \sum_{i\varepsilon M_o} a_i\overline{P}_i(\Lambda_i) = \sum_{i,j\varepsilon M_o} [a_i\overline{P}_i(\Lambda_i^* \cap \Lambda_j)-a_j\overline{P}_j(\Lambda_i^* \cap \Lambda_j)]$$

$$= \sum_{i,j\varepsilon M_o} \int_{\Lambda_i^* \cap \Lambda_j} \left[a_i \frac{d\overline{P}_i}{d\overline{P}_o}(\omega) - a_j \frac{d\overline{P}_j}{d\overline{P}_o}(\omega) \right] d\overline{P}_o$$

$$\geq 0 \qquad (7.67)$$

where the first equality follows from the fact that both $(\Lambda_o^*,\Lambda_1^*,\ldots,\Lambda_M^*)$ and $(\Lambda_o,\Lambda_1,\ldots,\Lambda_M)$ are partitions of Ω and the inequality follows from the definition of the sets Λ_i^*, $i\varepsilon M_o$. Then, because of the relation

$$\sum_{i\varepsilon M_o} a_i\overline{P}_i(\Omega-\Lambda_i) = 1 - \sum_{i\varepsilon M_o} a_i\overline{P}_i(\Lambda_i), \qquad (7.68)$$

we see that (7.67) is equivalent to

$$\sum_{i\varepsilon M_o} a_i\overline{P}_i(\Omega-\Lambda_i^*) \leq \sum_{i\varepsilon M_o} a_i\overline{P}_i(\Omega-\Lambda_i). \qquad (7.69)$$

Q.E.D.

As a result, the above theorem tells us that if the measures $\overline{P}_i$, $i\varepsilon M_o$, are each absolutely continuous with respect to $\overline{P}_o$ which is implied by the statement $P_i \equiv P_o$, $i\varepsilon M_o$, then knowledge of the Radon-Nikodym derivatives $d\overline{P}_i/d\overline{P}_o$, $i\varepsilon M_o$, is sufficient to provide an optimal partition of the observation space Ω. Thus, the problem reduces to that of obtaining these derivatives. The following theorem will allow us to extend these results to the general direction finding problem in which α_i's are not limited to a finite set of values.

Theorem 7.25: Let $\overline{P}_i$ and Λ_i^*, $i\varepsilon M_o$, be as in the previous theorem but with M no longer finite and define

$$\overline{N} = \{\omega; \sup_{j\epsilon M_o'} a_j \frac{d\overline{P}_j}{d\overline{P}_o}(\omega) \neq a_i \frac{d\overline{P}_i}{d\overline{P}_o}(\omega), i\epsilon M_o'\} \qquad (7.70)$$

where $M_o' = \{0,1,2,\ldots\}$. Also, define

$$\overline{\Lambda}_o^* = \Lambda_o^* \cup \overline{N}$$

and

$$\overline{\Lambda}_i^* = \Lambda_i^*, \ i\epsilon\{1,2,\ldots\}. \qquad (7.71)$$

Then $\{\overline{\Lambda}_i^*\}$ is a partition of Ω and

$$\sum_{i\epsilon M_o'} a_i \overline{P}_i(\Omega-\overline{\Lambda}_i^*) \leq \sum_{i\epsilon M_o'} a_i \overline{P}_i(\Omega-\Lambda_i) \qquad (7.72)$$

for all partitions $(\Lambda_o, \Lambda_1, \Lambda_2, \ldots)$ of Ω.

<u>Proof</u>: Clearly, the sets Λ_i^*, $i\epsilon M_o'$, are disjoint and measurable. However, their union $\bigcup_{i\epsilon M_o'} \Lambda_i^*$ is not necessarily equal to Ω, i.e., for every $\omega\epsilon\Omega$, $\sup_{j\epsilon M_o'} a_j (d\overline{P}_j)/(d\overline{P}_o)(\omega)$ may not be achieved for any $j\epsilon M_o'$ so that ω need not belong to any Λ_j^*. The set $\overline{N}$ is this set of exceptional ω points. Since $d\overline{P}_j/d\overline{P}_o \geq 0$ a.e.$(\overline{P}_o)$, Theorem 7.5 tells us

$$\int_\Omega \sum_{j\epsilon M_o'} a_j \frac{d\overline{P}_j}{d\overline{P}_o} d\overline{P}_o = \sum_{j\epsilon M_o'} a_j \int_\Omega \frac{d\overline{P}_j}{d\overline{P}_o} d\overline{P}_o = 1.$$

Thus,

$$\sum_{j\epsilon M_o'} a_j \frac{d\overline{P}_j}{d\overline{P}_o} < \infty \qquad \text{a.e. } (\overline{P}_o)$$

and

$$\lim_{j\to\infty} a_j \frac{d\overline{P}_j}{d\overline{P}_o} \to 0 \qquad \text{a.e.}(\overline{P}_o).$$

As a result, $\overline{P}_o(\overline{N}) = 0$ and the remainder of the proof follows that of the previous theorem.

Q.E.D.

Thus, even when the true angle of arrival can achieve any one of a countable set of values, the Radon-Nikodym derivative is the key to the optimal partition of the observation space Ω if $P_i \equiv P_o$ for all $i \varepsilon M_o'$.

In order to provide a proof of the next theorem which specifies, for any arbitrary index $i \varepsilon M_o'$, the functional form of the Radon-Nikodym derivative $d\overline{P}_i/d\overline{P}_o$ as well as the conditions under which the measure P_i is equivalent to P_o, we will need the results of several lemmas. These lemmas will require the use of two auxiliary random processes, $[Z^i]$ and $[Z^o]$, defined on the probability spaces $(\Omega, \mathcal{B}, P_i)$ and $(\Omega, \mathcal{B}, P_o)$, respectively, where i is any index $i \varepsilon M_o'$ for which the particular Radon-Nikodym derivative $d\overline{P}_i/d\overline{P}_o$ is desired. Before specifying these processes, however, suppose $[Z'] = \{[Z_t'];t\varepsilon T\}$ is defined on Ω by

$$[Z_t'([f])] = [f(t)] \tag{7.73}$$

for all $[f]\varepsilon\Omega$. It can be seen that $[Z']$ is a random process on $(\Omega, \mathcal{B}, P_i)$ for each $i\varepsilon M_o'$ and that $\mathcal{B}$ is the minimum σ-field for which this is true. It can also be seen, for any finite subset $\{t_1, t_2, \ldots, t_n\} \subset T$, the distribution function of $[Z_{t_1}'], [Z_{t_2}'], \ldots, [Z_{t_n}']$ when $[Z']$ is assumed defined on $(\Omega, \mathcal{B}, P_i)$ is equal to the distribution function of $[Y_{t_1}^i], [Y_{t_2}^i], \ldots, [Y_{t_n}^i]$. It then follows, $[Z']$ is a Gaussian process with a mean value and covariance function identical to that of the process $[Y^i]$ when $[Z']$ is defined on $(\Omega, \mathcal{B}, P_i)$. (These common mean value and covariance functions were discussed in Section 6.3.) Moreover, according to Theorem 7.9, there exist measurable processes defined on $\{(\Omega, \mathcal{B}, P_i); i\varepsilon M_o'\}$ which are equivalent to $[Z']$ when defined on $\{(\Omega,\mathcal{B},P_i);i\varepsilon M_o'\}$, respectively. The equivalent processes defined on $(\Omega,\mathcal{B},P_i)$ and $(\Omega,\mathcal{B},P_o)$ are the auxiliary processes referred to above and denoted by $[Z^i]$ and $[Z^o]$, respectively. Furthermore, due again to Theorem 7.9, we can assume without loss of generality $[Z^i(t,[f])] = [Z^o(t,[f])]$ for all $t\varepsilon T$ and $[f]\varepsilon\Omega$. (Even though $[Z^i]$ and $[Z^o]$ have identical values on the probability spaces, the separate superscripts will be retained to distinguish the underlying probability measure.) The equivalences mentioned above can then be applied to show $[Z^i]$ and $[Y^i]$ as well as $[Z^o]$ and $[Y^o]$ have identical mean value and covariance functions. Finally, let us define $\mathcal{B}'$ to be the minimum σ-field with respect to which $[Z_t^i]$ and $[Z_t^o]$ are measurable for all $t\varepsilon T$. It then follows, again due to the above equivalences, the elements generating the σ-field $\mathcal{B}'$ (and, hence, the elements of $\mathcal{B}'$) differ from the elements generating $\mathcal{B}$(and, hence, the elements of $\mathcal{B}$) by sets of measure zero with respect to $\overline{P}_i$ and $\overline{P}_o$, i.e., for any element in the generator of $\mathcal{B}'$ (and, hence in $\mathcal{B}'$), there exists an element in the generator of $\mathcal{B}$(and, hence, in $\mathcal{B}$) for which the symmetric difference of these two sets is a subset of a set in $\mathcal{B}$ whose P_i measure is zero and similarly for P_o. Thus, for any element $B\varepsilon\mathcal{B}$, there exists sets B_1', $\varepsilon\mathcal{B}'$, $B_2'\varepsilon\mathcal{B}'$, $N_1\subset N_1^i\varepsilon\mathcal{B}$ for which $P_i(N_1^i) = 0$, and $N_2\subset N_2^o\varepsilon\mathcal{B}$ for which $P_o(N_2^o) = 0$ such that $B = B_1'\Delta N_1$ and $B = B_2'\Delta N_2$. But this implies, defining $\hat{\mathcal{B}}'$ to be the class of sets of the form $B'\Delta N$ where $B'\varepsilon\mathcal{B}'$ and $N\subset N_i\varepsilon\mathcal{B}$, $N\subset N_o\varepsilon\mathcal{B}$, with $P_i(N_i) = P_o(N_o) = 0$,

$$\mathcal{B}\subset\hat{\mathcal{B}}' . \tag{7.74}$$

In the following lemmas, we will also have need of the integral operators R_i, $i\varepsilon M_o'$, defined on $\overline{L}_2[T]$ by

$$R_i([f]) = \int_T [r^i(t,v)][f(v)]dv \qquad (7.75)$$

where $[r^i(t,v)]$ is the matrix covariance function common to the processes $[Y^i]$ and $[Z^i]$ and $[f(v)]$ is any vector valued function from $\overline{L}_2[T]$ which was discussed in Section 7.2.2. (Note that $[r^i(t,v)][f(v)]$ denotes the matrix multiplication of $[r^i(t,v)]$ and $[f(v)]$.) The properties of these operators which will be of interest to us are summarized in the next lemma.

Lemma 7.1: The integral operator R_i defined in Eq. (7.75) takes elements of $\overline{L}_2[T]$ into elements of $\overline{L}_2[T]$ and is self-adjoint, positive definite and Hilbert-Schmidt.

Proof: See Appendix II for a proof of this lemma.

Note that since R_i is positive definite, self-adjoint and completely continuous, there exists a unique, positive, self-adjoint, bounded, square root operator $R_i^{1/2}$ defined on $\overline{L}_2[T]$ (see Definition 7.22). In addition,

$$(R_i^{1/2}([f]),R_i^{1/2}([f])) = (R_i([f]),[f]) > 0 \qquad (7.76)$$

for all $[f] = 0$ so that $R_i^{1/2}$ has a zero null space. Then, due to Theorem 7.17, $R_i^{-1/2} = (R_i^{1/2})^{-1}$ is densely defined.

The next three lemmas will be concerned with the relationships that exist between different probability measures defined on identical σ-fields. In particular, we shall be concerned with the σ-field $\mathcal{B}$ defined in Eq. (7.62) and the two probability measures, P_o and P_i, induced on $\mathcal{B}$ by the random processes $[Y^o]$ and $[Y^i]$ as defined in Eq. (7.63).

Lemma 7.2: Let $\hat{\mathcal{B}}$ be a σ-field of subsets of Ω such that

$$\hat{\mathcal{B}} \subset \mathcal{B}, \qquad (7.77)$$

and let $\hat{P}_o$ and $\hat{P}_i$ be the restrictions of P_o and P_i to $\hat{\mathcal{B}}$. Then,

$$\hat{P}_o \perp \hat{P}_i \Longrightarrow P_o \perp P_i \ . \tag{7.78}$$

<u>Proof</u>: If $\hat{P}_o \perp \hat{P}_i$, Definition 7.12 says that there exist disjoint sets $A, B \varepsilon \hat{\mathcal{B}}$ such that $A \cup B = \Omega$ and

$$\hat{P}_o(A) = \hat{P}_i(B) = 0.$$

But, due to (7.77), $A, B \varepsilon \mathcal{B}$ and, since $\hat{P}_o$ and $\hat{P}_i$ are restrictions of P_o and P_i,

$$P_o(A) = P_i(B) = 0$$

which implies $P_o \perp P_i$.

Q.E.D.

<u>Lemma 7.3:</u> Let $\hat{\mathcal{B}}$ be a σ-field of subsets of Ω such that $\hat{\mathcal{B}} \subset \mathcal{B}$, and define $\hat{\hat{\mathcal{B}}}$ to be the class of subsets of Ω of the form $\hat{B} \Delta N$ where $\hat{B} \varepsilon \hat{\mathcal{B}}$, $N \subset N_o \varepsilon \mathcal{B}$, $N \subset N_i \varepsilon \mathcal{B}$ and $P_o(N_o) = P_i(N_i) = 0$. (Note that $\hat{\hat{\mathcal{B}}}$ is contained in $\bar{\mathcal{B}}_o$ and $\bar{\mathcal{B}}_i$.) Let $\hat{P}_o$ and $\hat{P}_i$ again be the restrictions of P_o and P_i to $\hat{\mathcal{B}}$ and assume

$$\hat{P}_o \equiv \hat{P}_i \tag{7.79}$$

and

$$\mathcal{B} \subset \hat{\hat{\mathcal{B}}} \ . \tag{7.80}$$

Then,

(1) $P_o \equiv P_i$

(2) $\bar{\mathcal{B}}_o = \bar{\mathcal{B}}_i = \hat{\hat{\mathcal{B}}}$

(3) $d\bar{P}_i/d\bar{P}_o = d\hat{P}_i/d\hat{P}_o \quad$ a.e.$(\bar{P}_o)$.

Proof: It follows from (7.80) and the definition of $\hat{\mathcal{B}}$ that, for any $B \epsilon \mathcal{B}$, there exist sets $\hat{B}$ and N such that $\hat{B} \epsilon \hat{\mathcal{B}}$, $N \epsilon \overline{\mathcal{B}}_o$, $N \epsilon \overline{\mathcal{B}}_i$, $\overline{P}_o(N) = \overline{P}_i(N) = 0$ and

$$B = \hat{B} \Delta N .$$

Then if $P_o(B) = 0$, we have

$$P_o(B) = \overline{P}_o(B) = \overline{P}_o(\hat{B}) = \hat{P}_o(\hat{B}) = \hat{P}_i(\hat{B}) = \overline{P}_i(\hat{B}) = \overline{P}_i(B) = P_i(B)$$

because of (7.79) which implies $P_o << P_i$. Conversely, it can be shown that $P_i << P_o$ so the assertion of (1) follows.

To prove (2), note that (1) implies $\overline{\mathcal{B}}_o = \overline{\mathcal{B}}_i$. Moreover, it follows from (1) and (7.80) that

$$\overline{\mathcal{B}}_o \subset \hat{\hat{\mathcal{B}}} .$$

But, from the definition of $\hat{\hat{\mathcal{B}}}$, we have $\hat{\hat{\mathcal{B}}} \subset \overline{\mathcal{B}}_o$ so that

$$\overline{\mathcal{B}}_o = \hat{\hat{\mathcal{B}}} .$$

To prove (3), let $B \epsilon \overline{\mathcal{B}}_o$ and write

$$\int_B \frac{d\hat{P}_i}{d\hat{P}_o} d\overline{P}_o = \int_{\hat{B}\Delta N} \frac{d\hat{P}_i}{d\hat{P}_o} d\overline{P}_o = \int_{\hat{B}} \frac{d\hat{P}_i}{d\hat{P}_o} d\hat{P}_o = \hat{P}_i(\hat{B}) = \overline{P}_i(\hat{B}\Delta N)$$

where $B = \hat{B}\Delta N$ with $\hat{B} \epsilon \hat{\mathcal{B}}$ and $N \epsilon \overline{\mathcal{B}}_o$, $N \epsilon \overline{\mathcal{B}}_i$ with $\overline{P}_o(N) = \overline{P}_i(N) = 0$. It then follows from Theorem 7.14 that

$$\frac{d\hat{P}_i}{d\hat{P}_o} = \frac{d\overline{P}_i}{d\overline{P}_o} \qquad \text{a.e.}(\overline{P}_o) .$$

Q.E.D.

Lemma 7.4: Let $\{\theta_j; j\varepsilon M'\}$, $M' = \{1,2,\ldots\}$, be a sequence of jointly Gaussian random variables defined on the probability spaces $(\Omega,\mathcal{B},P_o)$ and $(\Omega,\mathcal{B},P_i)$ such that

$$E_o\{\theta_j\} = E_i\{\theta_j\} = 0$$
$$E_o\{\theta_j \; \theta_k\} = \gamma_j\delta_{jk}$$
$$E_i\{\theta_j \; \theta_k\} = \beta_j\delta_{jk}$$

where E_o and E_i denote the expectations with respect to P_o and P_i, respectively, δ_{jk} is the Kronecker delta, and $\gamma_j, \beta_j, j\varepsilon M'$, are arbitrary positive numbers. Let $\hat{\mathcal{B}}$ be the minimum σ-field with respect to which all the θ_j's are measurable, and let $\hat{P}_o$ and $\hat{P}_i$ be the restrictions of P_o and P_i on $\hat{\mathcal{B}}$. Then,

(1) either $\hat{P}_o \equiv \hat{P}_i$ or $\hat{P}_o \perp \hat{P}_i$,

(2) $\hat{P}_o \equiv \hat{P}_i$ iff

$$\sum_{j=1}^{\infty} \left(1 - \frac{\gamma_j}{\beta_j}\right)^2 < \infty ,$$

(3) if $\hat{P}_o \equiv \hat{P}_i$,

$$\frac{dP_i}{dP_o} = \exp\left\{\sum_{j=1}^{\infty}\left[\frac{1}{2}\left(\frac{1}{\gamma_j} - \frac{1}{\beta_j}\right)\theta_j^2 + \frac{1}{2}\log\frac{\gamma_j}{\beta_j}\right]\right\} \qquad \text{a.e.}(P_o) \; .$$

Proof: Let each $\mathcal{B}_j$, $j\varepsilon M' = \{1,2,\ldots,\}$, be the minimum σ-field with respect to which the random variable θ_j is measurable. Then, define Θ to be the class of sets of the form $B_{j_1} \cap B_{j_2} \cap \ldots \cap B_{j_n}$ where $B_{j_k} \varepsilon \mathcal{B}_{j_k}$, $k\varepsilon\{1,2,\ldots,n\}$, n is an arbitrary finite positive integer, and $j_k \varepsilon M'$. (Note that $\mathcal{B}(\Theta) = \hat{\mathcal{B}}$ so $\hat{P}_o$ and $\hat{P}_i$ are probability measures on $\mathcal{B}(\Theta)$.) Now consider the product probability* spaces $(\Omega^*,\mathcal{B}^*,P_o^*)$ and $(\Omega^*,\mathcal{B}^*,P_i^*)$

*See Ref. 30, p. 157.

where $\Omega^* = \prod_{j=1}^{\infty} \Omega_j, \Omega_j = \Omega$, $\mathcal{B}^* = \prod_{j=1}^{\infty} \mathcal{B}_j$ and $P_m^* = \prod_{j=1}^{\infty} P_{mj}, m\epsilon\{o,i\}$ with P_{mj} being the restriction of P_m to B_j. Remember, the σ-field $\mathcal{B}^*$ is generated by the class of sets $\mathcal{D}^*$ of the form $(B_1 x B_2 x \ldots)$ where $B_j \epsilon \mathcal{B}_j$, $j\epsilon M'$, and all but a finite number of the B_j equal Ω. Also, the product measure P_m^* is uniquely defined by specifying, for each set $(B_1 x B_2 x \ldots)$ in $\mathcal{D}^*$,

$$P_m^*(B_1 x B_2 x \ldots) = \hat{P}_m(B_{j_1}) \cdot \hat{P}_m(B_{j_2}) \ldots \hat{P}_m(B_{j_n})$$

where $B_{j_1}, B_{j_2}, \ldots, B_{j_n}$ are the sets $B_{j_k} \epsilon \mathcal{B}_{j_k}$, $j_k \epsilon M'$, which are not equal to Ω. We can now see, since the random variables $\{\theta_j\}$ are independent, there exists a natural mapping ζ from $\mathcal{D}^*$ onto $\mathcal{D}$

$$\zeta(B_1 x B_2 x \ldots) = B_{j_1} \cap B_{j_2} \cap \ldots \cap B_{j_n} ,$$

with $(B_1 x B_2 x \ldots)$ as described above, which preserves measures, i.e.,

$$P_m^*(B_1 x B_2 x \ldots) = \hat{P}_m(B_{j_1} \cap B_{j_2} \cap \ldots \cap B_{j_n})$$

for $M\epsilon\{o,i\}$. Then, since the probability* measure of the limit of an increasing sequence of sets is equal to the limit of the probabilities of the sets and since all sets** in $\mathcal{B}(\mathcal{D})$ and $\mathcal{B}(\mathcal{D}^*)$ are disjoint unions of sets in $\mathcal{D}$ and $\mathcal{D}^*$, respectively, it follows that $\hat{P}_o \equiv \hat{P}_i$ if $\hat{P}_o^* \equiv \hat{P}_i^*$ and $P_o \perp P_i$ if $P_o^* \perp P_i^*$. Assertion (1) is now an immediate consequence of Theorem 7.15.

Since P_{oj} and P_{ij} are Gaussian, we can write

$$\frac{dP_{ij}}{dP_{oj}} = \left(\frac{\gamma_j}{\beta_j}\right)^{1/2} \exp\left[\frac{1}{2}\left(\frac{1}{\gamma_j} - \frac{1}{\beta_j}\right) \theta_j^2\right] .$$

---

*See Ref. 30, p. 38.
**See Ref. 57, p. 224.

Then

$$\prod_{j=1}^{\infty} \int_{\Omega} \left(\frac{dP_{ij}}{dP_{oj}} \right)^{1/2} dP_{oj} = \prod_{j=1}^{\infty} \frac{(4\gamma_j\beta_j)^{1/4}}{(\gamma_j+\beta_j)^{1/2}}$$

which converges to a positive number if

$$\prod_{j=1}^{\infty} \frac{4\gamma_j\beta_j}{(\gamma_j+\beta_j)^2} \tag{7.81}$$

converges to a positive number. But (7.81) converges* to a positive number if and only if

$$\sum_{j=1}^{\infty} \left[1 - \frac{4\gamma_j\beta_j}{(\gamma_j+\beta_j)^2} \right] < \infty \tag{7.82}$$

and, since

$$1 - \frac{4\gamma_j\beta_j}{(\gamma_j+\beta_j)^2} = \frac{(1-\gamma_j/\beta_j)^2}{(1+\gamma_j/\beta_j)^2} ,$$

the inequality of (7.82) is satisfied if and only if

$$\sum_{j=1}^{\infty} (1-\gamma_j/\beta_j)^2 < \infty.$$

Assertion (2) then also follows from Theorem 7.15.

To prove (3), let $\mathcal{B}^n = \mathcal{B}(\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n)$ and note, for any $B_n \varepsilon \mathcal{B}^n$, there exists a B_n' which is a Borel set in n-dimensional Euclidean space such that

$$\hat{P}_i(B_n) = \hat{P}_i^n(B_n) = \int_{B_n'} p_i^n(\theta_1, \theta_2, \ldots, \theta_n) d\theta_1 d\theta_2 \ldots d\theta_n \tag{7.83}$$

*See Ref. 58, p. 381.

where $\hat{P}_i^n$ is the restriction of $\hat{P}_i$ to $\mathcal{B}^n$ and $p_i^n(\theta_1,\theta_2,\ldots,\theta_n)$ is the joint density, with respect to the probability measure P_i, of the Gaussian random variables $\theta_1,\theta_2,\ldots,\theta_n$. (See the discussion concerning Eqs. (7.3) and (7.11).) Since a similar expression can be written for $\hat{P}_o(B_n)$, (7.83) is also given by

$$\hat{P}_i(B_n) = \int_{B'_n} \frac{p_i^n(\theta_1,\theta_2,\ldots,\theta_n)}{p_o^n(\theta_1,\theta_2,\ldots,\theta_n)}\; p_o^n(\theta_1,\theta_2,\ldots,\theta_n)d\theta_1\, d\theta_2 \ldots d\theta_n$$

$$= \int_{B_n} p^n(\theta_1,\theta_2,\ldots,\theta_n)d\hat{P}_o^n$$

where $p_o^n(\theta_1,\theta_2,\ldots,\theta_n)$ is the joint density, with respect to the probability measure P_o, of $\theta_1,\theta_2,\ldots,\theta_n$ and $p^n = p_i^n/p_o^n$. However, due to Theorem 7.14 and Definition 7.11, we can also write

$$\hat{P}_i(B_n) = \int_{B_n} E_o^{\mathcal{B}^n}\left(\frac{d\hat{P}_i}{d\hat{P}_o}\right) d\hat{P}_o^n$$

where $E_o^{\mathcal{B}^n}(d\hat{P}_i/d\hat{P}_o)$ is uniquely defined a.e. $(\hat{P}_o^n)$ so that

$$E_o^{\mathcal{B}^n} \frac{d\hat{P}_i}{d\hat{P}_o} = p^n \qquad \text{a.e. } (\hat{P}_o^n).$$

(The subscript of $E_o^{\mathcal{B}^n}$ denotes the fact that the underlying probability measure is P_o.) Then, by Theorem 7.13, we have

$$\lim_{n\to\infty} p^n = \lim_{n\to\infty} E_o^{\mathcal{B}^n} \frac{d\hat{P}_i}{d\hat{P}_o} = \frac{d\hat{P}_i}{d\hat{P}_o} \qquad \text{a.e. } (\hat{P}_o) \tag{7.84}$$

since $d\hat{P}_i/d\hat{P}_o$ is measurable with respect to $\hat{\mathcal{B}}$. Assertion (3) now follows from (7.84) and the fact that

$$\frac{p_i^n(\theta_1,\theta_2,\cdots,\theta_n)}{p_o^n(\theta_1,\theta_2,\cdots,\theta_n)} = \exp\left\{\sum_{j=1}^{n}\left[\frac{1}{2}\left(\frac{1}{\gamma_j}-\frac{1}{\beta_j}\right)\theta_j^2 + \frac{1}{2}\log\frac{\gamma_j}{\beta_j}\right]\right\}$$

Q.E.D.

The following lemma shows that the "inner product" of the measurable random processes $[Z^i]$ and $[Z^o]$ discussed previously and any $\overline{L}_2[T]$ function results in a Gaussian random variable.

Lemma 7.5: If $[Z^i]$ and $[Z^o]$ are the K-dimensional, zero mean, measurable Gaussian processes discussed following Theorem 7.25 and $\{[g_j]\}$ is any sequence of elements from $\overline{L}_2[T]$, then the sequences of random variables $\{\theta_j^i\}$ and $\{\theta_j^o\}$ defined for all $[f]\epsilon\Omega$ by

$$\theta_j^i([f]) = \int_T [Z^i(t,[f])]^T[g_j(t)]dt$$

and

$$\theta_j^o([f]) = \int_T [Z^o(t,[f])]^T[g_j(t)]dt$$

are jointly Gaussian on $(\Omega,\mathscr{B},P_i)$ and $(\Omega,\mathscr{B},P_o)$, respectively. (Note that $\theta_j^i([f]) = \theta_j^o([f])$ for all $[f]\epsilon\Omega$ since $[Z^i(t,[f])] = [Z^o(t,[f])]$ for all $t\epsilon T$ and $[f]\epsilon\Omega$.)

Proof: See Appendix III for a proof of this lemma.

The next lemma specifies a condition, in terms of the integral operators defined in Eq. (7.75), under which the measures P_o and P_i are mutually singular, i.e., $P_o \perp P_i$.

Lemma 7.6: Let R_o and R_i be integral operators of the type defined by Eq. (7.75) and let $[Z^o]$ and $[Z^i]$ be the random processes referred to above. If either $R_i^{1/2}R_o^{-1/2}$ or $R_o^{1/2}\, R_i^{-1/2}$ is unbounded, then $P_o \perp P_i$.

Proof: Note that $R_i^{1/2}R_o^{-1/2}$ and $R_o^{1/2}R_i^{-1/2}$ are densely defined operators on $\overline{L}_2[T]$ and suppose $R_i^{1/2}R_o^{-1/2}$ is unbounded. Then, there exists a sequence of $\overline{L}_2[T]$ functions $\{[f_j]\}$ in the domain of $R_o^{-1/2}$ satisfying $||[f_j]|| = 1$ and $||R_i^{1/2}R_o^{-1/2}([f_j])|| \geq j^3$. Let

$$\theta_j^i([f]) = \frac{1}{j} \int_T [Z^i(t,[f])]^T [R_o^{-1/2}([f_j])](t)dt$$

and

$$\theta_j^o([f]) = \frac{1}{j} \int_T [Z^o(t,[f])]^T [R_o^{-1/2}([f_j])](t)dt$$

be random variables defined $(\Omega,\mathcal{B},P_i)$ and $(\Omega,\mathcal{B},P_o)$, respectively. (Note that these random variables are Gaussian due to the previous lemma.) Then, since $\theta_j^o([f]) = \theta_j^i([f])$ for every $[f]\varepsilon\Omega$, it can be seen that the function defined on Ω by

$$\theta_j([f]) = \theta_j^o([f]) = \theta_j^i([f])$$

is a Gaussian random variable on both $(\Omega,\mathcal{B},P_i)$ and $(\Omega,\mathcal{B},P_o)$. Moreover,

$$E_o\{\theta_j^2\} = \frac{1}{j^2}(R_o(R_o^{-1/2}[f_j]),R_o^{-1/2}[f_j]) = \frac{1}{j^2} \qquad (7.85)$$

and

$$E_i\{\theta_j\} = \frac{1}{j^2}(R_i(R_o^{-1/2}[f_j]),R_o^{-1/2}[f_j]) \geq \frac{1}{j^2} \qquad (7.86)$$

where E_o and E_i denote expectations with respect to the measures P_o and P_i. (To obtain (7.85) and (7.86), the order of integration with respect to the probability measure and the time variable was interchanged. This interchange can be justified by an argument similar to that used to justify the interchange in Eq. (II.4).) But, by Theorem 7.7, for any $\varepsilon > 0$, we have

$$P_o\{[f];\ |\theta_j([f])| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2 j^2}$$

so that,* by Theorem 7.8,

$$P_o\{[f];\ \overline{\lim_{j\to\infty}}\ |\theta_j([f])| \geq \varepsilon\} = 0$$

which implies

$$P_o\{[f];\ \overline{\lim_{j\to\infty}}\ |\theta_j([f])| = 0\} = 1.$$

Also, since each θ_j is Gaussian with respect to P_i by Lemma 7.5, for any finite positive n_o, we have

$$P_i\{[f];\ |\theta_j([f])| \leq n_o\} \leq \frac{1}{\sqrt{2\pi j^4}} \int_{-n_o}^{n_o} dx$$

$$= \frac{2n_o}{j^2\sqrt{2\pi}}$$

and, again by Theorem 7.8,

$$P_i\{[f];\ \overline{\lim_{j\to\infty}}\ |\theta_j([f])| > n_o\} = 1$$

which implies

$$P_i\{[f];\ \overline{\lim_{j\to\infty}}\ |\theta_j([f])| = \infty\} = 1.$$

Thus, if we let

$$A = \{[f];\ \overline{\lim_{j\to\infty}}\ |\theta_j([f])| = 0\},$$

we obtain

$$P_o(A) = P_i(\Omega - A) = 1$$

---

* $\overline{\lim_{j\to\infty}}\ |\theta_j([f])| = \inf_k \sup_{j\geq k} |\theta_j([f])|$.

and $P_o \perp P_i$.

A similar argument can be used to prove $P_o \perp P_i$ in the case where $R_o^{1/2}R_i^{-1/2}$ is assumed to be unbounded.

Q.E.D.

The following lemma will be necessary to the proof of Lemma 7.8.

Lemma 7.7: Let $\{[f_j]\}$ be a sequence of functions from $\overline{L}_2[T]$ and, for every $[f_j]$ in this sequence, suppose $\{[f_{jk}]\}$ is another sequence of functions from the domain of $R_o^{-1/2}$ such that

$$\lim_{k\to\infty} ||[f_j] - [f_{jk}]|| = 0. \qquad (7.87)$$

Then, if $R_i^{1/2}R_o^{-1/2}$ is bounded and if we defined a sequence of random variables $\{\theta_{jk}\}$ on Ω by

$$\theta_{jk} = \int_T [Z^i(t,[f])]^T[R_o^{-1/2}([f_{jk}])](t)dt = \int_T [Z^o(t,[f])]^T[R_o^{-1/2}([f_{jk}])](t)dt,$$

there exists a sequence of random variables $\{\theta_j\}$ which are measurable with respect to $\mathcal{B}$, jointly Gaussian with respect to P_o and P_i, possess finite variances, and such that

$$\theta_j = \underset{k\to\infty}{\text{l.i.m.}}\ \theta_{jk} \qquad (P_o,P_i).$$

Proof: According to Lemma 7.5, the random variables $\{\theta_{jk};j=1,2,\ldots\}$ are jointly Gaussian on both $(\Omega,\mathcal{B},P_i)$ and $(\Omega,\mathcal{B},P_o)$. In addition, we have

$$\begin{aligned} E_o\{\theta_{jk}\theta_{jn}\} &= (R_o^{-1/2}[f_{jk}],R_o(R_o^{-1/2}[f_{jn}])) \\ &= ([f_{jk}],[f_{jn}]). \end{aligned}$$

Again, the order of integration has been interchanged which is justifiable by an argument similar to that used to justify the interchange in Eq.

(II.4). Hence, we find

$$\lim_{k,n\to\infty} E_o\{|\theta_{jk}-\theta_{jn}|^2\} = \lim_{k,n\to\infty} \{||[f_{jk}]||^2 + ||[f_{jn}]||^2 - 2([f_{jk}],[f_{jn}])\} = 0 \tag{7.88}$$

since $\lim_{k\to\infty} ||[f_{jk}]||^2 = ||[f_j]||^2$ and $\lim_{k,n\to\infty} ([f_{jk}],[f_{jn}]) = ||[f_j]||^2$ as a result of (7.87). Also,

$$\begin{aligned} E_i\{\theta_{jk}\theta_{jn}\} &= (R_o^{-1/2}[f_{jk}],R_iR_o^{-1/2}[f_{jn}]) \\ &= (R_i^{1/2}R_o^{-1/2}[f_{jk}],R_i^{1/2}R_o^{-1/2}[f_{jn}]) \end{aligned}$$

and, because $R_i^{1/2}R_o^{-1/2}$ is bounded and therefore continuous,

$$\lim_{k,n\to\infty} E_i\{|\theta_{jk}-\theta_{jn}|^2\} = 0 . \tag{7.89}$$

Then, combining (7.88) and (7.89), we see that $\{\theta_{jk}\}_k$, for each j, is a mean fundamental sequence with respect to both P_o and P_i and, hence, with respect to the measure P_o+P_i. If Theorem 7.3 is now applied, we see that there also exists a sequence of random variables $\{\theta_j\}$, each of which is measurable with respect to $\mathcal{B}$ and square integrable on $(\Omega,\mathcal{B},P_o+P_i)$, such that

$$\theta_j = \underset{k\to\infty}{\text{l.i.m.}}\ \theta_{jk} \qquad (P_o+P_i)$$

which implies

$$\theta_j = \underset{k\to\infty}{\text{l.i.m.}}\ \theta_{jk} \qquad (P_o,P_i).$$

Then, since the integral of a positive function over $(\Omega,\mathcal{B},P_o+P_i)$ is equal to or greater than its integral over $(\Omega,\mathcal{B},P_o)$ and $(\Omega,\mathcal{B},P_i)$, we see that each θ_j is square integrable over $(\Omega,\mathcal{B},P_o)$ and $(\Omega,\mathcal{B},P_i)$. Moreover, since $E_o\{\theta_j\} = 0$ and $E_i\{\theta_j\} = 0$, θ_j has a finite variance with respect to both P_o and P_i.

Finally, since the sequence $\{\theta_{jk}\}_k$ converges in the mean to θ_j with respect to both measures, it converges in measure to θ_j with respect to both measures. This implies,* the characteristic function, with respect to either measure, of $\{\theta_{j_1},\theta_{j_2},\ldots,\theta_{j_n}\}$ is the limit of the corresponding characteristic function of the jointly Gaussian variables $\{\theta_{j_1k},\theta_{j_2k},\ldots,\theta_{j_nk}\}$ when $j_m \epsilon \{1,2,\ldots\}$ for $m \epsilon \{1,2,\ldots,n\}$ and n is an arbitrary positive integer. Then, since the mean values and covariances of these random variables approach finite limits as $k \to \infty$, it follows from the general form of Gaussian characteristic functions that the limiting characteristic function is also Gaussian.

Q.E.D.

<u>Lemma</u> <u>7.8</u>: If $R_i^{1/2}R_o^{-1/2}$ is bounded, then, for any sequence of functions $\{[f_j]\}$ from $\overline{L}_2[T]$, there exists a corresponding sequence of jointly Gaussian variables $\{\theta_j\}$, defined over Ω and measurable with respect to $\mathscr{B}$, such that

$$\begin{aligned} E_o\{\theta_j\} &= E_i\{\theta_j\} = 0 \\ E_o\{\theta_j\theta_k\} &= ([f_j],[f_k]) \\ E_i\{\theta_j\theta_k\} &= ([f_j],X^*X[f_k]) \end{aligned} \tag{7.90}$$

where X is the bounded extension of $R_i^{1/2}R_o^{-1/2}$ to the whole of $\overline{L}_2[T]$ and X* is the adjoint of X.

<u>Proof</u>: Since $R_o^{-1/2}$ is densely defined on $\overline{L}_2[T]$, there exists a sequence of $\overline{L}_2[T]$ functions $\{[f_{jk}]\}_k$, for each j, such that

*See Ref. 28, p. 169.

$$\lim_{k\to\infty} ||[f_j] - [f_{jk}]|| = 0.$$

This, in turn, implies

$$\lim_{m,n\to\infty} ([f_{jm}],[f_{kn}]) = ([f_j],[f_k]) \tag{7.91}$$

since* the inner product is a continuous function on $\overline{L}_2[T]x\overline{L}_2[T]$. Moreover, we can define, for each j, the sequence of functions $\{\theta_{jk}\}$ on Ω which satisfy

$$\begin{aligned}\theta_{jk} &= \int_T [Z^i(t,[f])]^T[R_o^{-1/2}([f_{jk}])](t)dt \\ &= \int_T [Z^o(t,[f])]^T[R_o^{-1/2}([f_{jk}])](t)dt.\end{aligned}$$

Then, by Lemma 7.7, there exists a sequence of random variables $\{\theta_j\}$ defined on Ω, each of which is measurable with respect to $\mathcal{B}$, jointly Gaussian with respect to P_o and P_i, square integrable on $(\Omega,\mathcal{B},P_o)$ and $(\Omega,\mathcal{B},P_i)$, and such that

$$\theta_j = \underset{k\to\infty}{\text{l.i.m.}}\, \theta_{jk} \qquad (P_o,P_i).$$

Let us now show that the sequence $\{\theta_j\}$ satisfies (7.90). Since $[Z^o]$ is a zero mean process and

$$\begin{aligned}\left|\int_\Omega (\theta_j-\theta_{jk})dP_o\right| &\le \int_\Omega |\theta_j-\theta_{jk}|dP_o \\ &\le \left\{\int_\Omega |\theta_j-\theta_{jk}|^2 dP_o\right\}^{1/2}\end{aligned}$$

where the last inequality follows from Theorem 7.6, we have

$$E_o\{\theta_j\} = \lim_{k\to\infty} E_o\{\theta_{jk}\} = 0.$$

---

*See Ref. 44, p. 108.

Similarly,

$$E_i\{\theta_j^2\} = 0.$$

Also, because $E_o\{\theta_j^2\} < \infty$ and $E_o\{\theta_{kn}^2\} < \infty$ for all j, k, and n,

$$\left| \int_\Omega (\theta_j\theta_k - \theta_{jm}\theta_{kn})dP_o \right| \leq \int_\Omega \{|\theta_j(\theta_k-\theta_{kn})| + |\theta_{kn}(\theta_j-\theta_{jm})|\}dP_o$$

$$\leq [E_o\{\theta_j^2\}]^{1/2}\left[\int_\Omega |\theta_k-\theta_{kn}|^2 dP_o\right]^{1/2} + [E_o\{\theta_{kn}^2\}]^{1/2}\left[\int_\Omega |\theta_j-\theta_{jm}|^2 dP_o\right]^{1/2}$$

which implies

$$E_o\{\theta_j\theta_k\} = \lim_{m,n\to\infty} E_o\{\theta_{jm}\theta_{kn}\} . \qquad (7.92)$$

Similarly, it can be shown

$$E_i\{\theta_j\theta_k\} = \lim_{m,n\to\infty} E_i\{\theta_{jm}\theta_{kn}\} . \qquad (7.93)$$

Then, applying Eqs. (7.91), (7.92), and (7.93), we find

$$E_o\{\theta_j\theta_k\} = \lim_{m,n\to\infty} ([f_{jm}],[f_{kn}])$$

$$= ([f_j],[f_k])$$

and

$$E_i\{\theta_j\theta_k\} = \lim_{m,n\to\infty} (R_i^{1/2}R_o^{-1/2}[f_{jm}], R_i^{1/2}R_o^{-1/2}[f_{kn}])$$

$$= (X[f_j],X[f_k])$$

where X is the bounded (continuous) extension of $R_i^{1/2}R_o^{-1/2}$.

Q.E.D.

In the next lemma, we shall obtain an expansion for the processes $[Z^o]$ and $[Z^i]$ in terms of a countable set of Gaussian random variables, each of which is measurable with respect to $\mathcal{B}$.

Lemma 7.9: Assume $R_i^{1/2}R_o^{-1/2}$ is bounded and denote its bounded extension to the whole of $\bar{L}_2[T]$ by X. Then, if the identity* is denoted by I and if the self-adjoint operator I-X*X is completely continuous,

$$[Z^o] = \underset{n\to\infty}{\text{l.i.m.}} \left[\sum_{j=1}^{n} [R_o^{1/2}[\phi_j]]\eta_j + \sum_{j=1}^{n} [R_o^{1/2}[\bar{\phi}_j]]\bar{\eta}_j \right] (txP_o) \qquad (7.94)$$

and**

$$[Z^i] = \underset{n\to\infty}{\text{l.i.m.}} \left[\sum_{j=1}^{n} [R_o^{1/2}[\phi_j]]\eta_j + \sum_{j=1}^{n} [R_o^{1/2}[\bar{\phi}_j]]\bar{\eta}_j \right] (txP_i) \qquad (7.95)$$

where $\{[\phi_j]\}$ are the orthonormal eigenfunctions corresponding to nonzero eigenvalues of I-X*X, $\{[\bar{\phi}_j]\}$ is an orthonormal basis for the null space of I-X*X, $\{\eta_j\}$ and $\{\bar{\eta}_j\}$ are square integrable, measurable with respect to $\mathcal{B}$, jointly Gaussian random variables that satisfy

$$\eta_j = \underset{k\to\infty}{\text{l.i.m.}} \int_T [Z^i(t,[f])]^T[R_o^{-1/2}([\phi_{jk}])](t)dt$$

$$\bar{\eta}_j = \underset{k\to\infty}{\text{l.i.m.}} \int_T [Z^i(t,[f])]^T[R_o^{-1/2}([\bar{\phi}_{jk}])](t)dt$$

with $\{[\phi_{jk}]\}_k$ and $\{[\bar{\phi}_{jk}]\}_k$ being elements of $\bar{L}_2[T]$ chosen such that they are elements of the domain of $R_o^{-1/2}$ and such that

$$\lim_{k\to\infty} ||[\phi_j] - [\phi_{jk}]|| = 0$$

$$\lim_{k\to\infty} ||[\bar{\phi}_j] - [\bar{\phi}_{jk}]|| = 0,$$

---

*The identity operator maps every element of its domain into itself.

**If $[Z]$ and $[Z_n]$ are vectors,

$$[Z] = \underset{n\to\infty}{\text{l.i.m.}}[Z_n](txP_o) \iff \lim_{n\to\infty} \int_{\Omega xT} [[Z]-[Z_n]]^T[[Z]-[Z_n]]d(txP_o) = 0.$$

and $(t\times P_o)$ and $(t\times P_i)$ are the product measures on $(\Omega\times T)$.

Proof: Note that X is uniquely defined since $R_i^{1/2}R_o^{-1/2}$ is densely defined and bounded (Theorem 7.16). Also $\{\eta_j\}$ and $\{\overline{\eta}_j\}$ exist as a result of Lemma 7.7 and

$$\begin{aligned}
E_o\{\eta_j\eta_k\} &= ([\phi_j],[\phi_k]) \\
E_o\{\overline{\eta}_j\overline{\eta}_k\} &= ([\overline{\phi}_j],[\overline{\phi}_k]) \\
E_i\{\eta_j\eta_k\} &= ([\phi_j],X^*X[\phi_k]) \\
E_i\{\overline{\eta}_j\overline{\eta}_k\} &= ([\overline{\phi}_j],X^*X[\overline{\phi}_k])
\end{aligned} \tag{7.96}$$

by Lemma 7.8. In addition, $\{[\phi_j]\}$ and $\{[\overline{\phi}_j]\}$ exist since $I-X^*X$ is completely continuous and self-adjoint (Theorem 7.20).

Now consider

$$I_n^o = \int_T \int_\Omega \left[[Z^o] - \sum_{j=1}^n [R_o^{1/2}[\phi_j]]\eta_j - \sum_{j=1}^n [R_o^{1/2}[\overline{\phi}_j]]\overline{\eta}_j\right]^T \left[[Z^o] - \sum_{j=1}^n [R_o^{1/2}[\phi_j]]\eta_j - \sum_{j=1}^n [R_o^{1/2}[\overline{\phi}_j]]\overline{\eta}_j\right] dP_o dt.$$

(In this case, the interated integral $\int_T\int_\Omega$ is equal to $\int_{T\times\Omega}$ due to Theorem 7.10.) Then, performing the Ω integration and substituting (7.96), we find

$$\begin{aligned}
I_n^o = {} & Tr\left\{\int_T [r^o(t,t)]dt\right\} - 2\sum_{j=1}^n ([R_o^{1/2}[\phi_j]],[E_o\{[Z^o]\eta_j\}]) \\
& - 2\sum_{j=1}^n ([R_o^{1/2}[\phi_j]],[E_o\{[Z^o]\overline{\eta}_j\}]) \\
& + \sum_{j=1}^n ([R_o^{1/2}[\phi_j]],[R_o^{1/2}[\phi_j]]) + \sum_{j=1}^n ([R_o^{1/2}[\overline{\phi}_j]],[R_o^{1/2}[\overline{\phi}_j]])
\end{aligned} \tag{7.97}$$

where the symbol Tr{ } has been used to denote the trace of the indicated matrix. Furthermore, since

$$\left|\int_\Omega [Z_t^o]_q(\eta_j-\eta_{jk})dP_o\right| \leq \left\{\int_\Omega [Z_t^o]_q^2 \; dP_o\right\}^{1/2} \left\{\int_\Omega |\eta_j-\eta_{jk}|^2 dP_o\right\}^{1/2},$$

we have

$$\begin{aligned}
[E_o\{[Z_t^o]\eta_j\}] &= \lim_{k\to\infty} \int_\Omega \int_T [Z_t^o][Z_s^o]^T[R_o^{-1/2}[\phi_{jk}]](s)\,ds\,dP_o \\
&= \lim_{k\to\infty} [R_o^{1/2}[\phi_{jk}]] \\
&= [R_o^{1/2}[\phi_j]] \qquad (7.98)
\end{aligned}$$

when the order of Ω and T integration is interchanged. ($[Z_t^o]_q$ denotes the qth element of the vector $[Z_t^o]$.) This interchange can be justified by an argument similar to that used in the previous lemmas. Similarly,

$$[E_o\{[Z_t^o]\overline{\eta}_j\}] = [R_o^{1/2}[\overline{\phi}_j]]. \qquad (7.99)$$

Now, since R_o is completely continuous and has a zero null space, there exists an othonormal basis for $\overline{L}_2[T]$, $\{[\psi_k^o]\}$, which satisfies

$$[R_o[\psi_k^o]] = \lambda_k^o[\psi_k^o]$$

and

$$[r^o(t,s)] = \lim_{n\to\infty} \sum_{k=1}^{n} \lambda_k^o[\psi_k^o(t)]\cdot[\psi_k^o(s)]^T$$

with the sum converging uniformly (Theorem 7.22). As a result, we have

$$\begin{aligned}
\mathrm{Tr}\left\{\int_T [r^o(t,t)]dt\right\} &= \mathrm{Tr}\left\{\lim_{n\to\infty} \sum_{k=1}^{n} \lambda_k^o \int_T [\psi_k^o(t)]\cdot[\psi_k^o(t)]^T\right\} dt \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \lambda_k^o \qquad (7.100)
\end{aligned}$$

To justify the last statement, see the discussion concerning Eq. (II.11) and note that

$$\lambda_k^o = ([\psi_k^o],[R_o[\psi_k^o]]) > 0.$$

This last summation of Eq. (7.100), however, can also be written as

$$\sum_{k=1}^{\infty} ([R_o[\psi_k^o]],[\psi_k^o]) = \sum_{k=1}^{\infty} ([R_o^{1/2}[\psi_k^o]],[R_o^{1/2}[\psi_k^o]])$$

which becomes

$$\sum_{k=1}^{\infty} \sum_{j=1}^{\infty} [([R_o^{1/2}[\psi_k^o]],[\phi_j])^2 + ([R_o^{1/2}[\psi_k^o]],[\overline{\phi}_j])^2] \qquad (7.101)$$

when $[R_o^{1/2}[\psi_k^o]]$ is expanded in terms of the orthonormal basis $\{\{[\phi_j]\}, \{[\overline{\phi}_j]\}\}$. Moreover, since $R_o^{1/2}$ is self-adjoint and the order of summation can* be interchanged, (7.101) becomes

$$\sum_{j=1}^{\infty} \{([R_o^{1/2}[\phi_j]],[R_o^{1/2}[\phi_j]]) + ([R_o^{1/2}[\overline{\phi}_j]],[R_o^{1/2}[\overline{\phi}_j]])\}. \qquad (7.102)$$

Then, substituting (7.98), (7.99), and (7.102) into (7.97), we find

$$I_n^o = \sum_{j=n}^{\infty} ||[R_o^{1/2}[\phi_j]]||^2 + \sum_{j=n}^{\infty} ||[R_o^{1/2}[\overline{\phi}_j]]||^2$$

and

$$\lim_{n\to\infty} I_n^o = 0$$

since (7.100) which equals (7.102) is finite.

In a similar manner, it can be shown that

$$I_n^i = \sum_{k=1}^{\infty} \lambda_k^i - \sum_{j=1}^{n} ([R_o^{1/2}[\phi_j]],[R_o^{1/2}[\phi_j]])([X^*X[\phi_j]],[\phi_j])$$

$$- \sum_{j=1}^{n} ([R_o^{1/2}[\overline{\phi}_j]],[R_o^{1/2}[\overline{\phi}_j]])(X^*X[\overline{\phi}_j],[\overline{\phi}_j])$$

---

*See Ref. 59, p. 161.

where λ_k^i satisfies

$$[R_i[\psi_k^i]] = \lambda_k^i[\psi_k^i]$$

and $\{[\psi_k^i\}$ is an orthonormal basis for $\overline{L}_2[T]$. Also,

$$\sum_{k=1}^{\infty} \lambda_k^i = \sum_{k=1}^{\infty} ([R_i[\psi_k^i]],[\psi_k^i])$$

$$= \sum_{k=1}^{\infty} ([R_o^{1/2}X^*XR_o^{1/2}[\psi_k^i]],[\psi_k^i]). \tag{7.103}$$

Then, expanding $[R_o^{1/2}[\psi_k^i]]$ in terms of the basis $\{\{[\phi_j]\ ,\ \{[\overline{\phi}_k]\}\}$, (7.103) becomes

$$\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} ([R_o^{1/2}[\psi_k^i]],[\phi_j])^2([X^*X[\phi_j]],[\phi_j])$$

$$+ \sum_{k=1}^{\infty}\sum_{k=1}^{\infty} ([R_o^{1/2}[\psi_k^i]],[\overline{\phi}_j])^2([X^*X[\overline{\phi}_j]],[\overline{\phi}_j]).$$

Finally since $R_o^{1/2}$ is self-adjoint, interchanging the order of summation* results in

$$\sum_{k=1}^{\infty} \lambda_k^i = \sum_{j=1}^{\infty} ([R_o^{1/2}[\phi_j]],[R_o^{1/2}[\phi_j]])([X^*X[\phi_j]],[\phi_j])$$

$$+ ([R_o^{1/2}[\overline{\phi}_j]],[R_o^{1/2}[\overline{\phi}_j]])([X^*X[\overline{\phi}_j]],[\overline{\phi}_j]),$$

and we see again that

$$\lim_{n\to\infty} I_n^i = 0 .$$

Q.E.D.

The final lemma of this section proves the σ-field $\mathcal{B}$ is contained in a slight enlargment of the σ-field generated by the random vari-

---

*See Ref. 59, p. 161.

ables $\{\eta_j\}$ and $\{\bar{\eta}_j\}$ defined in Lemma 7.9.

Lemma 7.10: Under the hypothesis of Lemma 7.9,

$$\mathcal{B} \subset \hat{\hat{\mathcal{B}}}$$

where $\hat{\hat{\mathcal{B}}}$ is the σ-field of sets of the form $\Lambda\Delta N$, $\Lambda\varepsilon\hat{\mathcal{B}}$, $N\subset N_o\varepsilon\mathcal{B}$, $P_o(N_o) = 0$, $N\subset N_i\varepsilon\mathcal{B}$, $P_i(N_i) = 0$, and $\hat{\mathcal{B}}$ is the minimum σ-field with respect to which the random variables $\{\eta_j\}$ and $\{\bar{\eta}_j\}$ are measurable.

Proof: Let us first prove $\mathcal{B}'\subset\hat{\hat{\mathcal{B}}}$. (Remember, $\mathcal{B}'$ is the minimum σ-field with respect to which $[Z_t^o]$ and $[Z_t^i]$ are measurable for all $t\varepsilon T$.) To prove this proposition, is suffices to prove $[Z(t,[f])] = [Z^o(t,[f])] = [Z^i(t,[f])]$ is measurable for every $t\varepsilon T$.

Define

$$[S_n(t,[f])] = \sum_{j=1}^{n} \{[R_o^{1/2}[\phi_j]](t)\eta_j([f]) + [R_o^{1/2}[\bar{\phi}_j]](t)\bar{\eta}_j([f])\}.$$

Then, since $\{[S_n(t,[f])]\}$ converges to $[Z(t,[f])]$ in the mean with respect to $(t\mathrm{x}P_o)$ and $(t\mathrm{x}P_i)$ by the previous lemma, $[S_n(t,[f])]$ also converges in the mean to $[Z(t,[f])]$ with respect to the measure $(t\mathrm{x}(P_o+P_i))$. As a result, there exists a subsequence $\{[S_{n_k}(t,[f])]\}$ which converges to $[Z(t,[f])]$ a.e. $(t\mathrm{x}(P_o+P_i))$ and, hence, a.e. $(t\mathrm{x}P_o)$ and a.e. $(t\mathrm{x}P_i)$. (See Theorem 7.1.) Thus, if

$$\breve{N} = \{(t,[f]);[Z(t,[f])] \neq \overline{\lim_{n_k\to\infty}} [S_{n_k}(t,[f])]\},$$

we have

$$(t\mathrm{x}P_o)(\breve{N}) = (t\mathrm{x}P_i)(\breve{N}) = 0.$$

Then, due to Theorem 7.10, there exists a set $T'\subset T$ such that, for $t\varepsilon T'$,

$$P_o(\tilde{N}_t) = P_i(\tilde{N}_t) = 0 \qquad (7.104)$$

when $\tilde{N}_t$ is the section of $\tilde{N}$ determined by t and $t(T') = T_f$, i.e., the measure of the set T' equals the measure of the interval $[0,T_f]$. Now define

$$B_t = \{[f];\ Z(t,[f])\varepsilon A\}$$

and

$$\hat{B}_t = \{[f];\ \overline{\lim_{n_k\to\infty}}\ S_{n_k}(t,[f])\varepsilon A\}$$

for $t\varepsilon T'$ and A an arbitrary Borel set in K-dimensional Euclidean space. (Note* that $B_t \varepsilon \mathcal{B}'$ and $\hat{B}_t \varepsilon \hat{\mathcal{B}}$). Then, since

$$B_t = (B_t \cap \tilde{N}_t) \cup (B_t \cap \tilde{N}_t^C)$$

$$B_t = (\hat{B}_t \cap \tilde{N}_t) \cup (\hat{B}_t \cap \tilde{N}_t^C)$$

and

$$B_t \cap \tilde{N}_t^C = \hat{B}_t \cap \tilde{N}_T^C$$

if we denote the complement of N_t by N_t^C, and since

$$P_o(B_t \cap \hat{N}) = P_o(\hat{B}_t \cap \tilde{N}_t) = P_i(B_t \cap \tilde{N}_t) = P_i(\hat{B}_t \cap \tilde{N}_t) = 0$$

due to (7.104), when we form $B_t \Delta \hat{B}_t$, we find

$$P_o(B_t \Delta \hat{B}_t) = P_i(B_t \Delta \hat{B}_t) = 0.$$

It can now be seen, since

$$B_t = \hat{B}_t \Delta (\hat{B}_t \Delta B_t),$$

$B_t \varepsilon \hat{\hat{\mathcal{B}}}$ for $t\varepsilon T'$, i.e., $[Z(t,[f])]$ is $\hat{\hat{\mathcal{B}}}$-measurable for almost every $t\varepsilon T$.

*See Ref. 30, p. 84.

Let us now prove $[Z(t,[f])]$ is $\hat{\hat{\mathcal{B}}}$-measurable for every $t\varepsilon T$. Since $[r^o(t,s)]$ and $[r^i(t,s)]$ are continuous by Assumption A6.3.1 of Section 6.3 and T' is dense in T, there exists, for every $t\varepsilon T$, a sequence $\{t_n\}$, $t_n \varepsilon T'$, converging to t such that

$$\lim_{n\to\infty} E_o\{||[Z(t,[f])] - [Z(t_n,[f])]||^2\} = 0$$

and

$$\lim_{n\to\infty} E_i\{||[Z(t,[f])] - [Z(t_n,[f])]||^2\} = 0.$$

Hence, there exists a subsequence $\{t_{n_k}\}$ such that $[Z(t_{n_k},[f])]$ converges to $[Z(t,[f])]$ a.e. (P_o) and a.e. (P_i). Then, since each $[Z(t_{n_k},[f])]$ is $\hat{\hat{\mathcal{B}}}$-measurable for every t_{n_k}, the same argument as used above shows $[Z(t,[f])]$ is $\hat{\hat{\mathcal{B}}}$-measurable for every $t\varepsilon T$.

If we now return to Eq. (7.75), we see

$$\mathcal{B} \subset \hat{\mathcal{B}}'$$

and, due to the definition of $\hat{\mathcal{B}}'$ and the above discussion,

$$\hat{\mathcal{B}}' \subset \hat{\hat{\mathcal{B}}}$$

which together imply

$$\mathcal{B} \subset \hat{\hat{\mathcal{B}}} \ .$$

Q.E.D.

We now have all the lemmas that are necessary to prove the principal theorem of this chapter. This theorem specifies the conditions under which $P_o \equiv P_i$, $i\varepsilon M_o'$, and the form of the Radon-Nikodym derivatives, $d\overline{P}_i/d\overline{P}_o$, $i\varepsilon M_o'$.

Theorem 7.26: The following three properties relating P_o and P_i are true:

(1) Either $P_o \equiv P_i$ or $P_o \perp P_i$.

(2) $P_o \equiv P_i$ if and only if $R_i^{1/2}R_o^{-1/2}$ is bounded and I-X*X is a Hilbert-Schmidt operator where X is the bounded extension of $R_i^{1/2}R_o^{-1/2}$ to all of $\overline{L}_2[T]$.

(3) If $P_o \equiv P_i$,

$$\frac{d\overline{P}_i}{d\overline{P}_o} = \lim_{n\to\infty} \exp\left\{\frac{1}{2}\sum_{j=1}^{n}\left[\left(\frac{\rho_j}{\rho_{j-1}}\right)\eta_j^2 - \log(1-\rho_j)\right]\right\} \quad \text{a.e.}(\overline{P}_o)$$

where ρ_j, $j\varepsilon\{1,2,\ldots\}$, are the eigenvalues of I-X*X and η_j, $j\varepsilon\{1,2,\ldots\}$, are defined as in Lemma 7.9 with $[\phi_j]$ being the eigenfunctions of I-X*X.

Proof: Let us first prove (2).

Necessity: Assume $P_o \equiv P_i$. Then, from Lemma 7.6 and Theorem 7.16, $R_i^{1/2}R_o^{-1/2}$ is bounded and possesses a unique bounded extension to all of $\overline{L}_2[T]$ which we are denoting by X. Furthermore, it is easily seen that X*X is bounded, self-adjoint and positive. Theorem 7.21 then guarantees the existence of a spectral decomposition of X*X which we can write as $\int_0^\infty \nu dE_\nu$ where E_ν represents a resolution of the identity. We shall now show, by contradiction, X*X has a purely discrete spectrum.

Suppose for some $\varepsilon > 0$, $I-E_{1+\varepsilon}$ is infinite dimensional. Then, there exists a sequence of real numbers $\{\nu_j\}$ such that $1+\varepsilon \le \nu_1 \le \nu_2 \le \ldots$ and a sequence of orthonormal functions $\{[f_j]\}$, from $\overline{L}_2[T]$, such that

$$(E_\nu'' - E_\nu')[f_j] = [f_j]$$

where $\nu' \le \nu_j < \nu''$ and

$$(E_{\nu''}-E_{\nu'})[f_j] = 0$$

when $\nu_j < \nu' < \nu''$ or $\nu' < \nu'' \leq \nu_j$. It then follows

$$([f_j],X^*X[f_k]) = \int_0^\infty \nu d(E_\nu[f_j],[f_k])$$

$$\geq \delta_{jk}\nu_j$$

$$\geq \delta_{jk}(1+\varepsilon) . \qquad (7.105)$$

But, due to Lemma 7.8, there also exists a sequence of random variables $\{\theta_j\}$ which are measurable with respect to $\mathcal{B}$, jointly Gaussian with respect to both P_0 and P_i and such that

$$E_0\{\theta_j\} = E_i\{\theta_j\} = 0$$

$$E_0\{\theta_j\theta_k\} = \delta_{jk}$$

$$E_i\{\theta_j\theta_k\} = ([f_j],X^*X[f_k]). \qquad (7.106)$$

Now let $\mathcal{B}''$ be the minimum σ-field with respect to which the sequence $\{\theta_j\}$ is measurable, and let P_0'' and P_i'' be the restrictions of P_0 and P_i to $\mathcal{B}''$. Then, from Lemma 7.4 and Eqs. (7.105) and (7.106), $P_0'' \perp P_i''$ since

$$\sum_{j=1}^\infty \left(1 - \frac{E_0\{\theta_j^2\}}{E_i\{\theta_j^2\}}\right)^2 = \infty . \qquad (7.107)$$

But, by Lemma 7.2, we also have $P_0 \perp P_i$ which is a contradiction. Therefore, $I-E_{1+\varepsilon}$ is finite dimensional for every $\varepsilon > 0$. Similarly, it can be shown that $E_{1-\varepsilon}$ is finite dimensional. Hence, X*X has a purely discrete spectrum with 1 being the only possible limit point of the spectrum. This implies I-X*X also has a discrete spectrum and its only possible limit point is zero. Therefore, by Corollary 7.1, I-X*X is completely continuous.

If we now look at the equation

$$(I-X^*X)[\phi] = \rho[\phi],$$

we see that the eigenfunctions of I-X*X are also eigenfunctions of X*X with the eigenvalues of X*X being one minus the corresponding eigenvalue of I-X*X. As a result, the $\{\eta_j\}$ and $\{\overline{\eta}_j\}$ defined in Lemma 7.9 can be shown to have the following properties:

$$E_o\{\eta_j\} = E_o\{\overline{\eta}_j\} = E_i\{\eta_j\} = E_i\{\overline{\eta}_j\} = 0,$$

$$E_o\{\eta_j\eta_k\} = ([\phi_j],[\phi_k]) = \delta_{jk},$$

$$E_o\{\overline{\eta}_j\overline{\eta}_k\} = ([\overline{\phi}_j],[\overline{\phi}_k]) = \delta_{jk} ,$$

$$E_o\{\eta_j\overline{\eta}_k\} = ([\phi_j],[\overline{\phi}_k]) = 0,$$

$$E_i\{\eta_j\overline{\eta}_k\} = ([\phi_j],X^*X[\overline{\phi}_k]) = 0,$$

$$E_i\{\eta_j\eta_k\} = ([\phi_j],X^*X[\phi_k] = (1-\rho_j)\delta_{jk},$$

$$E_i\{\overline{\eta}_j\overline{\eta}_k\} = ([\overline{\phi}_j],X^*X[\overline{\phi}_k]) = \delta_{jk}. \tag{7.108}$$

To verify these properties, an argument analogous to that used to prove Eq. (7.90) of Lemma 7.8 suffices.

Now let $\mathcal{B}$ be the minimum σ-field with respect to which $\{\eta_j\}$ and $\{\overline{\eta}_j\}$ are measurable and let $\hat{P}_o$ and $\hat{P}_i$ be the restrictions of P_o and P_i to $\mathcal{B}$. Then, it follows from Lemma 7.4 that either $\hat{P}_o \equiv \hat{P}_i$ or $\hat{P}_o \perp \hat{P}_i$, and $\hat{P}_o \equiv \hat{P}_i$ if and only if

$$\sum_{j=1}^{\infty} \left(1 - \frac{1}{1-\rho_j}\right)^2 < \infty . \tag{7.109}$$

Furthermore, from Lemma 7.2, $\hat{P}_o \perp \hat{P}_i \Rightarrow P_o \perp P_i$. But since $P_o \equiv P_i$ from the hypothesis, we must have $\hat{P}_o \equiv \hat{P}_i$. Hence, (7.109) is satisfied, or equivalently,

$$\sum_{j=1} \rho_j^2 < \infty \tag{7.110}$$

which implies, by Definition 7.24, that I-X*X is a Hilbert-Schmidt operator.

Sufficiency: Assume I-X*X is a Hilbert-Schmidt operator on $\overline{L}_2[T]$. We can then establish $\{\eta_j\}$, $\{\overline{\eta}_j\}$ and (7.108) as previously done. Moreover, since I-X*X is Hilbert-Schmidt, (7.110) and hence (7.109) is satisfied. It then follows from Lemma 7.4 that $\hat{P}_o \equiv \hat{P}_i$ which, together with Lemmas 7.3 and 7.10, imply

$$P_o \equiv P_i \ .$$

Lew us now prove (1). Assume P_o and P_i are not equivalent. Then, according to (2) proved above, one of the following three cases must hold:

(a) $R_i^{1/2}R_o^{-1/2}$ is unbounded,

(b) I-X*X is bounded but not completely continuous,

(c) I-X*X is completely continuous but not Hilbert-Schmidt.

In case (a), $P_o \perp P_i$ by Lemma 7.6. In case (b), X*X has a spectral representation and either $I-E_{1+\varepsilon}$ or $E_{1+\varepsilon}$ must be infinite dimensional for some $\varepsilon > 0$. Then, $P_o \perp P_i$, as shown in the paragraph which contains Eq. (7.107). In case (c), I-X*X has the eigenvalues and eigenfunctions $\{\rho_j\}$ and $\{[\phi_j]\}$. Also, the random variables $\{\eta_j\}$ and $\{\overline{\eta}_j\}$, as described previously, are well defined. But, since I-X*X is not Hilbert-Schmidt, (7.110), and, hence, (7.109) do not hold. Then, according to Lemma 7.4, $\hat{P}_o \perp \hat{P}_i$ and, from Lemma 7.2, $P_o \perp P_i$. Therefore, we conclude that if P_o and P_i are not equivalent, then, they must be singular. This trivially implies that if P_o and P_i are not singular, then, they must be equivalent.

Let us now prove (3). This assertion, however, is an immediate consequence of Lemma 7.4 (3) and Lemma 7.3 (3).

Q.E.D.

The next two theorems provide, under certain conditions, an alternate functional form for the Radon-Nikodym derivatives, $d\overline{P}_i/d\overline{P}_o$, $i\varepsilon M_o$. This alternate form will allow us to determine, by the use of linear filters and correlators, in which set of the optimal partition of the observation space a particular sample function belongs.

<u>Theorem 7.27</u>: The measures P_o and P_i are equivalent if an only if there exists a Hilbert-Schmidt operator U defined on $\overline{L}_2[T]$ which satisfies

$$R_o^{1/2} U R_o^{1/2} = R_o - R_i \ . \tag{7.111}$$

<u>Proof</u>: If P_o and P_i are equivalent, let

$$U' = I - X^*X$$

where X is the bounded extension of $R_i^{1/2}R_o^{-1/2}$. ($R_i^{1/2}R_o^{-1/2}$ is bounded due to Lemma 7.6.). Then, since U' can be seen to satisfy (7.111) and, by Theorem 7.26, U' is Hilbert-Schmidt, we see that the conditions of the theorem are necessary.

To prove these conditions are sufficient, assume U is a Hilbert-Schmidt operator that satisfies (7.111). Then U also satisfies

$$R_i = R_o^{1/2} (I-U) R_o^{1/2}$$

and, as a result, on the dense subset of $\overline{L}_2[T]$ where $R_o^{-1/2}R_iR_o^{-1/2}$ is defined, $R_o^{-1/2}R_iR_o^{-1/2}$ is equal to I-U. Furthermore, since U is completely continuous by Theorem 7.24, U is bounded as is I-U so that $R_o^{-1/2}R_iR_o^{-1/2}$ is also bounded. Finally, since the bounded extension of a densely defined bounded operator is unique by Theorem 7.16, we see that

$$X^*X = I-U.$$

Thus, $I-X^*X$ is Hilbert-Schmidt and the application of Theorem 7.26 completes the proof.

Q.E.D.

Theorem 7.28: If there exists a bounded self-adjoint operator H_i on $\overline{L}_2[T]$ satisfying

$$R_oH_iR_i = R_i-R_o, \tag{7.112}$$

then P_o and P_i are equivalent. Moreover,

$$\lim_{n\to\infty} \exp\left[\frac{1}{2}\sum_{j=1}^{n} \log(1-\rho_j)\right] = D_i$$

exists and is finite and

$$\frac{dP_i}{d\overline{P}_o} = D_i^{-1} \exp\{ \frac{1}{2} (H_i([f]),[f])\} \quad \text{a.e.}(\overline{P}_o)$$

where $\{\rho_j\}$ are the eigenvalues of $I-X^*X$ and $[f]$ is any observation from the observation space Ω. (If $[f]\notin\overline{L}_2[T]$, $(H_i([f]),[f])$ is defined to be zero.)

Proof: If (7.112) is satisfied, we can write

$$R_i = R_o(I+H_iR_i) = (I+R_iH_i)R_o$$

which shows that $R_o^{-1/2}R_iR_o^{-1/2}$ is bounded and densely defined on the subset of $\overline{L}_2[T]$ where $R_o^{-1/2}$ is defined. As a result, we can define an operator U to be the uniquely defined bounded extension of

$$I - R_o^{-1/2}R_iR_o^{-1/2} = - R_o^{1/2}H_iR_iR_o^{-1/2} = - R_o^{-1/2}R_iH_iR_o^{-1/2} .$$

Furthermore, since R_o is positive definite and completely continuous, we can assume the sequence of eigenfunctions $\{[\psi_j^o]\}$ of R_o form an orthonormal basis for $\overline{L}_2[T]$. Then, if we let $\{\lambda_j^o\}$ be the sequence of eigenvalues that correspond to $\{[\psi_j^o]\}$, we find

$$[N(U)]^2 = \sum_{j=1}^{\infty} ||U([\psi_j^o])||^2 = \sum_{j=1}^{\infty} (R_o^{1/2}H_iR_iR_o^{-1/2}([\psi_j^o]), R_o^{-1/2}R_iH_iR_oR_o^{1/2}([\psi_j^o]))$$

$$= \sum_{j=1}^{\infty} (\lambda_j^o)^{-1/2}(\lambda_j^o)^{1/2}(R_o^{1/2}H_iR_i([\psi_j^o]), R_o^{-1/2}R_iH_i([\psi_j^o]))$$

$$= \sum_{j=1}^{\infty} (H_iR_i([\psi_j^o]), R_iH_i([\psi_j^o]))$$

$$= tr(R_iH_iR_iH_i)$$

$$< N(R_i)N(H_iR_iH_i) \leq ||H_i||^2[H(R_i)]^2 < \infty$$

when the results* of Theorem 7.23 are applied. Furthermore U satisfies (7.111) and, from Theorem 7.27, we see that P_o and P_i are equivalent.

Now let X be the unique bounded extension of $R_i^{1/2}R_o^{-1/2}$. (Again, $R_i^{1/2}R_o^{-1/2}$ is bounded due to Lemma 7.6 and the fact that $P_o \equiv P_i$.) It can be seen that $U = I-X^*X$ since U is the unique bounded extension of $I-R_o^{-1/2}R_iR_o^{-1/2}$. Furthermore, if $A \subset \overline{L}_2[T]$ denotes the dense subset of $\overline{L}_2[T]$ on which $R_o^{-1/2}$ is defined, we have

$$\begin{aligned}(X^*X)(R_o^{1/2}H_iR_o^{1/2}) &= (R_o^{-1/2}R_iR_o^{-1/2})(R_o^{1/2}H_iR_o^{1/2}) \\ &= R_o^{-1/2}R_iH_iR_o^{1/2} \\ &= R_o^{-1/2}R_iH_iR_oR_o^{-1/2} \quad \text{on } A \\ &= R_o^{-1/2}R_iR_o^{-1/2} - I \quad \text{on } A \\ &= X^*X - I \qquad (7.113)\end{aligned}$$

*The symbol N(U) denotes the Hilbert-Schmidt norm of the operator U.

since $(X^*X)\ (R_o^{1/2}H_iR_o^{1/2})$ is bounded and the bounded extension of $I-R_o^{-1/2}R_iR_o^{-1/2}$ is unique. Similarly, we find

$$(R_o^{1/2}H_iR_o^{1/2})(X^*X) \quad = \quad X^*X - I$$

which, together with (7.113), implies

$$I - R_o^{1/2}H_iR_o^{1/2} \quad = \quad (X^*X)^{-1}$$

or

$$R_o^{1/2}H_iR_o^{1/2} \quad = \quad I - (X^*X)^{-1}\ . \tag{7.114}$$

Then, if $\{[\phi_j]\}$ and $\{\rho_j\}$ are the eigenfunctions and eigenvalues, respectively, of the Hilbert-Schmidt operator I-X*X and $\{\bar{\phi}_j\}$ is a basis for its null space, we have

$$(X^*X)^{-1}[\phi_j] \quad = \quad \left(\frac{1}{1-\rho_j}\right)[\phi_j]$$

and

$$(X^*X)^{-1}[\bar{\phi}_j] \quad = \quad [\bar{\phi}_j]\ .$$

Finally, this implies, because of (7.114),

$$(H_iR_o^{1/2}[\phi_j],R_o^{1/2}[\phi_k]) \quad = \quad \left(\frac{\rho_j}{\rho_{j-1}}\right)\ \delta_{jk}\ , \tag{7.115}$$

$$(H_iR_o^{1/2}[\bar{\phi}_j],R_o^{1/2}[\phi_k]) = \quad 0, \tag{7.116}$$

and

$$(H_iR_o^{1/2}[\bar{\phi}_j],R_o^{1/2}[\bar{\phi}_k]) \quad = \quad 0. \tag{7.117}$$

Returning now to Lemma 7.9, we see that

$$[Z^o] \quad = \underset{n\to\infty}{\text{l.i.m.}}[Z_n^o] \qquad (t\times P_o) \tag{7.118}$$

where

$$[Z_n^o] = \sum_{j=1}^{n} \{[R_o^{1/2}([\phi_j])]n_j + R_o^{1/2}([\overline{\phi}_j])]\overline{n}_j\} . \qquad (7.119)$$

Then, if we define $[\overline{Z}_n^o]$ by $[\overline{Z}_n^o] = [Z^o]-[Z_n^o]$ and let H_i be zero off $\overline{L}_2[T]$, we find

$$\lim_{n\to\infty} \int_{Tx\Omega} [H_i([\overline{Z}_n^o])]^T[H_i([\overline{Z}_n^o])]d(txP_o) = \lim_{n\to\infty} \int_\Omega \int_T [H_i([\overline{Z}_n^o])][H_i([\overline{Z}_n^o])]dt\,dP_o$$

$$\leq \lim_{n\to\infty} ||H_i||^2 \int_\Omega \int_T [\overline{Z}_n^o]^T[\overline{Z}_n^o]dt\,dP_o$$

$$= \lim_{n\to\infty} ||H_i||^2 \int_{Tx\Omega} [\overline{Z}_n^o]^T[\overline{Z}_n^o]d(txP_o),$$

$$= 0$$

by the application of Theorem 7.10, the definition of the norm of H_i, and Eq. (7.118). But this implies

$$H_i([Z^o]) = \underset{n\to\infty}{\text{l.i.m.}}\ H_i([Z_n^o]) \qquad (txP_o) \qquad (7.120)$$

Furthermore, as shown in Appendix IV, we can write*

*Note that $m(P_o)$ denotes convergence in measure.

$$\int_T [Z^o(t,[f])]^T [H_i([Z^o])](t)\,dt = \lim_{n\to\infty} \int_T [Z_n^o(t,[f])]^T [H_i([Z_n^o])](t)\,dt \quad m(\overline{P}_o)$$

$$= \lim_{n\to\infty} \sum_{j=1}^{n} \frac{\rho_j}{\rho_{j-1}} \eta_j^2 \qquad m(\overline{P}_o) \quad (7.121)$$

when we make use of Eqs. (7.115)-(7.120).

Let us now obtain the relation that exists between $\int_T [Z^o(t,[f])]^T [H_i([Z^o])](t)\,dt$ and $d\overline{P}_i/d\overline{P}_o$. In Theorem 7.26, it was shown that

$$\lim_{n\to\infty} \{\ell_i^n - \log D_i^n\} = \log \frac{d\overline{P}_i}{d\overline{P}_o} \qquad \text{a.e.}(\overline{P}_o)$$

where

$$\ell_i^n = \frac{1}{2} \sum_{j=1}^{n} \left(\frac{\rho_j}{\rho_{j-1}}\right) \eta_j^2$$

and

$$D_i^n = \prod_{j=1}^{n} (1-\rho_j)^{1/2} .$$

Let us now assume the sequence $\{\log D_i^n\}$ does not converge to a finite number. Then, since $\{\ell_i^n - \log D_i^n\}$ converges a.e.$(\overline{P}_o)$ to the finite valued function $d\overline{P}_i/d\overline{P}_o$, $\{\ell_i^n\}_n$ must converge only on a set of measure zero. Thus, for every n_o and $\varepsilon > 0$,

$$\overline{P}_o\{[f]; |\ell_i^n - \ell_j^m| > \varepsilon\} = 1 \qquad (7.122)$$

for some $n, m \geq n_o$. But, from (7.121) we see, for every $\varepsilon > 0$, there exists an n_o such that if $n, m \geq n_o$, we have

$$\overline{P}_o\{[f]; |\ell_i^n - \ell_i^m| > \varepsilon\} < \varepsilon$$

which is a contradiction when we look at (7.122). As a result, $\{\log D_i^n\}$

must converge to a finite limit, call it $\log D_i$. Furthermore, since $\{\ell_i^n\}_n$ converges in measure to both $\frac{1}{2}\int_T [Z^o(t,[f])]^T [H_i([Z^o])](t)dt$ and $\log \frac{d\overline{P}_i}{d\overline{P}_o}$ + $\log D_i$, we see from Theorem 7.2 that

$$\frac{d\overline{P}_i}{d\overline{P}_o} = D_i^{-1} \exp\left\{\frac{1}{2}\int_T [Z^o(t,[f])]^T [H_i([Z^o])](t)dt\right\} \qquad \text{a.e.}(\overline{P}_o).$$

But, as can be seen from the proof of Theorem 7.9, we can write

$$[Z^o(t,[f])] = [Z'(t,[f])] = [f(t)]$$

for all $[f]\varepsilon\Omega$ which are continuous on $[0,T_f]$. Moreover, since the noise process $[N(t)]$ is separable and there exist two constants $C > 0$ and $\delta > 0$ such that

$$\mathrm{Tr}\{[r^o(t,t)] + [r^o(t',t')] - [r^o(t,t')] - [r^o(t',t)]\} < C|t-t'|^\delta$$

for all $t,t'\varepsilon T$ as stated in Assumption A6.3.1, there also exists* a set $N\subset\Omega$ for which $\overline{P}_o(N) = 0$ and such that all $[f]\varepsilon\Omega-N$ are continuous. Finally, this implies

$$\frac{d\overline{P}_i}{d\overline{P}_o} = D_i^{-1}\exp\{\tfrac{1}{2}(H_i([f]),[f])\} \qquad \text{a.e.}(\overline{P}_o)$$

when $(H^i[f],[f])$ is defined to be zero if $[f]\notin\overline{L}_2[T]$.

Q.E.D.

7.4 SUMMARY AND DISCUSSION OF RESULTS

In this chapter, we have obtained a decision scheme for providing an estimate of the direction of arrival of a stochastic signal which is received by an array of receiving elements in the presence of noise. This scheme makes use of the a priori probabilities of the possible angles of

*See Ref. 50, p. 98.

arrival and requires knowledge of the covariance functions of the signal and noise processes. Moreover, it is an optimal decision scheme in the sense that it results in a minimum probability of error as discussed in Assumption A6.3.1 of Section 6.3. (This scheme is also optimal with respect to the "Maximum A Posteriori Probability" optimality criterion mentioned in Assumption A6.3.1 so that it results in the specification of the direction of arrival which is "most probable" based on the received signals and the available a priori information. Optimally with respect to the "MAP" optimality criterion is discussed in Appendix V.)

The decision scheme is as follows:

(1) Determine, for each possible direction of arrival α_i, a bounded self-adjoint operator H_i which satisfies the operator equation

$$R_o H_i R_i = R_i - R_o \qquad (7.123)$$

where R_o and R_i are the integral operators defined in Eq. (7.75). Determine also, for each i, $D_i = \lim_{n\to\infty} \prod_{j=1}^{n} (1-\rho_j)^{1/2}$ where $\{\rho_j\}$ is the sequence of eigenvalues of the bounded extension of the operator $I-R_o^{-1/2}R_i R_o^{-1/2}$.

(2) Evaluate, for each i, the expression

$$I_i = \{Pr\{\alpha_i\}\} \left[D_i^{-1} \exp \left\{ \frac{1}{2} (H_i([y]),[y]) \right\} \right] \qquad (7.124)$$

where [y] denotes the vector of signals received by the array on the interval $[0,T_f]$ and $Pr\{\alpha_i\}$ is the a priori probability of α_i being the true angle of arrival. In addition, let $I_o = Pr\{\alpha_o\}$ where $Pr\{\alpha_o\}$ is the a priori probability of receiving noise alone.

(3) Finally, make the decision α_i is the true direction of arrival if i is the smallest index of I_i for which $I_i \geq I_j$ for all j. (If $I_o \geq I_j$ for all j, make the decision noise alone is present.)

This decision scheme possesses certain anomalies when the "no signal present" hypothesis has a nonzero a priori probability, however. For instance, when the number of possible angles of arrival become large, each of the a priori probabilities $Pr\{\alpha_i\}$ approach zero. This implies the decision "no signal present" will be made for "nearly all" received samples [y] if $Pr\{\alpha_o\} \neq 0$. This difficulty has arisen because we are comparing the hypothesis "no signal present" with the individual hypothesis "signal at an angle α_i" whose a priori probability is going to zero. An alternate, but more reasonable, decision procedure first makes a choice between the hypothesis "no signal present" and the composite hypothesis "signal at one of the angles α_i." Then, if the latter is chosen as the true hypothesis, it makes a decision as to which from the sequence of hypotheses "signal at an angle α_i" is the true hypothesis. If this alternate decision procedure is adopted and if each of these decision are made so as to minimize the probability of error, it follows easily that step (3) of the above decision scheme should be replaced by the following:

(3') Make the decision "no signal present" if

$$Pr\{\alpha_o\} \geq \sum_{i \in M'} \{Pr\{\alpha_i\}\}\{D_i^{-1} \exp\{\tfrac{1}{2}(H_i([y]),[y])\}\} .$$

Otherwise, make the decision α_i is the true direction of arrival if i is the smallest index of I_i for which $I_i \geq I_j$ for all j.

It should be added that, if no such solution H_i exists for the Eq. (7.123), it is not clear what procedure is optimal for obtaining an estimate of the direction of arrival of the signal. However, in most practical problems it appears that (7.123) has the required solution.

Assuming the sequence of operators $\{H_i\}$ and constants $\{D_i\}$ can be found, attempts to apply the above procedure in a practical direction finding problem will still meet several difficulties, however. In the first place, the received signals normally can have an infinite number of possible directions of arrival. Even if we assume this number is finite, say n_o, it is still necessary to have either n_o processing systems operating in parallel to evaluate (7.124) or the signals must be recorded and then processed at a later time. In either case, the optimal procedure appears to be an impractical approach to the problem. This procedure does, however, suggest suboptimal techniques. For instance, the interval of observation $[0,T_f]$ can be divided into n_o' disjoint subintervals so that Eq. (7.124), for various values of i, can be sequentially evaluated over these subintervals. Although this does not use all available information, it still appears to be a plausible technique when it is impractical to record the signals or have a great redundancy of equipment. It must be realized, though, in most practical problems $n_o > n_o'$ so this technique necessitates the need for a policy which selects the particular I_i, $i \varepsilon \{1,2,\ldots,n_o\}$, to be calculated during each of the n_o' subintervals of $[0,T_f]$. A search for policies that are optimal would probably follow very closely the discussion presented in the first five chapters of this thesis.

In at least two special cases of tremendous practical interest, however, the above difficulties do not exist. These cases, which restrict the class of possible noise and signal covariance functions to those which are stationary and "band-limited," are the subject of the next two chapters.

CHAPTER 8

SPECIAL CASES

8.1 INTRODUCTION

In the previous chapter, we derived an optimal decision scheme for choosing between the hypothesis of "no signal present" and the sequence of hypotheses {"signal at an angle α_i"}. This decision scheme required us first to obtain the operators $\{H_i\}$ and then to calculate, for each i, the value of

$$\ell_i = (H_i([y]),[y]) \qquad (8.1)$$

where [y] represents the observation. Unfortunately, even if we assume the sequence of operators $\{H_i\}$ is known, this necessitates the need for either recording the observation [y] so that ℓ_i, for all α_i, can be calculated at a later time, or a bank of filters, one for each possible α_i, so that the entire sequence $\{\ell_i\}$ can be calculated simultaneously. In this chapter, we shall consider the special cases discussed in Section 6.4 and Sections 6.4.1-6.4.2. The solutions that will be obtained, while being more restrictive in their application, are, nevertheless, of great practical importance since the restricting assumptions typify, in many ways, the environment of realistic direction finding problems and since a greatly simplified implementation is possible. We shall find, in these special cases, a single matrix of $2K^2$ numbers provide all the information given by the sequence $\{\ell_i\}$. (Remember K denotes the number of elements in the antenna array.) Each of these $2K^2$ numbers can be obtained as the output of a realizable filter operating on the received data. We will be able, for these special cases, to eliminate not only the need for recording the received signals or having available a large bank of parallel processing systems but also the need for solving individually for each of the operators in the sequence $\{H_i\}$.

In addition to the sequence $\{\ell_i\}$, the sequence of constants $\{D_i\}$ must be available before an optimal estimate can be made. The problems associated with the calculation of these constants will also be considered in this chapter.

8.2 INDEPENDENT, IDENTICALLY DISTRIBUTED NOISE

Let us consider, first, the problem in which Assumptions A6.4.1 and A6.4.1.1 together with A6.3.1-A6.3.3 are valid descriptions of the environment of the direction finding system. Before obtaining a solution for this problem, however, let us investigate the consistence of these assumptions. In particular, let us shown the existence of constants $C > 0$ and $\delta > 0$ such that

$$\{\sum_{k\varepsilon K} [r^o(t,t)]_{kk} + [r^o(v,v)]_{kk} - [r^o(t,v)]_{kk} - [r^o(v,t)]_{kk}\} < C|t-v|^\delta \tag{8.2}$$

when $t, v\varepsilon T$ is consistent with the "narrow-band" assumption. Since the noise process driving each array element is now stationary, independent of all others, and possesses a covariance function $r_N(\tau)$, if we let $\bar{K} = 1/2K$, Eq. (8.2) reduces to

$$|r_N(0) - r_N(\tau)| < \bar{K}C|\tau|^\delta \quad . \tag{8.3}$$

Furthermore, since $r_N(\tau)$ is continuous and bounded, a question as to the existence of the constants $C > 0$ and $\delta > 0$ such that Eq. (8.3) is satisfied arises only when $\tau \to 0$. Moreover, if Eq. (8.3) is satisfied for $\delta > 1$

and any $\tau < 1$, it is automatically satisfied for $\delta = 1$ and any $\tau < 1$. Thus, we need only investigate the implications of Eq. (8.3) for $C > 0$ and $0 < \delta \leq 1$. Suppose, now,

$$\lim_{\omega\to\infty} \omega^2 \; F_N(\omega) \; < \bar{K}C \qquad (8.4)$$

where $F_N(\omega)$ is the Fourier transform of $r_N(\tau)$. Then, applying the Initial Value Theorem* from the Theory of Fourier Transforms, we find that Eq. (8.4) implies

$$\lim_{\tau\to 0} \left| \frac{r_N(0)-\tau_N(\tau)}{\tau} \right| < \bar{K}C$$

which implies

$$|r_N(0) - r_N(\tau)| < \bar{K}C|\tau|^{\delta}$$

for $0 < \delta \leq 1$ and τ sufficiently close to zero. Thus, we see the conditions under which there exist constants $C > 0$ and $\delta > 0$ such that Eq. (8.3) is satisfied are related to the manner in which the power spectrum falls off for large ω and, therefore, are consistent with the narrow-band assumption.

Turning now to the problem of specifying an optimal estimation procedure when the above stated assumptions are valid, we see, according to the discussion of the previous chapter, an optimal estimate requires the use of a bounded self-adjoint operator H_i which satisfies

$$R_o H_i R_i \;\; = \;\; R_i \; - \; R_o.$$

But, referring to Assumption A6.4.1, we find there exists a symmetric kernel $[h_i^T(t,s)]$ which satisfies Eq. (6.6) and which is square integrable

*See Ref. 60, p. 267.

on TxT. Then, since $[h_i^T(t,s)]$ is symmetric and square integrable, the integral operator H_i^T with $[h_i^T(t,s)]$ as its kernel is self-adjoint and Hilbert-Schmidt which, in turn, implies H_i^T is self-adjoint and bounded. The proof establishing these properties follows exactly the proof presented in Appendix II where R_i was shown to be self-adjoint and Hilbert-Schmidt. Moreover, since $[h_i^T(t,s)]$ satisfies Eq. (6.6), it is easily shown that H_i^T satisfies

$$R_o H_i^T R_o = R_i - R_o$$

so H_i^T represents a realization of the desired operator H_i.

At this point, it is instructive to note that there actually are covariance functions $[r^i(t,s)]$ and $[r^o(t,s)]$ for which we can find symmetric square integrable kernels satisfying Eq. (6.6). For example, suppose the orthonormal sequences of eigenfunctions for the operators R_o and R_i are identical and given by $\{[\psi_i^o]\}$ while the corresponding eigenvalues are $\{\lambda_i^o\}$ and $\{\lambda_j^i\}$, respectively. Then, it is easily shown that

$$\sum_{j=1}^{\infty} \frac{(\lambda_j^i - \lambda_j^o)}{\lambda_j^i \lambda_j^o} [\psi_j^o(t)][\psi_j^o(s)]^T$$

is a solution of Eq. (6.6), is symmetric, and, if

$$\sum_{j=1}^{\infty} \left(\frac{\lambda_i^i - \lambda_j^o}{\lambda_j^i \lambda_j^o} \right)^2 < \infty ,$$

is square integrable on TxT.

Returning now to Eq. (6.6), the fact that a kernel with the desired properties is known to exist does not provide us with a method for obtaining this kernel. On the other hand, Eq. (6.7) which has a similar form can be solved easily for $[h_i^\infty(t-s)]$ by the use of Fourier transforms when the signal and noise covariance functions are known to be stationary. The assumption which states that $[h_i^T(t,s)]$ can be approximated by $[h_i^\infty(t-s)]$ for $t, s\varepsilon T$ when T_f becomes large is then intended to replace the necessity for solving a difficult integral equation, Eq. (6.6), with the necessity for solving the much simpler integral equation, Eq. (6.7). To justify this assumption on a basis other than its expedience, though, let us refer to Eq. (7.124) where it can be seen that the usefulness of the operator H_i^T stems from the fact that it allows us to obtain the statistic

$$\ell_i = (H_i[y],[y])$$

since H_i^T is a realization of the desired H_i. Thus if the integral operator H_i^∞ whose kernel is $[h_i^\infty(t-s)]$ for $t, s\varepsilon T$ can be shown to provide "substantially" the same statistics when substituted for H_i^T, it would be unnecessary to make this assumption and H_i^∞ could be used directly when making an angle of arrival estimate. Unfortunately, attempts to produce conditions on the observation interval $[0,T_f]$ and the covariance functions of the signal and noise processes which guarantee this property have not met with any degree of success. Intuitively, however, this assumption is not unreasonable when we note that $H_i[y]$ can be obtained as the output of a linear filter whose impulse response is $[h_i^T(t,s)]$ and if we remember that the optimal Wiener impulse response for finite data,* whose defining integral equation is very similar to Eq. (6.6) if white noise is present, approaches the solution for the infinite interval with only minor variations near the endpoints of the interval.**

Turning now to the problem of solving Eq. (6.7), the kernel

*See Ref. 43, p. 240.
**See Ref. 61, p. 1087.

$[h_i^\infty(t-s)]$, due to Eqs. (6.3), (7.75) and Assumption A6.4.1, satisfies

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[r^o(u-t)][h_i^\infty(t-s)]\{[r^{(i)}(s-v)]+[r^o(s-v)]\}dtds = [r^{(i)}(u-v)]$$

or, equivalently,

$$\sum_{j,k\varepsilon K}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[r^o(u-t)]_{mj}[h_i^\infty(t-s)]_{jk}\{[r^{(i)}(s-v)]_{kn}+[r^o(s-v)]_{kn}\}dtds = [r^{(i)}(u-v)]_{mn}$$

for all m, $n\varepsilon K$. Furthermore, if we let $u+\tau_m(\alpha_i) = u'$, $v+\tau_n(\alpha_i) = v'$, $t+\tau_j(\alpha_i) = t'$ and $s+\tau_k(\alpha_i) = s'$ and assume

$$[h_i^\infty(t-s)]_{jk} = h(t-s+\tau_j(\alpha_i) - \tau_k(\alpha_i)), \tag{8.5}$$

the above equation reduces to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{r_N(u'-t')\}\{h(t'-s')\}\{Kr_S(s'-v')+r_N(s'-v')\}dt'ds' = r_S(u'-v') \tag{8.6}$$

when we apply Eq. (6.2) and Assumption A6.4.1. (The scalar K denotes the total number of antenna elements in the array while $\tau_j(\alpha_i)-\tau_k(\alpha_i)$, as defined in Assumption A6.3.1, is the propogation time of a signal between the jth and kth array element if α_i is its true direction of arrival.) Thus, by anticipating the form of the kernel $[h_i^\infty(t-s)]$, the requirement for solving K simultaneous integral equations can be replaced by the need for solving only <u>one</u> integral equation. By introducing the proper delays as indicated in Eq. (8.5), the kernel $[h_i^\infty(t-s)]$ can be constructed from the solution $h(t'-s')$ of Eq. (8.6). Moreover, taking the Fourier transform of both sides of Eq. (8.6), the Fourier transform of $h(\tau)$ is found to be

$$F_h(\omega) = \frac{F_S(\omega)}{F_N(\omega)[F_N(\omega)+KF_S(\omega)]} \tag{8.7}$$

where $F_h(\omega)$, $F_S(\omega)$, and $F_N(\omega)$ are the Fourier transforms of $h(\tau)$, $r_S(\tau)$, and $r_N(\tau)$, respectively.

To obtain a physical interpretation of the operator H_i^∞, let us observe Eqs. (8.5) and (8.7) where it can be seen to perform two distinct operations. First of all, H_i^∞ delays the individual components of [y]. These delays, as illustrated in Fig. 8.1, are such that the various components of [y] are in phase if α_i is the true direction of arrival. (Remember, T_f is large compared with the propagation time of signals across the array so the fact that the delayed signals are not "quite" defined over the same interval can be neglected.) These delayed signals are then added and filtered by a nonrealizable filter whose impulse response is $h(\tau)$. Looking at Eq. (8.7), it can be seen that this filter has an impulse response which is very similar to the impulse response it would have if its output was required to be a best (minimum mean square) estimate of the signal received by each array element when no delays are present. (The Fourier transform of the impulse response of such a Wiener filter would have the form

$$\frac{F_S(\omega)}{F_N(\omega)+KF_S(\omega)} \quad .)$$

Finally, the filtered signal is again delayed to produce the vector $[H_i^\infty[y]]$ whose kth component is in phase with the signal received by the kth array element when α_i is the true direction of arrival. As a result, we see that $(H_i^\infty[y],[y])$ is the sum over j, $k\varepsilon K$ of the correlation of the signals received by the jth and kth antenna, the jth antenna sig-

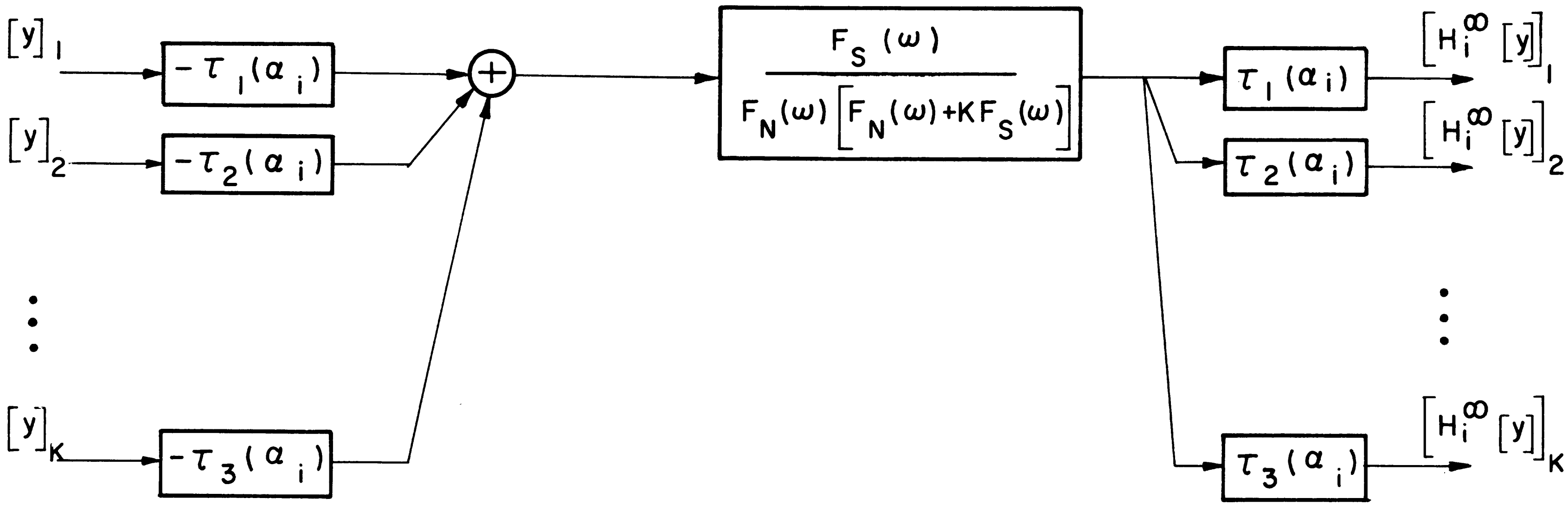

Fig. 8.1. Operations performed by H_i^{∞}.

nal being filtered and delayed so that it is in phase with the kth antenna signal when α_i is the true direction of arrival. Thus, when we calculate the sequence $\{(H_i^\infty[y],[y])\}$ which approximates the sequence $\{\ell_i\}$, we are actually "scanning" the array and obtaining a measure of the signal energy received by the array from each possible direction of arrival α_i.

Let us now see now the band-limited condition of Assumption A6.4.1 affects Eq. (8.7). As shown by Kelley, Reed, and Root,[40] when the Fourier transform of the signal covariance function has the form illustrated in Fig. 6.3, it can be written as

$$r_S(\tau) = \mathrm{Re}(r_{S_o}(\tau)e^{j\omega_o\tau}) \tag{8.8}$$

where $r_{S_o}(\tau)$ is the inverse Fourier transform of $F_S^+(\omega+\omega_o)$, $F_S^+(\omega)$ being defined by

$$\begin{aligned} F_S^+(\omega) &= F_S(\omega) \qquad \omega \geq 0 \\ &= 0 \qquad\qquad \omega < 0 , \end{aligned}$$

and where Re( ) denotes the "real part" of the indicated quantity. It can be seen that $F_S^+(\omega+\omega_o)$ is a "low-pass" function so $r_{S_o}(\tau)$, when compared with $\cos \omega_o\tau$, is a slowly varying function of τ. Kelley, Reed, and Root[40] have also shown

$$r_{S_o}(\tau) = \bar{r}_{S_o}(-\tau) \tag{8.9}$$

where the bar indicates the complex conjugate. Then, since*

*See Ref. 43, p. 365.

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j(\omega-\omega_o)\tau} d\tau = \delta(\omega-\omega_o)$$

with δ representing the Dirac delta function, we can write

$$\begin{aligned} F_S(\omega) &= \mathcal{F}\{\mathrm{Re}(r_{S_o}(\tau)e^{j\omega_o\tau})\} \\ &= \mathcal{F}\{(1/2)r_{S_o}(\tau)e^{j\omega_o\tau} + (1/2)\overline{r}_{S_o}(\tau)e^{-j\omega_o\tau}\} \\ &= \pi\{F_{S_o}(\omega)*\delta(\omega-\omega_o) + F_{S_o}(-\omega)*\ (\omega+\omega_o)\} \end{aligned} \tag{8.10}$$

by applying Eq. (8.9) and letting $F_{S_o}(\omega) = \mathcal{F}\{r_{S_o}(\omega)\}$. (The symbol $\mathcal{F}\{\ \}$ denotes the Fourier transform of the indicated quantity while "*" denotes convolution.) Furthermore, substituting (8.10) and a similar expression for $F_N(\omega)$ into (8.7), we find

$$F_h(\omega) = F_{h_1}(\omega)*\delta(\omega-\omega_o) + F_{h_2}(\omega)*\delta(\omega+\omega_o) \tag{8.11}$$

where

$$F_{h_1}(\omega) = \frac{F_{S_o}(\omega)}{F_{N_o}(\omega)[F_{N_o}(\omega)+KF_{S_o}(\omega)]}, \tag{8.12}$$

$$F_{h_2}(\omega) = \frac{F_{S_o}(-\omega)}{F_{N_o}(-\omega)[F_{N_o}(-\omega)+KF_{S_o}(-\omega)]}, \tag{8.13}$$

and $F_{N_o}(\omega) = F_N^+(\omega+\omega_o)$, $F_N^+(\omega)$ being defined by

$$\begin{aligned} F_N^+(\omega) &= F_N(\omega) \qquad \omega \geq 0 \\ &= 0 \qquad\qquad \omega < 0. \end{aligned}$$

Finally, since $F_{h_2}(\omega) = F_{h_1}(-\omega)$, the inverse Fourier transform of $F_h(\omega)$, $h(\tau)$, is given by

$$h(\tau) = \mathrm{Re}(h_1(\tau)e^{j\omega_o\tau})$$

$$= \mathrm{Re}(h_1(\tau))\cos\omega_o\tau - \mathrm{Im}(h_1(\tau))\sin\omega_o\tau \qquad (8.14)$$

where $h_1(\tau)=\mathcal{F}^{-1}(F_{h_1}(\omega))$ is the inverse Fourier transform of $F_{h_1}(\omega)$ and Im( ) denotes the "imaginary part" of the indicated quantity. Thus, according to Eq. (8.14), the kernel of the desired integral operator can be separated into high frequency and low frequency terms, $h_1(\tau)$ being slowly varying when compared with $\cos\omega_o\tau$ and $\sin\omega_o\tau$. Note that $h_1(\tau)$ is real if the signal and noise power spectrums $F_{S_o}(\omega)$ and $F_{N_o}(\omega)$ are symmetric about the point $\omega = 0$. This will be true when $F_S(\omega)$ and $F_N(\omega)$ are symmetric about $\omega = \omega_o$ and fall off sufficiently fast so $F_S(0) \approx 0$ and $F_N(0) \approx 0$. As a result, we can interpret $\mathrm{Im}(h_1(\tau))$ as a term which compensates for any lack of symmetry that might exist in these spectrums.

Let us now investigate what happens to the sequence $\{(H_i^\infty[y],[y])\}$ when H_i^∞ is the integral operator whose kernel is obtained by substituting (8.14) into (8.5). Since $h_1(\tau)$ is slowly varying compared with $\cos\omega_o\tau$ and the propagation time of signals across the array is small compared with the inverse of the bandwidth (Assumption A6.4.1), Eq. (8.5) becomes

$$[h_i^\infty(\tau)]_{jk} \approx h_{1r}(\tau)\cos(\omega_o\tau+\omega_o(\tau_j(\alpha_i)-\tau_k(\alpha_i))$$

$$- \; h_{1i}(\tau)\sin(\omega_o\tau+\omega_o(\tau_j(\alpha_i)-\tau_k(\alpha_i)) \qquad (8.15)$$

where "≈" denotes an approximation, $h_{1r}(\tau) = \mathrm{Re}(h_1(\tau))$, and $h_{1i}(\tau) = \mathrm{Im}(h_1(\tau))$. Then, applying the identities*

$$\cos(\psi_1+\psi_2) = \cos\psi_1 \cos\psi_2 - \sin\psi_1 \sin\psi_2 \tag{8.16}$$

$$\sin(\psi_1+\psi_2) = \sin\psi_1 \cos\psi_2 + \cos\psi_1 \sin\psi_2, \tag{8.17}$$

we find

$$\ell_i \approx \mathrm{Tr}\{[A_1(\alpha_i)]^T[Y_1]\} - \mathrm{Tr}\{[A_2(\alpha_i)]^T[Y_2]\} \tag{8.18}$$

where Tr{} denotes the trace of the indicated quantity while $[A_1(\alpha_i)]$, $[A_2(\alpha_i)]$, $[Y_1]$, and $[Y_2]$ are KxK matrices whose jkth elements are given by

$$[A_1(\alpha_i)]_{jk} = \cos\omega_o(\tau_j(\alpha_i) - \tau_k(\alpha_i)) , \tag{8.19}$$

$$[A_2(\alpha_i)]_{jk} = \sin\omega_o(\tau_j(\alpha_i) - \tau_k(\alpha_i)) , \tag{8.20}$$

$$[Y_1]_{jk} = \int_T \int_T [y(t)]_j[y(s)]_k\{h_{1r}(t-s)\cos\omega_o(t-s) - h_{1i}(t-s)\sin\omega_o(t-s)\}dtds, \tag{8.21}$$

and

$$[Y_2]_{jk} = \int_T \int_T [y(t)]_j[y(s)]_k\{h_{1r}(t-s)\sin\omega_o(t-s) + h_{1i}(t-s)\cos\omega_o(t-s)\}dtds, \tag{8.22}$$

respectively.

Observe, now, the fact that the matrices $[Y_1]$ and $[Y_2]$ are independent

*See Ref. 62, p. 344.

of the parameter α_i. The utility of this fact becomes apparent when we realize, if $[Y_1]$ and $[Y_2]$ are known, the calculation of our approximation to the sequence of coefficients $\{\ell_i\}$ reduces to a simple matrix operation on known matrices which is easily handled by a digital computer. To approximate ℓ_i, for example, we need only form the matrices $[A_1(\alpha_i)]$ and $[A_2(\alpha_i)]$ which are known functions of the parameter α_i and substitute into Eq. (8.18). Therefore, if we know $[Y_1]$ and $[Y_2]$, it is unnecessary, as in the previous chapter, to record the received time signals or have a large number of parallel systems, each calculating ℓ_i for a particular α_i. The matrices $[Y_1]$ and $[Y_2]$, each with K^2 elements, contain all the information necessary to construct an approximation of the sequence $\{\ell_i\}$. This means we are able, at any later time, to "scan" the array and obtain a measure of the signal energy received by the array from <u>any</u> particular direction of arrival during the time interval $[0,T_f]$. The advantage of this method of operation becomes apparent when we compare it with the normal mode of operation of existing direction finding systems where scanning takes place sequentially and, therefore, only a <u>single</u> direction is under surveillance during a particular time interval. (Similar results have been observed by Ksienski[63] for the "sure signal" case where the signal is representable as a sinusoid of known frequency but whose amplitude and phase, while being constant, are unknown in magnitude.)

Let us now investigate the matrices $[Y_1]$ and $[Y_2]$ further and determine just how they may be evaluated. In Eq. (8.21), let us split the integral over the two-dimensional subspace $(t,s)\varepsilon T x T$ into the sum of two integrals, one over each of the regions I and II indicated in Fig. 8.2. We can then write (a similar expression can be found for $[Y_2]_{jk}$)

$$[Y_1]_{jk} = \int_0^{T_f}\int_0^t [y(t)]_j [y(s)]_k \{h_{1r}(t-s)\cos \omega_o(t-s) - h_{1i}(t-s)\sin \omega_o(t-s)\}dtds$$
$$+ \int_0^{T_f}\int_0^s [y(t)]_j [y(s)]_k \{h_{1r}(s-t)\cos \omega_o(s-t) - h_{1i}(s-t)\sin \omega_o(s-t)\}dtds \qquad (8.23)$$

since $h(\tau) = h(-\tau)$, which can be realized by the use of the hardware illustrated in Fig. 8.3. In this figure, the jth and kth received signals are fed into filters with the realizable impulse response

$$\hat{h}_1(\tau) = h_{1r}(\tau) \cos \omega_o\tau - h_{1i}(\tau)\sin \omega_o\tau \qquad \tau \geq 0$$
$$= 0 \qquad \tau < 0 \qquad (8.24)$$

and then correlated with the kth and jth received signals, respectively. The outputs of these correlators are then added to obtain the matrix element $[Y_1]_{jk}$. Thus, by the use of K^2 devices of the type illustrated in Fig. 8.3, we are able to produce the K^2 elements of the matrix $[Y_1]$. Moreover, a similar procedure shows the jkth element of the matrix $[Y_2]$ can be realized by the use of the device illustrated in Fig. 8.4 where the block containing $\hat{h}_2$ is a realizable linear filter whose impulse response is given by

$$\hat{h}_2(\tau) = h_{1r}(\tau)\sin \omega_o\tau + h_{1i}(\tau)\cos \omega_o\tau \qquad \tau \geq 0$$
$$= 0 \qquad \tau < 0 . \qquad (8.25)$$

Note that the outputs of the correlators are subtracted in this case.

The element of the matrices $[Y_1]$ and $[Y_2]$ can also be produced by using mixers and local oscillators rather than the band-pass filters

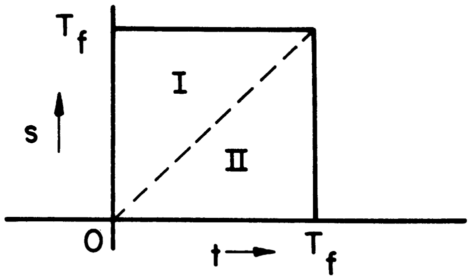

Fig. 8.2. Regions of integration for Eq. (8.23).

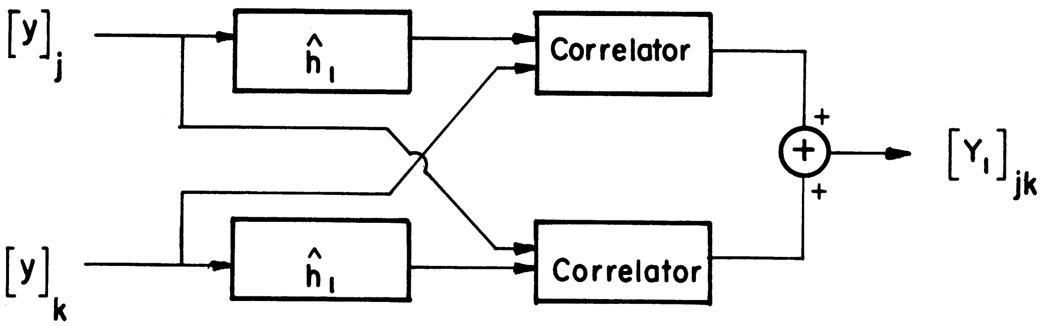

Fig. 8.3. Device for computing $[Y_1]_{jk}$.

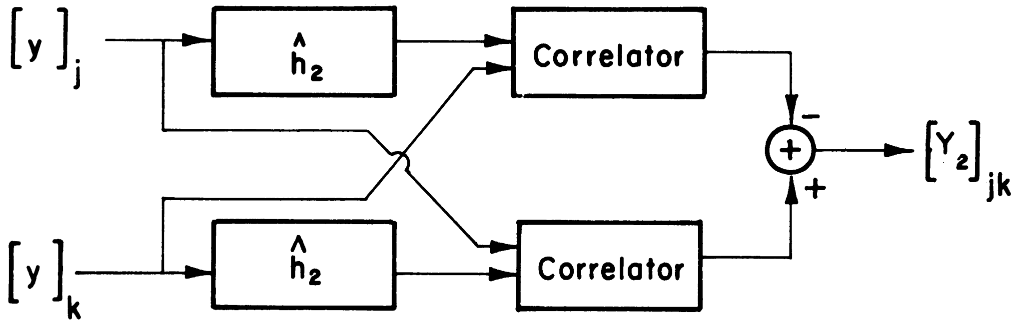

Fig. 8.4. Device for computing $[Y_2]_{jk}$.

suggested in the previous paragraph. To show this, substitute (8.16) and (8.17) into (8.21) and (8.22) to obtain

$$
\begin{aligned}
[Y_1]_{jk} = & \int_0^{T_f}\int_0^t \{[y(t)]_j^c[y(s)]_k^c + [y(t)]_j^s[y(s)]_k^s\} h_{1r}(t-s)dsdt \\
& + \int_0^{T_f}\int_0^t \{[y(t)]_j^c[y(s)]_k^s - [y(t)]_j^s[y(s)]_k^c\} h_{1i}(t-s)dsdt \\
& + \int_0^{T_f}\int_0^s \{[y(t)]_j^c[y(s)]_k^c + [y(t)]_j^s[y(s)]_k^s\} h_{1r}(s-t)dtds \\
& + \int_0^{T_f}\int_0^s \{[y(t)]_j^s[y(s)]_k^c - [y(t)]_j^c[y(s)]_k^s\} h_{1i}(s-t)dtds
\end{aligned}
\tag{8.26}
$$

where

$$[y(t)]_k^c = [y(t)]_k \cos \omega_o t \tag{8.27}$$

and

$$[y(t)]_k^s = [y(t)]_k \sin \omega_o t \ . \tag{8.28}$$

In Fig. 8.5, the hardware and circuitry necessary to implement Eq. (8.26) is illustrated. The blocks containing either $\hat{h}_{1r}$ or $\hat{h}_{1i}$ represent realizable low-pass filters with impulse responses of either

$$
\begin{aligned}
\hat{h}_{1r}(\tau) &= h_{1r}(\tau), \quad \tau \geq 0, \\
&= 0 \text{ otherwise}
\end{aligned}
\tag{8.29}
$$

or

$$
\begin{aligned}
\hat{h}_{1i}(\tau) &= h_{1i}(\tau), \quad \tau \geq 0, \\
&= 0 \text{ otherwise.}
\end{aligned}
\tag{8.30}
$$

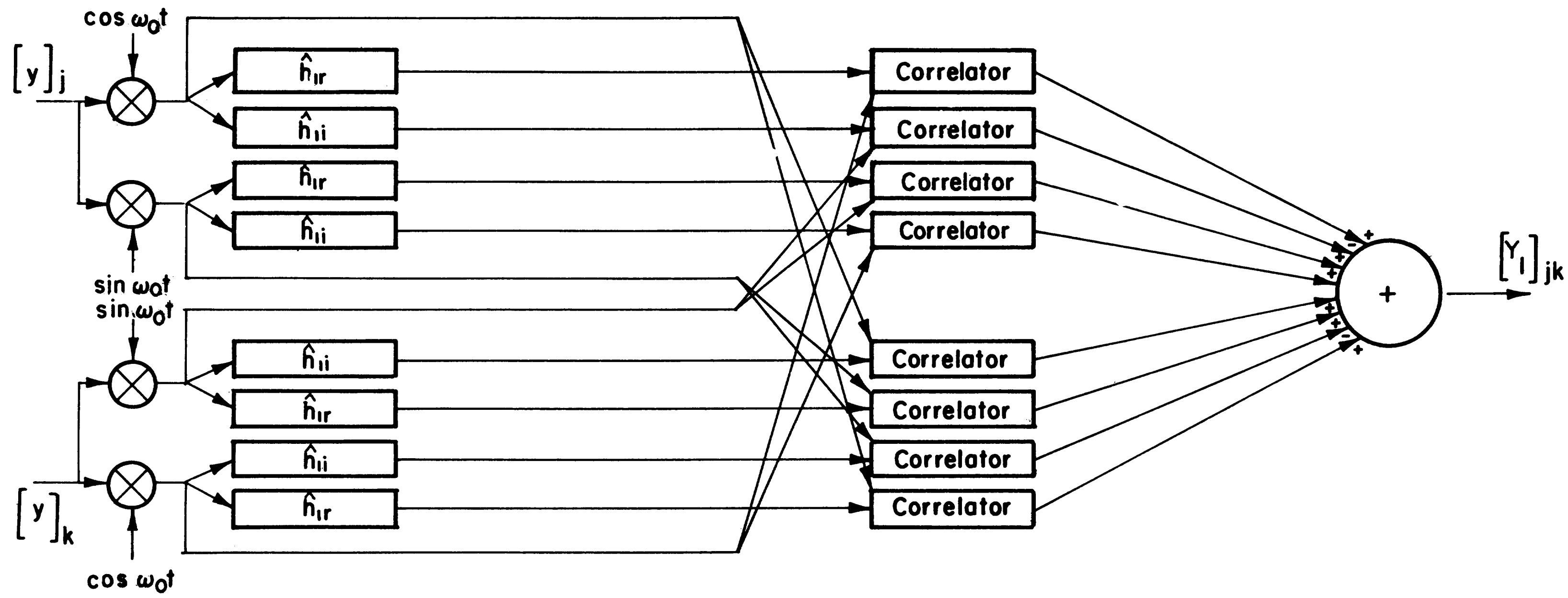

Fig. 8.5. Alternate device for computing $[Y_1]_{jk}$.

Even though Fig. 8.5 is more complicated than Fig. 8.3, from a practical standpoint, this latter approach appears to have an advantage over the one indicated in Fig. 8.3 (at least for microwave signals) in that mixers and local oscillators are usually easier to build and regulate than are band-pass filters.

A similar technique can be used to realize the elements of the matrix $[Y_2]$. Proceeding as above, we find

$$\begin{aligned}[Y_2]_{jk} = &\int_0^{T_f}\int_0^{t} \{[y(t)]_j^s[y(s)]_k^c - [y(t)]_j^c[y(s)]_k^s\}h_{1r}(t-s)dsdt \\ &+ \int_0^{T_f}\int_0^{t} \{[y(t)]_j^c[y(s)]_k^c + [y(t)]_j^s[y(s)]_k^s\}h_{1i}(t-s)dsdt \\ &- \int_0^{T_f}\int_0^{s} \{[y(t)]_j^c[y(s)]_k^s - [y(t)]_j^s[y(s)]_k^c\}h_{1r}(s-t)dtds \\ &- \int_0^{T_f}\int_0^{s} \{[y(t)]_j^c[y(s)]_k^c + [y(t)]_j^s[y(s)]_k^s\}h_{1i}(s-t)dtds \qquad (8.31)\end{aligned}$$

which can be implemented as illustrated in Fig. 8.6.

Let us now turn our attention to the problem of calculating the sequence $\{D_i\}$ which is necessary to the decision scheme presented in Section 7.4. Although techniques are available for estimating each D_i when the sequence of eigenvalues $\{\rho_j\}$ corresponding to each particular "α_i" is known, this is an unsatisfactory approach since, normally, we do not know $\{\rho_j\}$. The ideal situation would be where $D(\alpha_i)$, $D(\alpha_i) = D_i$ for all i, is some simple function of the kernels of the operators R_i and R_o since these functions are known. Although a search for such a function has not met with success, it is possible, for the special cases being considered in this chapter, to determine properties of this function which provide

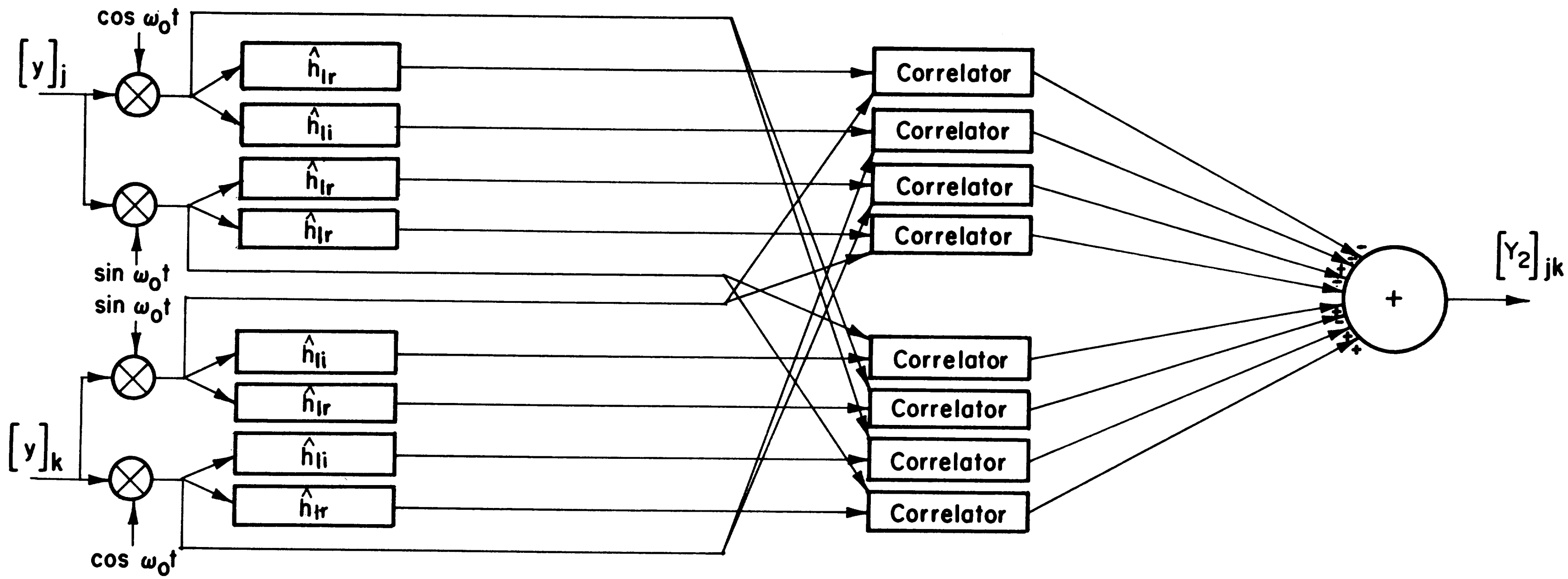

Fig. 8.6. Alternate device for computing $[Y_2]_{jk}$.

"essentially" the same information as would be available if $D(\alpha_i)$ was known exactly. In particular, we can show it is reasonable to assume $D(\alpha_i)$ is independent of α_i so that only the detection problem requires knowledge of the exact form of $D(\alpha_i)$. But, since the a priori probability of noise alone is chosen somewhat arbitrarily anyway, it can be seen that, even in the detection problem, little is gained by knowing the exact numerical value of $D(\alpha_i)$. Thus, an experimental approach which sets the level $D(\alpha_i)$ so as to obtain desired "false alarm" and "detection" probabilities appear most tractable for a real direction finding situation.

To demonstrate the relationship that exists between $D(\alpha_i)$ and α_i, return to Theorem 7.28 where $D_i = D(\alpha_i)$ is defined. According to this theorem,

$$D_i^2 = \prod_{j=1}^{\infty} (1-\rho_j)$$

where $\{\rho_j\}$ is the sequence of eigenvalues corresponding to the operator $I-X^*X$. But, since X^{-1} is the unique bounded extension of $R_o^{1/2}R_i^{-1/2}$ by Lemma 7.6 and Theorem 7.16, it follows that

$$D_i^{-2} = \prod_{j=1}^{\infty} \rho_j' \qquad (8.32)$$

where

$$\rho_j' = \frac{1}{1-\rho_j}$$

and $\{\rho_j'\}$ is the sequence of eigenvalues corresponding to $(X^*X)^{-1}$ which is unique bounded extension of $R_i^{-1/2}R_oR_i^{-1/2}$. The infinite product of the eigenvalues of an operator as specified in Eq. (8.32) is sometimes referred* to as its "determinant" so that

*See Ref. 49, p. 1029.

$$D_i^{-2} = \det\{(X^*X)^{-1}\}$$

where det{} denotes the determinant of the indicated operator. Moreover, since

$$R_i^{-1/2}R_oR_i^{-1/2} = R_i^{-1/2}R_oR_i^{-1}R_i^{1/2} = R_i^{-1/2}(I-R_oH_i)R_i^{1/2}$$

when we apply Eq. (7.112), it can be seen that the eigenfunctions and eigenvalues of $(X^*X)^{-1}$, $R_oR_i^{-1}$, and $I-R_oH_i$ are identical so

$$\det\{(X^*X)^{-1}\} = \det\{R_oR_i^{-1}\} = \det\{I-R_oH_i\} .$$

Let us now investigate the effect of substituting H_i^∞ for H_i which has been assumed to approximate H_i throughout this chapter. Suppose the operator Sf_i^+ is defined on $\overline{L}_2[T]$ so that, for each $k\varepsilon K$ and $[f]\varepsilon\overline{L}_2[T]$,

$$[Sf_i^+([f])]_k = [f(t-\tau_k(\alpha_i))]_k \quad t\varepsilon[\tau_k(\alpha_i),T_f]$$

$$= [f(t+T_f-\tau_k(\alpha_i)]_k \quad t\varepsilon[0,\tau_k(\alpha_i))$$

where the subscripted k denotes the kth component of the indicated element of $\overline{L}_2[T]$. It can be seen that Sf_i^+ maps $\overline{L}_2[T]$ into $\overline{L}_2[T]$ by shifting the individual components of the $\overline{L}_2[T]$ elements forward in time with a slight distortion near the endpoints to ensure an $\overline{L}_2[T]$ range. In a similar manner, let us define the operator Sf_i^- by

$$[Sf_i^-([f])]_k = [f(t+\tau_k(\alpha_i))]_k \quad t\varepsilon[0,T_f-\tau_k(\alpha_i)]$$

$$= [f(t-T_f+\tau_k(\alpha_i))]_k \quad t\varepsilon(T_f-\tau_k(\alpha_i),T_f]$$

for each $k\varepsilon K$ and $[f]\varepsilon\overline{L}_2[T]$. Referring now to the discussion of Fig. 8.1, we see, since $T_f \gg \max_k \tau_k(\alpha_i)$, the operator H_i^∞ can be "approximated" by

$$Sf_i^+ \, H \, Sf_i^-$$

where H is the integral operator on $\overline{L}_2[T]$ into $\overline{L}_2[T]$ for which the jkth element of its matrix kernel is given by h(t-s). This approximation for H_i^∞ appears valid and we shall not pursue further the manner in which it converges to H_i^∞ since H_i^∞ is, itself, an approximation to H_i. If we now form $I-R_o Sf_i^+ HSf_i^-$, we find

$$I - R_o Sf_i^+ HSf_i^- \;=\; Sf_i^+(I-R_o H)Sf_i^-$$

since R_o has a diagonal kernel. As a result, it can be seen $I-R_o Sf_i^+ HSf_i^-$ and $I-R_o H$ have identical eigenvalues and eigenfunctions since

$$Sf_i^+ Sf_i^- \;=\; Sf_i^- Sf_i^+ \;=\; I \; .$$

But this implies

$$\det\{I-R_o Sf_i^+ HSf_i^-\} \;=\; \det\{I-HR_o\}$$

which is easily seen to be independent of α_i. Finally, since $Sf_i^+ HSf_i^-$ "approximates" H_i, we see that it is reasonable to assume $I-R_o H_i$ and, hence, D_i is also independent of α_i.

An alternate and possible more intuitive argument for assuming $D(\alpha_i)$ is independent of α_i can be given if we rewrite Eq. (7.124) in the form

$$I_i \;=\; Pr\{\alpha_i\} \exp\{\, \tfrac{1}{2}\, \{(H_i[y],[y]) - B(\alpha_i)\}\} \qquad (8.33)$$

where $B(\alpha_i) = 2 \log D_i$. In Eq. (8.33), it can be seen that $B(\alpha_i)$ serves to introduce a bias of the statistic $(H_i[y],[y])$. Moreover, the approximation to this statistic developed above, as shown in the next chapter, is the sum of two statistics which depend on α and α_i, the true and assumed directions of arrival, respectively. One of these statistics achieves its maximum value when $\alpha_i = \alpha$ while the other is a noise term which tends to distort the sum causing the maximum of our approximation of the sequence $\{(H_i[y],[y])\}$ to be achieved when α_i is different from α. Suppose, rather than requiring $B(\alpha_i)$ to equal 2 log D_i, we substitute for $B(\alpha_i)$ the function of α_i which equals the expected value of the distorting noise term so that the term in the exponent of Eq. (8.33) is unbiased by the noise, i.e., the expected value of this term is maximum at the true direction of arrival. This approach, though possibly suboptimal, has merit since we would expect, if our original estimation technique was any good, approximately the same answers should be obtained using either bias. The advantage of using the latter bias is that it can be obtained easily. Moreover, as we shall find in the next chapter, the expected value of this noise term is independent of α_i which tends to reenforce the statements made in the previous paragraph concerning the independence of $D(\alpha_i)$ and α_i.

The summarize, the discussion of this chapter has produced a technique for approximating the sequence $\{I_i\}$ which is the crucial element of the optimal decision scheme presented in Section 7.4. This technique requires us first to obtain the KxK matrices $[Y_1]$ and $[Y_2]$ whose elements can be realized by the use of the devices illustrated in Figs. 8.3 and 8.4 or Figs. 8.5 and 8.6. Then, substituting into Eq. (8.18), the sequence $\{\ell_i\}$ can be easily approximated for any desired set of possible angles of arrival $\{\alpha_i\}$. Finally, it is a simple matter to substitute this approximation to $\{\ell_i\}$ into Eq. (7.124) letting $D(\alpha_i)$ be a preassigned constant level to be determined as discussed above and obtain the sequence $\{I_i\}$.

8.2.1 Example

It is instructive at this point to consider an example which applies the above techniques to the design of a signal processor. Suppose K elements of a receiving array are equally spaced along the y-axis as illustrated in Fig. 8.7. Suppose, also, there is a Gaussian signal source located in the x-y plane at an angle α relative to the y-axis and possessing a power spectrum

$$F_S(\omega) = \frac{E_S}{(\omega-\omega_o)^2+b_o^2} \tag{8.34}$$

where E_S is a constant which depends on the signal strength and $b_o << \omega_o$. (This last inequality will allow us to apply the "band-limited" condition discussed in Section 6.4.) A wavefront due to this signal is shown in Fig. 8.7 together with a unit vector $\underline{u}$ which is drawn orthogonal to it. Finally, suppose the observation interval $[0,T_f]$ is long compared with $1/b_o$ (b_o being a measure of the bandwidth of the signal) while the noise corrupting each received signal is Gaussian, independent, and identically distributed with a power spectrum which is much wider than the power spectrum of the signal. The noise spectrum can then be approximated as shown in Fig. 8.8 where E_N is the height of the spectrum and $a_o >> b_o$. It can now be seen that this model satisfies all the hypotheses of the Special Case Solution presented in Section 8.2 except possibly the conditions relating to the existence of a symmetric square integrable solution to Eq. (6.6) which can be approximated by the solution to Eq. (6.7). However, since these conditions are very difficult to verify, let us design a signal processor according to the techniques presented in the previous section even though it may be suboptimal. We can then evaluate its performance later by calculating the expected value of the square of the

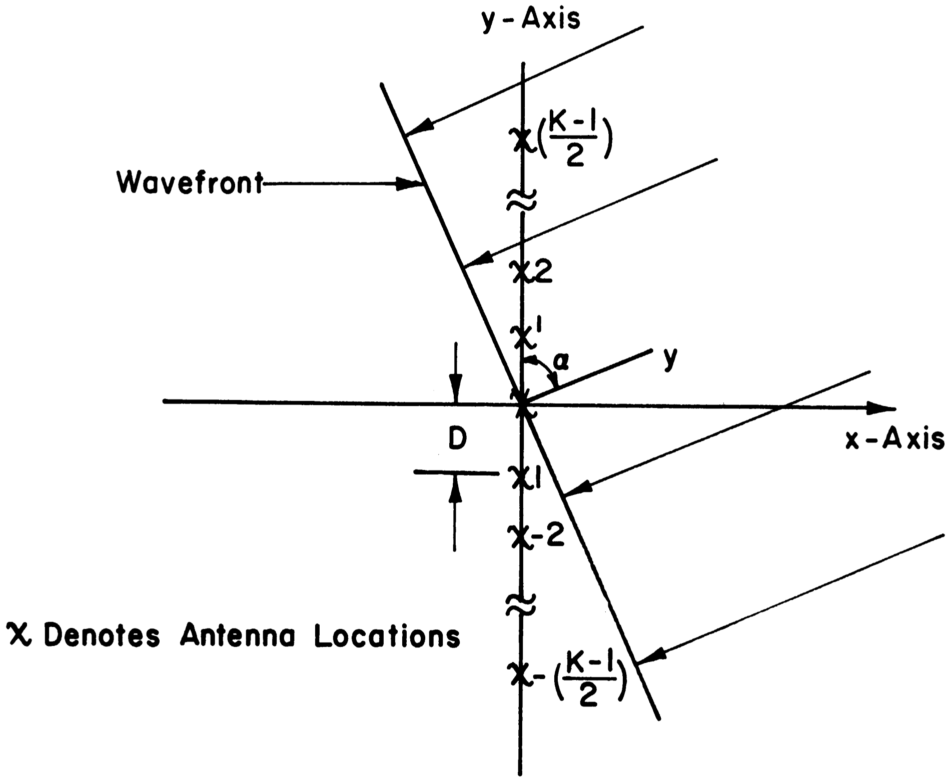

Fig. 8.7. Geometry of the receiving array relative to the incident wave.

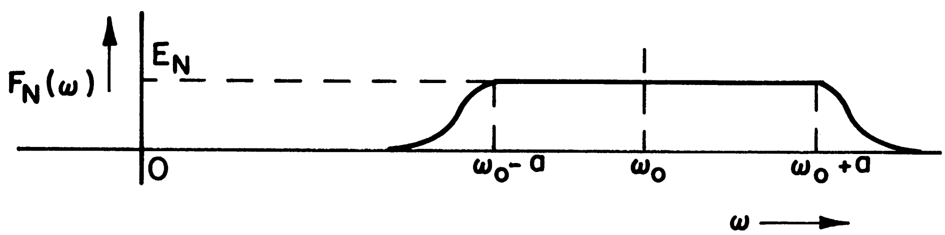

Fig. 8.8. Noise power spectrum.

resulting error. Observing the forms of $F_S(\omega)$ and $F_N(\omega)$, it can be seen that

$$F_{S_o}(\omega) = \frac{E_S}{\omega^2+b_o^2} \qquad (8.35)$$

while $F_{N_o}(\omega)$ has the form illustrated in Fig. 8.9. Then, substituting these relations into (8.12), we obtain

$$F_{h_1}(\omega) = \frac{(E_S/E_N^2)}{\omega^2+b_o^2+KE_S/E_N} \ . \qquad (8.36)$$

If we now take the inverse Fourier transform of (8.36), we find

$$h_{1i}(\tau) = 0 \qquad (8.37)$$

and

$$h_{1r}(\tau) = h_1(\tau)$$

$$= \frac{E_S}{2E_N^2\sqrt{b_o^2+KE_S/E_N}} \exp\{-|\tau|\sqrt{b_o^2+KE_S/E_N}\} \ . \qquad (8.38)$$

Finally, substituting (8.37) and (8.38) into (8.29) and (8.30) we obtain

$$\hat{h}_{1i}(\tau) = 0 \qquad (8.39)$$

and

$$\hat{h}_{1r}(\tau) = \frac{E_S}{2E_N^2\sqrt{b_o^2+KE_S/E_N}} \exp\{-\tau\sqrt{b_o^2+KE_S/E_N}\}, \ \tau \geq 0 \ , \qquad (8.40)$$

$$= 0 \text{ otherwise}$$

which are the impulse responses of the low-pass filters required in Figs. 8.5 and 8.6 to provide the elements of the matrices $[Y_1]$ and $[Y_2]$.

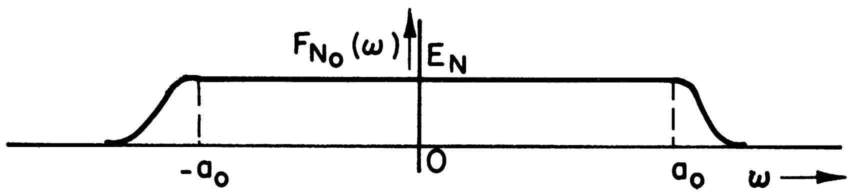

Fig. 8.9. Low-frequency noise power spectrum.

All that remains now, assuming we have the devices for realizing $[Y_1]$ and $[Y_2]$, is to provide a method for approximating the sequence $\{\ell_i\}$ by the use of Eq. (8.18) where, for this case,

$$[A_1(\alpha_i)]_{jk} = \cos[(k-j)\frac{\omega_o D}{c}\cos\alpha_i] \qquad (8.41)$$

and

$$[A_2(\alpha_i)]_{jk} = \sin[(k-j)\frac{\omega_o D}{c}\cos\alpha_i] \qquad (8.42)$$

since

$$\tau_j(\alpha_i) = -j(\omega_o D/c)\cos\alpha_i . \qquad (8.43)$$

(In Eqs. (8.41) and (8.42) we have indexed the antennas from $-(K-1)/2$ to $(K-1)/2$ as shown in Fig. 8.7 with the element at the origin being the 0th antenna.) Then, if all possible angles of arrival have equal a priori probability, our estimate of the true angle of arrival becomes the α_i corresponding to the ℓ_i possessing the largest approximation. (The modification of the above statement to handle the case where not all possible angles of arrival are equally probable is obvious.) But, assuming a digital computer is available, it can be seen that no additional hardware is necessary to produce this estimate since all the operations mentioned in this paragraph can be performed easily by such a computer. (If the detection problem is also of interest, the experimental technique mentioned above can be applied to obtain a level to be substituted for $D(\alpha_i)$.)

An evaluation of the performance of the above estimation technique will be discussed in Section 9.3.1.

8.3 INDEPENDENT, IDENTICALLY DISTRIBUTED (EXCEPT FOR AN AMPLITUDE FACTOR) NOISE

Let us now consider the problem in which Assumptions A6.4.1 and A6.4.2.1 together with A6.3.1-A6.3.3 are valid descriptions of the environment of the direction finding system. As noted in Section 6.4.2, this problem is very similar to the problem discussed in Section 8.2. In this case, however, we shall allow the noise levels to vary between antenna elements.

Following the procedure of the previous section, if we let

$$[h_i^{\infty}(t-s)]_{jk} = \frac{h'(t-s+\tau_j-\tau_k)}{a_j a_k} , \tag{8.44}$$

we find the Fourier transform of $h'(\tau)$, $F_{h'}(\omega)$, must satisfy

$$F_{h'}(\omega) = \frac{F_S(\omega)}{F_N(\omega)[F_N(\omega)+K'F_S(\omega)]} \tag{8.45}$$

where $K' = \sum_{j\varepsilon K} (1/a_j)$. Observe, the only difference between Eq. (8.45) and Eq. (8.7) is the replacement of the scalar K by the scalar K'. If we then solve for $h'_{1r}(\tau)$ and $h'_{1i}(\tau)$ as in Section 8.2, we also find

$$\ell_i \approx \mathrm{Tr}[A_1'(\alpha_i)]^T[Y_1'] - \mathrm{Tr}[A_2'(\alpha_i)]^T[Y_2']$$

where $[Y_1']$ and $[Y_2']$ are defined as $[Y_1]$ and $[Y_2]$ in Eq. (8.21) and Eq. (8.22) with $h'_{1r}(\tau)$ and $h'_{1i}(\tau)$ replacing $h_{1r}(\tau)$ and $h_{1i}(\tau)$, respectively,

$$[A_1'(\alpha_i)]_{jk} = \frac{[A_1(\alpha_i)]_{jk}}{a_j a_k} , \tag{8.46}$$

and where

$$[A_2'(\alpha_i)]_{jk} = \frac{[A_2(\alpha_i)]_{jk}}{a_j a_k} . \qquad (8.47)$$

It can now be seen that the entire solution obtained in Section 8.2 carries over to this problem with only minor variations. As a result, we shall not pursue this problem further.

CHAPTER 9

ERROR ANALYSIS

9.1 INTRODUCTION

In this chapter we shall consider the variance of the error resulting from the use of the estimation techniques discussed in the two previous chapters. Expressions relating this variance to the number of elements in the antenna array and the "signal-to-noise" ratio will be presented for the example of Section 8.2.1 in which the array elements are collinear and uniformly spaced.

9.2 ERROR ANALYSIS (General Case)

In Section 7.4, it was found that the optimal estimate of the angle of arrival corresponded to the α_i which maximized the sequence $\{I_i\}$ where

$$I_i = \{Pr\{\alpha_i\}\}\{\exp \frac{1}{2} \{(H_i[y],[y]) - 2 \log D_i\}\} \qquad (9.1)$$

with H_i being dependent of α_i and [y] denoting the observation. Let us now consider the error that results from the use of this technique when the a priori probability distribution of the angles of arrival is uniform, i.e., $Pr\{\alpha_i\} = Pr\{\alpha_j\}$ for all i and j. This choice of distribution has been made as a matter of convenience and it is possible to extend our discussion, without difficulty, to cases where this distribution is other than uniform. Moreover, since the exponential function is a monotone function of its argument, the α_i that maximizes $\{I_i\}$ also maximizes the sequence $\{\ell(\alpha_i)-B(\alpha_i)\}$ when $\ell(\alpha_i) = (H_i[y],[y])$ and $B(\alpha_i) = 2 \log D_i$. As a result, our optimality criterion reduces to that of choosing the α_i which maximizes the sequence $\{\ell(\alpha_i)-B(\alpha_i)\}$.

Now, to obtain an estimate of the resulting error, let us assume there is a continuous range of possible angles of arrival. Furthermore, let us assume the noise is "small" so that the optimal estimate $\hat{\alpha}$ is approximately equal to the true value α. If we then expand $\ell(\alpha_i)-B(\alpha_i)$ in a Taylor series, about α, retaining only the first three terms, we find

$$\begin{aligned}\ell(\alpha_i) - B(\alpha_i) \approx\ & [(H_i[y],[y])-B(\alpha_i)]\Big|_{\alpha_i=\alpha} \\ & + \left(\frac{\partial}{\partial\alpha_i}[(H_i[y],[y])-B(\alpha_i)]\Big|_{\alpha_i=\alpha}\right)(\alpha_i-\alpha) \\ & + \left(\frac{\partial^2}{\partial\alpha_i^2}[(H_i[y],[y])-B(\alpha_i)]\Big|_{\alpha_i=\alpha}\right)(\alpha_i-\alpha)^2/2 \end{aligned} \tag{9.2}$$

so that the optimal estimate is given by

$$\hat{\alpha} = \alpha + \varepsilon \tag{9.3}$$

where

$$\varepsilon \approx \frac{\dfrac{\partial}{\partial\alpha_i}[(H_i[y],[y])-B(\alpha_i)]\Big|_{\alpha_i=\alpha}}{\dfrac{\partial^2}{\partial\alpha_i^2}[(H_i[y],[y])-B(\alpha_i)]\Big|_{\alpha_i=\alpha}} \;. \tag{9.4}$$

Finally, if we replace the sample function [y] in Eq. (9.4) by the random process $[Y^\alpha]$, the various statistics of this approximation for the error can be considered. In particular, the second moment of this approximation can be used as a measure of the performance of the system for the limiting case where the received signal energy is "large" compared with the received noise energy.

The above procedure for obtaining a measure of the system performance, although intuitively meaningful, leaves several questions unanswered however. For instance, "Over what range of values of α_i and α does Eq. (9.4) represent a "good" estimate of the error?" The more fundamental question concerned with the legality of the power series expansion also exists since the functional relationship between $\ell(\alpha_i)$ and α_i is not known at this point. Finally, the problem of calculating the second moment of this approximation is not trivial since it is the ratio of two random variables.

In the next section, we shall examine the error resulting from the use of the estimation technique developed in Section 8.2. To obtain a measure of the system performance, the procedure discussed in the previous paragraphs will be modified slightly due to its above stated difficulties.

9.3 ERROR ANALYSIS (Special Case Solution Presented in Section 8.2)

Let us now restrict our attention to the Special Case Solution of Section 8.2. Recall, in this case, Assumptions A6.4.1 and A6.4.1.1 together with A6.3.1-A6.3.3 characterize the operating conditions of the direction finding system. Let us further assume the Fourier transforms of the signal and noise covariance functions are symmetric about the frequency ω_o, i.e., $F_S(\omega_o+\omega) = F_S(\omega_o-\omega)$ and $F_N(\omega_o+\omega) = F_N(\omega_o-\omega)$ for $\omega \geq 0$. Because of this last assumption, it follows from Eq. (8.12) that $h_1(\tau)$ is real and, hence, $h_{1i}(\tau) = 0$. (Actually, allowing $h_{1i}(\tau)$ to be nonzero does not significantly alter the error calculation other than causing an increase in the necessary bookkeeping. It might also be added that this assumption is valid in most realistic problems.) But this implies the matrix elements $[Y_1]_{jk}$ defined by Eq. (8.12) can be written as

$$[Y_1]_{jk} = [S_1]_{jk} + [N_1]_{jk} \tag{9.5}$$

where

$$[S_1]_{jk} = \int_T \int_T [s(t)]_j [s(s)]_k \{h_{1r}(t-s)\cos \omega_o(t-s)\} dtds \qquad (9.6)$$

and

$$[N_1]_{jk} = \int_T \int_T \{[s(t)]_j [n(s)]_k + [n(t)]_j [s(s)]_k + [n(t)]_j [n(s)]_k\}$$
$$(x)\ \{h_{1r}(t-s)\cos \omega_o(t-s)\} dtds\ . \qquad (9.7)$$

(A similar expression can be written for the matrix element $[Y_2]_{jk}$ defined in Eq. (8.22).) In Eq. (9.5), $[S_1]_{jk}$ represents the pure signal component of $[Y_1]_{jk}$ while $[N_1]_{jk}$ represents its distortion due to the presence of noise.

Suppose, now, the sample function of the signal process during the observation interval $[0,T_f]$ is given by

$$s(t) = s_1(t)\cos \omega_o t + s_2(t)\sin \omega_o t \qquad (9.8)$$

where $s_1(t) \approx s_1(t+\tau)$ and $s_2(t) \approx s_2(t+\tau)$ if $\tau \leq \max_{ij} \tau_j(\alpha_i)$, i.e., the signals $s_1(t)$ and $s_2(t)$ do not vary significantly in the time it takes a wavefront to propagate across the array. This is a very realistic form for a sample function since, in most communication problems, it can be written as

$$s(t) = A(t)\cos(\omega_o t+\phi(t))$$
$$= [A(t)\cos \phi(t)]\cos \omega_o t - [A(t)\sin \phi(t)]\sin \omega_o t$$

where the "widths" of

$$\frac{1}{T_f} \left| \int_0^{T_f} A(t)e^{-j\omega t}\, dt \right|^2$$

and

$$\frac{1}{T_f} \left| \int_0^{T_f} \phi(t) e^{-j\omega t} dt \right|^2 ,$$

as a function of ω, indicate the maximum rate at which $A(t)$ and $\phi(t)$ can vary with time. Moreover, if $A(t)$ and $\phi(t)$ are samples from ergodic* processes (an assumption often made in such problems) and T_f is large compared to the reciprocal of these "widths," it is reasonable to approximate these "widths" by the "width" of $F_S(\omega)$ which implies (Assumption A6.4.1) $A(t)$ and $\phi(t)$ as well as $A(t)\cos\phi(t)$ and $A(t)\sin\phi(t)$ do not vary significantly in the time it takes a wavefront to propagate across the array. Let us now investigate the error that results when the sample function of the signal process during the interval $[0,T_f]$ is given by Eq. (9.8).

If Eq. (9.8) represents the sample of the signal process during $[0,T_f]$, the sample received by the jth array element during this interval is given by

$$[y(t)]_j \approx s_1(t)\cos\omega_o(t+\tau_j(\alpha)) + s_2(t)\sin\omega_o(t+\tau_j(\alpha)) + [n(t)]_j \tag{9.9}$$

where α represents the true direction of arrival of the signal. This follows from the above stated fact that $s_1(t) \approx s_1(t+\tau)$ and $s_2(t) \approx s_2(t+\tau)$ if $\tau \leq \max\limits_{ij} \tau_j(\alpha_i)$. Then, substituting Eq. (9.9) into Eq. (9.6), we find

$$[S_1]_{jk} \approx (\tfrac{1}{4})\, S[A_1(\alpha)]_{jk} \tag{9.10}$$

*See Ref. 43, p. 67.

where

$$S = \int_T \int_T \{s_1(t)s_1(s) + s_2(t)s_2(s)\}h_{1r}(t-s)dtds \qquad (9.11)$$

and $[A_1(\alpha)]_{jk}$ is defined in Eq. (8.19). (To obtain this approximation, we have used Eqs. (8.16) and (8.17) and the fact that $s_1(\tau)$, $s_2(\tau)$ and $h_{1r}(\tau)$ are slowly varying when compared with $\cos \omega_o\tau$.) From Eq. (9.11), it can be seen that S is dependent on the received signal energy while being independent of the true direction of arrival of the signal. Moreover, if we write

$$[Y_2]_{jk} = [S_2]_{jk} + [N_2]_{jk} \qquad (9.12)$$

where $[S_2]_{jk}$ and $[N_2]_{jk}$ are defined as $[S_1]_{jk}$ and $[N_1]_{jk}$ with $\sin \omega_o(t-s)$ replacing $\cos \omega_o(t-s)$ in Eqs. (9.6) and (9.7), respectively, we find also

$$[S_2]_{jk} \approx (-\tfrac{1}{4})S[A_2(\alpha)]_{jk} \; . \qquad (9.13)$$

(The matrix element $[A_2(\alpha)]_{jk}$ is defined in Eq. (8.20).) As a result, Eq. (8.18) reduces to

$$\ell(\alpha_i) \approx (\tfrac{1}{4})S[\mathrm{Tr}\{[A_1(\alpha_i)]^T[A_1(\alpha)]+[A_2(\alpha_i)]^T[A_2(\alpha)]\}]$$

$$+ \mathrm{Tr}\{[A_1(\alpha_i)]^T[N_1] - [A_2(\alpha_i)]^T[N_2]\} \; . \qquad (9.14)$$

Let us now investigate the variation of the terms on the right of Eq. (9.14) as a function of α_i. If we let $\ell'(\alpha_i)$ be the function enclosed in brackets which multiplies S, i.e., if we let

$$\ell'(\alpha_i) = \{[A_1(\alpha_i)]^T[A_1(\alpha)] - [A_2(\alpha_i)]^T[A_2(\alpha)]\} \; , \qquad (9.15)$$

the application of Eq. (8.16) reduces this expression to

$$\ell'(\alpha_i) = \sum_{j,k\varepsilon K} \cos[\omega_o(\tau_j(\alpha_i)-\tau_k(\alpha_i)-\tau_j(\alpha)+\tau_k(\alpha)] \tag{9.16}$$

which is easily seen to be maximum when α_i equals the true angle of arrival α. It then follows that the first term in Eq. (9.14) is the product of a term which depends on the signal energy and a term which depends on the geometry of the array in such a manner that it achieves a maximum value when α_i equals α. The second term of Eq. (9.14), on the other hand, distorts this approximation of $\ell(\alpha_i)$ so the α_i which maximizes $\ell'(\alpha_i)$, i.e., the true direction of arrival α, is not necessarily the α_i which maximizes the approximation. Thus, in order to compute the error that results from the use of the estimation techniques derived in Chapter 8, it is necessary to investigate this distortion since our estimate of the true direction of arrival is the α_i which maximizes $\ell(\alpha_i)-B(\alpha_i)$.

At this point, let us again consider the functional form of the "bias" $B(\alpha_i)$. In Chapter 8, two arguments were given to demonstrate the independence of $B(\alpha_i)$ and α_i in the Special Case Solutions of that chapter. One of these arguments made reference to the results we have just derived. At that time, it was stated that an approximation could be written for $\ell(\alpha_i)$ which was the sum of two statistics; the first statistic was maximum when α_i and the true angle of arrival α were equal while the other statistic tended to distort the sum so the maximum of the approximation was not necessarily α. This property has now been verified. In addition, it was stated that a reasonable (though possible suboptimal) form for $B(\alpha_i)$ is the expected value of the distorting term in the approximation of $\ell(\alpha_i)$ since an unbiased estimate then results, i.e., if $B(\alpha_i)$ has this dependence on α_i, our estimate equals the true angle of arrival α when the distorting term is equal to its expected value. If we now refer to Eqs. (9.7) and (9.12) where $[N_1]_{jk}$ and $[N_2]_{jk}$ are defined and replace the sample functions $[n]_j$ and $[n]_k$ by the processes $[N]_j$ and $[N]_k$, respectively, it can be seen that the expected value of these matrix elements are zero except when j equals k. Moreover, if this result is substituted into Eq. (9.14), it is easily seen that the expected value of the second term of this equation (the distorting term) is independent of α_i which then justifies the statements made in Chapter 8 concerning this independence.

It is now possible to estimate, for a limiting case of special interest, the error that results from the application of the estimation technique derived in the previous chapter. In particular, if we consider the case where the received signal energy is much greater than the received noise energy, it follows from the above discussion that the first term in Eq. (9.14) dominates and the maximum of this sum is achieved when α_i is "near" α since $\ell'(\alpha_i)$ is a continuous function of α_i which achieves its maximum value at $\alpha_i = \alpha$. Moreover, because of the form of $\ell'(\alpha_i)$ as a function of α_i, we can approximate $\ell'(\alpha)$ in the vicinity of α by the first three terms of its power series expansion about α. (In the next section where a specific array geometry is considered, it will be possible to determine the range of values of α_i over which the approximation is valid.) As a result, Eq. (9.14), in the vicinity of α, becomes

$$\ell(\alpha_i) \approx (\tfrac{1}{4})S\ell'(\alpha) + (\tfrac{1}{4})S \frac{\partial^2}{\partial\alpha_i^2} [\ell'(\alpha_i)]\Big|_{\alpha_i=\alpha} \frac{(\alpha_i-\alpha)^2}{2} + N_i \quad (9.17)$$

where

$$N_i = \mathrm{Tr}\{[A_1(\alpha_i)]^T[N_1] - [A_2(\alpha_i)]^T[N_2]\} . \quad (9.18)$$

(Note that the second term of the power series expansion of $\ell'(\alpha_i)$ about α is not present in Eq. (9.17) because the slope of $\ell'(\alpha_i)$ is zero at $\alpha_i = \alpha$.) Then, taking the partial derivative of $\ell(\alpha_i)-B(\alpha_i)$ with respect to α_i and setting it equal to zero to obtain a necessary condition for $\ell(\alpha_i)-B(\alpha_i)$ to achieve its maximum, we find

$$\hat{\alpha} = \alpha + \varepsilon$$

where

$$\frac{\varepsilon}{4} \approx \frac{\left.\frac{\partial N_i}{\partial \alpha_i}\right|_{\alpha_i=\hat{\alpha}}}{S \left.\frac{\partial^2}{\partial \alpha_i^2} [\ell'(\alpha_i)]\right|_{\alpha_i=\alpha}} \tag{9.19}$$

and $\hat{\alpha}$ represents the optimal estimate. This approximation for the error, as stated above, remains valid as long as ε is not greater than the maximum value of $|\alpha_i-\alpha|$ for which the first three terms of the power series expansion of $\ell'(\alpha_i)$ remains a valid approximation of $\ell'(\alpha_i)$.

To obtain a measure of the error in the above case, let us now evaluate the expected value of the square of the error approximation given by Eq. (9.19). Applying the identities of Eqs. (8.16)-(8.17) and the fact that $s_1(\tau)$, $s_2(\tau)$, $h_{1r}(\tau)$, and $r_{N_o}(\tau)$ are slowly varying when compared with $\sin \omega_o\tau$ and $\cos \omega_o\tau$, we find

$$E\{\varepsilon^2\} \approx \{S \frac{\partial^2}{\partial \alpha_i^2} [\ell'(\alpha_i)]|_{\alpha_i=\alpha}\}^{-2}\{4\overline{N}^{1} \sum_{j,k,m\varepsilon K} \{\cos \omega_o(\tau_j(\alpha)-\tau_k(\alpha)-\tau_j(\hat{\alpha})+\tau_k(\hat{\alpha}))\}$$
$$(x)\{\frac{\partial}{\partial\hat{\alpha}} (\tau_j(\hat{\alpha})-\tau_k(\hat{\alpha}))\} \{\frac{\partial}{\partial\hat{\alpha}} (\tau_m(\hat{\alpha})-\tau_k(\hat{\alpha}))\}+2\overline{N}^2 \sum_{j,k\varepsilon K} \{\frac{\partial}{\partial\hat{\alpha}} (\tau_j(\hat{\alpha})-\tau_k(\hat{\alpha}))\}^2\} \tag{9.20}$$

where*

*See Ref. 64, p. 71 for a discussion of the expectation of a product of Gaussian random variables.

$$\overline{N}^1 = \int_T \int_T \int_T \int_T \{s_1(t)s_1(u)+s_2(t)s_2(u)\} r_{N_o}(s-v)h_{1r}(t-s)h_{1r}(u-v)dtdsdudv \tag{9.21}$$

and

$$\overline{N}^2 = \int_T \int_T \int_T \int_T r_{N_o}(t-u)r_{N_o}(s-v)h_{1r}(t-s)h_{1r}(u-v)dtdsdudv \ . \tag{9.22}$$

(Remember, this expectation has been calculated assuming $s_1(t)$ and $s_2(t)$ are known functions of time with only the noise processes being allowed a random nature.) Thus, the above measure of system performance is seen to be a function of terms which are independent of the geometry of the array, i.e., $\overline{N}^1$, $\overline{N}^2$, and S, and terms which depend <u>only</u> on the geometry, i.e., $\frac{\partial^2}{\partial \alpha_i^2} [\ell'(\alpha_i)]\Big|_{\alpha_i=\alpha}$ and the two indicated summations.

Before evaluating a specific system by the use of Eq. (9.20), let us derive an additional property of the function $\ell'(\alpha_i)$ which appears in this equation. Applying the identity of Eq. (8.16) to Eq. (9.16), we find

$$\ell'(\alpha_i) = \left[\sum_{k\epsilon K} \cos \omega_o(\tau_k(\alpha_i)-\tau_k(\alpha))\right]^2 + \left[\sum_{k\epsilon K} \sin \omega_o(\tau_k(\alpha_i)-\tau_k(\alpha))\right]^2 . \tag{9.23}$$

Let us now assume the receiving antenna array possesses the geometrical symmetry, when a vector is drawn from a particular point (a point of symmetry) to any element of the array, there exists another element of the array for which a vector drawn from the point of symmetry to this element is the negative of the first vector. The position occupied by the center element of a linear array of an odd number of equally spaced elements is an example of such a point of symmetry. If this symmetry exists, with a little reflection we can see the second term in Eq. (9.23) is zero. For every element in this sum, there exists another element also in this sum whose value is the negative of the first. Thus, for this case, we have

$$\ell'(\alpha_i) = \left[\sum_{k\epsilon K} \cos \omega_o(\tau_k(\alpha_i)-\tau_k(\alpha))\right]^2 . \qquad (9.24)$$

9.3.1 Example

In this section we shall evaluate the performance of the system which was designed in Section 8.2.1. This evaluation will use Eq. (9.20) to demonstrate the variation of the expected value of the squared error with various parameters such as the "signal-to-noise" ratio and the total number of antennas in the receiving array.

Let us begin by simplifying the expression for $\ell'(\alpha_i)$ which, in this case, has the form presented in Eq. (9.24). Substituting Eq. (8.43) into Eq. (9.24), we find

$$\ell'(\alpha_i) = \left\{\sum_{k=-K'}^{K'} \cos kF(\alpha_i,\alpha)\right\}^2 \qquad (9.25)$$

where $K' = (K-1)/2$ and

$$F(\alpha_i,\alpha) = \frac{\omega_o D}{c}(\cos \alpha - \cos \alpha_i) . \qquad (9.26)$$

But Eq. (9.25) can also be written as

$$\ell'(\alpha_i) = \left\{\mathrm{Re}\left[\exp\{-jK'F(\alpha_i,\alpha)\} \sum_{k=0}^{2K'} \exp\{jkF(\alpha_i,\alpha)\}\right]\right\}^2 \qquad (9.27)$$

where j is the complex number $\sqrt{-1}$ and Re[ ] denotes the "real part" of the indicated quantity. Then, since the sum on the right side of Eq. (9.27) is a Geometrical Progression,* it reduces to

*See Ref. 61, p. 317.

$$\left\{ \frac{[1-\exp\{j(2K'+1)F(\alpha_i,\alpha)\}]}{[1-\exp\{jF(\alpha_i,\alpha)\}]} \right\}^2$$

so that $\ell'(\alpha_i)$ becomes

$$\ell'(\alpha_i) = \left\{ \text{Re} \left[\frac{\exp\{j(K' + \frac{1}{2})F(\alpha_i,\alpha)\} - \exp\{-j(K' + \frac{1}{2})F(\alpha_i,\alpha)\}}{\exp\{j(1/2)F(\alpha_i,\alpha)\} - \exp\{-j(1/2)F(\alpha_i,\alpha)\}} \right] \right\}^2$$

$$= \left\{ \frac{\sin[(K' + \frac{1}{2})F(\alpha_i,\alpha)]}{\sin[\ (1/2)F(\alpha_i,\alpha)]} \right\}^2 . \qquad (9.28)$$

Finally, taking the second partial derivative of the expression in Eq. (9.28) with respect to α_i and evaluating it at $\alpha_i = \alpha$, we obtain

$$\frac{\partial^2}{\partial \alpha_i^2} [\ell'(\alpha_i)]\Big|_{\alpha_i=\alpha} = -(D'\sin\alpha)^2 \left(\frac{2K'+1}{6}\right)^2 [(2K'+1)^2-1]$$

$$\approx -(D'\sin\alpha)^2 \frac{K^4}{6} , \qquad (9.29)$$

since $(2K'+1) = K >> 1$, where $D' = \omega_o D/c$.

It is now possible to answer the question that was posed earlier concerning the range of values of α_i and α for which the first three terms of the power series expansion of $\ell'(\alpha_i)$ remains a valid approximation of $\ell'(\alpha_i)$. Comparing the value of $\ell'(\alpha_i)$ as specified in Eq. (9.28) with its series approximation, we find the approximation to be "good" as long as

$$|\cos\alpha_i - \cos\alpha| < \frac{c}{\omega_o DK} \qquad (9.30)$$

and

$$|1 - \cos\alpha| > \frac{2c}{\omega_o DK} . \qquad (9.31)$$

The region specified by Eq. (9.30) is approximated by the region between half-power points of the antenna pattern of the array when operating as a phased array and oriented in the α direction. However, α must be greater than a "beam-width" from the $\alpha = 0$ direction as stated by Eq. (9.31). (If $\alpha = 0$, the third term of the power series expansion is zero as can be seen from Eq. (9.29).)

Turning now to the summation multiplying the $\overline{N}^1$ term in Eq. (9.20), we find

$$\Big| \sum_{j,k,m=-K'}^{K'} \{\cos \omega_o(\tau_j(\hat{\alpha})-\tau_m(\hat{\alpha})-\tau_j(\hat{\alpha})+\tau_m(\hat{\alpha}))\}\{ \frac{\partial}{\partial\hat{\alpha}} [\omega_o(\tau_j(\hat{\alpha})-\tau_k(\hat{\alpha}))]\}$$

$$\{ \frac{\partial}{\partial\hat{\alpha}} [\omega_o(\tau_m(\hat{\alpha})-\tau_k(\hat{\alpha}))\}\Big|$$

$$\leq \sum_{j,k,m=-K'}^{K'} \Big|\{ \frac{\partial}{\partial\hat{\alpha}} [\omega_o(\tau_j(\hat{\alpha}))-\tau_k(\hat{\alpha}))]\}\{ \frac{\partial}{\partial\hat{\alpha}} [\omega_o(\tau_m(\hat{\alpha})-\tau_k(\hat{\alpha}))]\}\Big|$$

$$= (D'\sin \hat{\alpha})^2 \Sigma_1 \qquad (9.32)$$

where

$$\Sigma_1 = \sum_{j,k,m=K'}^{K'} |(j-k)(m-k)| \ . \qquad (9.33)$$

Similarly, the summation multiplying the $\overline{N}^2$ term in Eq. (9.20) is given by

$$\sum_{j,k=-K'}^{K'} \{ \frac{\partial}{\partial\hat{\alpha}} [\omega_o(\tau_j(\hat{\alpha})-\tau_k(\hat{\alpha}))]\}^2 = (D'\sin \hat{\alpha})^2 \Sigma_2 \qquad (9.34)$$

where

$$\Sigma_2 = \sum_{j,k=-K'}^{K'} (j-k)^2 \ . \qquad (9.35)$$

We are now in a position to present an explicit expression which bounds the expected value of the squared error in this example. Substituting Eqs. (9.29), (9.32), and (9.34) into Eq. (9.20), we obtain the relation

$$E\{\varepsilon^2\} < \left(\frac{6\sqrt{2}}{D'\sin\alpha}\right)^2 \left[2\,\frac{\overline{N}^1}{S^2}\left(\frac{\Sigma_1}{K^8}\right) + \frac{\overline{N}^2}{S^2}\left(\frac{\Sigma_2}{K^8}\right)\right] \qquad (9.36)$$

which shows clearly how an upper bound on $E\{\varepsilon^2\}$ varies with the amount of signal and noise present as well as the number of elements in the antenna array. (The quantities $\overline{N}^1$ and $\overline{N}^2$ will be shown later to be positive.) The terms $(\overline{N}^1/S^2)$ and $(\overline{N}^2/S^2)$ represent generalized "signal-to-noise" ratios while the terms (Σ_1/K^8) and (Σ_2/K^8) depend only on the size of the array.

In order to give Eq. (9.36) a numerical interpretation, let us further assume $s_1(t)$ and $s_2(t)$ are samples from independent Gaussian processes whose power spectrums are each equal to $F_{S_O}(\omega)$. If this is the case, the sample function $s(t)$ defined in Eq. (9.8) is a sample function from a Gaussian process whose power spectrum is $F_S(\omega)$. Let us also assume $s_1(t)$ and $s_2(t)$ have Fourier transforms whose squared magnitude equal the expectation of the squared magnitude of these transforms over the probability space. We can then write*

$$\lim_{T_f\to\infty} \left|\mathcal{F}\left\{\frac{1}{T_f}\, s_1(t')\right\}\right|^2 = \lim_{T_f\to\infty} \left|\mathcal{F}\left\{\frac{1}{T_f}\, s_2(t')\right\}\right|^2 = F_{S_O}(\omega)$$

where $\mathcal{F}\{\ \}$ is the Fourier transform of the indicated quantities and $t' = t - T_f/2$ with $s_1(t')$ and $s_2(t')$ defined to be zero outside the interval $[-T_f/2, T_f/2]$. Moreover, since T_f has been assumed "large" compared with the reciprocal of the "bandwidth" of $F_{S_O}(\omega)$, it is reasonable to approximate $|\mathcal{F}\{s_1(t)\}|^2$ and $|\mathcal{F}\{s_2(t)\}|^2$ by $T_f F_{S_O}(\omega)$.

*See Ref. 43, p. 108.

Substituting the above approximation for $|\mathcal{F}\{s_1(t)\}|^2$ and $|\mathcal{F}\{s_2(t)\}|^2$ together with Eq. (8.36) into Eq. (9.11), we find

$$S \approx T_f \int_{-\infty}^{\infty} F_{S_o}(\omega) F_{h_1}(\omega) d\omega$$
$$= \pi \frac{T_f}{K} \left(\frac{E_S}{E_N}\right) \left\{ \frac{1}{b_o} - \frac{1}{\sqrt{b_o^2 + KE_S/E_N}} \right\} . \tag{9.37}$$

To evaluate the quantities $\overline{N}^1$ and $\overline{N}^2$ which are defined in Eqs. (9.21) and (9.22), let us first realize that $h_{1r}(\tau)$ has "approximately" the same form as $r_{S_o}(\tau)$. Then, since $r_{N_o}(\tau)$ has a Fourier transform which remains flat until $\omega \approx a_o$ and since $a_o >> b_o$ (see Fig. 8.9), it follows that $r_{N_o}(\tau)$ is very "narrow" compared with $r_{S_o}(\tau)$ and $h_{1r}(\tau)$. Applying this fact, Eqs. (9.21) and (9.22) reduce to

$$\overline{N}^1 \approx E_N \int_T \int_T \int_T \{s_1(t)s_1(u)+s_2(t)s_2(u)\} h_{1r}(t-v) h_{1r}(u-v) dt du dv \tag{9.38}$$

and

$$\overline{N}^2 \approx (E_N)^2 \int_T \int_T \{h_{1r}(t-s)\}^2 \, dt ds, \tag{9.39}$$

respectively. Then, substituting the approximation for $|\mathcal{F}\{s_1(t)\}|^2$ and $|\mathcal{F}\{s_2(t)\}|^2$ discussed above together with Eq. (8.38) into Eq. (9.38), we find

$$\overline{N}^2 \approx E_N T_f \int_{-\infty}^{\infty} F_{S_o}(\omega) \{F_{h_1}(\omega)\}^2 d\omega$$
$$= \pi \frac{T_f}{K^2} \left(\frac{E_S}{E_N}\right) \left\{ \frac{1}{b_o} - \frac{1}{\sqrt{b_o^2 + KE_S/E_N}} - \frac{1}{2} \frac{K(E_S/E_N)}{(b_o + KE_S/E_N)^{3/2}} \right\} . \tag{9.40}$$

(It can be seen from Eqs. (9.39) and (9.40) that $\overline{N}^1$ and $\overline{N}^2$ are positive quantities as stated earlier.) Moreover, if we let $\tau = t-s$, Eq. (9.39) becomes

$$\overline{N}^2 \approx (E_N)^2 \int_0^{T_f} \int_{t-T_f}^{t} \{h_{1r}(\tau)\}^2 d\tau dt$$

$$= \frac{\sqrt{2}\, T_f}{(b_o^2+KE_S/E_N)^{3/2}} \left(\frac{E_S}{E_N}\right)^2 \left[1 + \frac{\exp\{-2T_f\sqrt{b_o^2+KE_S/E_N}\}}{2T_f\sqrt{b_o^2+KE_S/E_N}} - \frac{1}{2T_f\sqrt{b_o^2+KE_S/E_N}}\right]$$

$$\approx \frac{\sqrt{2}\, T_f}{(b_o^2+KE_S/E_N)^{3/2}} \left(\frac{E_S}{E_N}\right)^2 . \qquad (9.41)$$

To obtain this last result, we have substituted Eq. (8.38) and performed the indicated integration. Finally, since we are considering the large signal case, Eqs. (9.37), (9.40), and (9.41) reduce to

$$S \approx \pi \frac{T_f}{Kb_o} \left(\frac{E_S}{E_N}\right), \qquad (9.42)$$

$$\overline{N}^1 \approx \frac{\pi T_f}{K^2 b_o} \left(\frac{E_S}{E_N}\right), \qquad (9.43)$$

and

$$\overline{N}^2 \approx \frac{\sqrt{2}\, T_f}{K^{3/2}} \left(\frac{E_S}{E_N}\right)^{1/2}, \qquad (9.44)$$

respectively.

The only quantities in Eq. (9.36) which still require further simplification are the summations Σ_1 and Σ_2. But, since*

$$\sum_{k=0}^{K'} k = \frac{K'(K'+1)}{2} ,$$

*See Ref. 62, p. 317.

$$\sum_{k=-K'}^{K'} k^2 = K' \frac{(K'+1)2K'+1)}{3} ,$$

and K >> 1, it follows easily that

$$\Sigma_1 \leq 3(2K'+1)(K')^2(K'+1)^2 + \frac{K'(K'+1)(2K'+1)^3}{3}$$
$$\approx \frac{K^5}{4} \tag{9.45}$$

and

$$\Sigma_2 = 2K' \frac{(K'+1)(2K'+1)^2}{3}$$
$$\approx \frac{K^4}{6} . \tag{9.46}$$

Finally, substituting Eqs. (9.42)-(9.46) into Eq. (9.36), we obtain the desired relation

$$E\{\varepsilon^2\} < \frac{1}{\pi}\left(\frac{1}{K^3}\right)\left(\frac{6}{D'\sin \alpha}\right)^2\left(\frac{b_o}{T_f}\right)\left(\frac{E_N}{E_S}\right) \tag{9.47}$$

which illustrates an upper bound on the expected value of the squared error as a function of the parameters E_N, E_S, K, b_o, T_f, D', and α. Remember, however, this expression is valid only in the "large" signal case where α is greater than a "beamwidth" from the α = 0 direction.

CHAPTER 10

SUMMARY AND CONCLUSIONS

The research described in this dissertation is concerned with the determination of optimal procedures for estimating directions of arrival of signals emitted by stochastic sources. To make these estimates, it is assumed that an array of K omnidirectional antennas is available for gathering data. Two distinct approaches to the problem are pursued which are described in Chapter 1 along with a general discussion of existing Direction Finding techniques. The first approach which is investigated is Chapters 2-5 applies "Stochastic Optimal Control Theory" and assumes the array is operating as a phased array whose control laws for directing the pointing angle and specifying the beam width of the array are desired. In addition, the forms of optimal filters for processing the received signals of the phased array are investigated. In Chapters 6-9, the array functions only as an information gathering device and "Estimation Theory" is applies to determine the processing that is necessary to produce optimal direction-of-arrival estimates.

In Chapter 2, a greatly simplified model of a phased array direction finding system is constructed in hopes of obtaining a model which approximates the operating conditions of a realistic system and which allows a meaningful analysis. In Chapter 3, this model is investigated under the open-loop operation conditions where the control laws are required to be independent of past controls and past signals received by the array and, thus, are chosen a priori. Results are obtained for this case with the added assumption that the a priori probability density of the possible angles of arrival is concentrated in a "narrow" region about the a priori expected value of this angle. These results indicate the system should operate with the array pointing angle directed so the maximum slope of the antenna pattern is oriented in the a priori expected direction of arrival. In addition, if the signal level is unknown and the antenna pattern is symmetric, the two orientations which satisfy this condition should be used for equal time intervals. In Chapter 4, this model is investigated under the closed-loop conditions where the control laws are allowed to be dependent on past controls and past signals received by the array. An optimal filter of the signals together with an integral equation whose solution represents the optimal controls are derived. In Chapter 5, the results of Chapters 2-4 are discussed more fully.

In Chapter 6, a new model for a direction finding system is constructed which employed the antenna array only to gather information. Various special cases which restrict the possible forms of the signal and noise processes of this model are also discussed. These special cases all require the signal and noise processes to be narrow-band and the observation interval to be long compared with the reciprocal of the bandwidth. Chapter 7 contains a derivation of an optimal estimation technique for this model without the narrow-band and observation interval constraints. It is found that the Radon-Nikodym derivative of a particular Gaussian probability measure with respect to another Gaussian probability measure plays a key role in this technique. Forms for this derivative are found which allow an evaluation by the use of physical devices. In Chapter 8, the narrow-band and observation interval constraints are introduced and result in a greatly simplified solution for cases which are of engineering importance. Finally, Chapter 9 contains an analysis of the errors that result when the estimation techniques of Chapter 7 and Chapter 8 are applied.

Comparing the two approaches, we find the application of "Estimation Theory" techniques results in a solution which allows implementation with existing hardware while the application of "Stochastic Optimal Control Theory" techniques produces easily implemented results only in very restricted situations. The difficulty associated with the control theory approach stems from the infancy of the existing theory of "Stochastic Optimal Control" when nonlinearities are present. "Estimation Theory," on the other hand, possesses many powerful tools which are applicable to the Direction Finding Problem.

To summarize, the contributions of this dissertation are then the following:

1. Control Theory and Estimation Theory Techniques have been applied to the Direction Finding Problem with the result that only the latter have been found to produce a solution which is easily implemented.

2. The work of T. T. Kadota on the "Optimal Reception of M-ary Gaussian Signals in Gaussian Noise" has been refined and extended to the vector valued signal case so that it can be applied in the design of DF systems employing arrays of omnidirectional receiving antennas.

3. In two narrow-band signal cases, a relatively easily implemented technique has been found for evaluating the Radon-Nikodym derivatives which are called for by the estimation procedure following from the above extension.

Numerous other special cases exist which restrict the applicability of the general solution presented in Chapter 7 while remaining of great practical interest. For example, the introduction of multiple narrow-band signal sources produces a problem whose analysis is different from that required when a single signal source is present. Also, the presence of noise sources which possess direction properties typify many realistic situations. Both of these cases, conceptually at least, fit easily into the framework of the analysis presented in Chapter 7. However, it is not clear at this time that simplified solutions of the form derived in Chapter 8 are possible. Finally, there are several unanswered theoretical questions which are discussed in the text of this report and which deserve further investigation.

APPENDIX I

A RELATIONSHIP BETWEEN σ-FIELDS

Let $[Y] = \{[Y];t\varepsilon T\}$ be a K-dimensional vector valued random process on the probability space $(\tilde{\Omega}, \tilde{\mathcal{B}}, \tilde{P})$ and let Ω be the space of all real, K-dimensional, vector valued functions on the interval $[0,T_f]$. Furthermore, let $\mathcal{D}$ be the class of sets of the form

$$D = \{[f]\varepsilon\Omega;[f(t_1)] \leq \underline{a}_1, [f(t_2)] \leq \underline{a}_2, \ldots, [f(t_n)] \leq \underline{a}_n\}$$

where n is any arbitrary finite positive integer, $t_i \varepsilon T$ for $i\varepsilon\{1,2,\ldots n\}$, and $\underline{a}_j$, $j\varepsilon\{1,2,\ldots,n\}$, is any real vector in K-dimensional Euclidean space. Define $\tilde{\mathcal{D}}$ to be the class of sets of the form

$$\tilde{D} = [Y]^{-1}(D)$$

where $D\varepsilon\mathcal{D}$. Then

$$\mathcal{B}(\tilde{\mathcal{D}}) = [Y]^{-1}(\mathcal{B}(\mathcal{D})) .$$

<u>Proof</u>: Let $\tilde{\mathcal{B}} = [Y]^{-1}(\mathcal{B}(\mathcal{D}))$. Since

$$[Y]^{-1}(\bigcup_{j\varepsilon J} B_j) = \bigcup_{j\varepsilon J} [Y]^{-1}(B_j) ,$$

$$[Y]^{-1}(\Omega - B) = [Y]^{-1}(\Omega) - [Y]^{-1}(B) ,$$

$$[Y]^{-1}(\Omega) = \tilde{\Omega} ,$$

for J, a countable set, and B, $B_j \varepsilon \mathcal{B}(\mathcal{D})$, we see that $\tilde{\mathcal{B}}$ is a σ-field. Also, $\tilde{\mathcal{D}}$ is easily seen to be contained in $\tilde{\mathcal{B}}$. Therefore

$$\mathcal{B}(\tilde{\mathcal{D}}) \subset \tilde{\mathcal{B}} .$$

Now let $\mathcal{M}$ be the class of sets $M \subset \Omega$ such that

$$[Y]^{-1}(M) \varepsilon \mathcal{B}(\tilde{\mathcal{D}}) .$$

It can be seen that $\mathcal{M}$ is also a σ-field with $\mathcal{D} \subset \mathcal{M}$ so that

$$\mathcal{B}(\mathcal{D}) \subset \mathcal{M}. \tag{I.1}$$

Thus, we have

$$\mathcal{B}(\tilde{\mathcal{D}}) \subset \tilde{\mathcal{B}} = [Y]^{-1}(\mathcal{B}(\mathcal{D})) \subset [Y]^{-1}(\mathcal{M}) = \mathcal{B}(\tilde{\mathcal{D}})$$

or

$$\mathcal{B}(\tilde{\mathcal{D}}) = \tilde{\mathcal{B}}.$$

Q.E.D.

APPENDIX II

PROPERTIES OF THE OPERATOR R_i

The integral operator R_i defined in Eq. (7.75) takes elements of $\overline{L}_2[T]$ into elements of $\overline{L}_2[T]$ and is self-adjoint, positive definite, and Hilbert-Schmidt.

Proof: Let us first show, if $[g] = R_i([f])$ where $[f]\epsilon\overline{L}_2[T]$, then $[g]\epsilon\overline{L}_2[T]$. If $[f]_k$ denotes the kth component of $[f]$, the jth component of $[g]$ given by

$$[g]_j = \sum_{k\epsilon K} \int_T [r^i(t,s)]_{jk} \cdot [f(s)]_k ds .$$

Then, due to Theorem 7.6, we have

$$[g]_j^2 \leq \sum_{k\epsilon K} \int_T [r^i(t,s)]_{jk}^2 ds \int_T [f(s)]_k^2 ds$$

and

$$\int_T [g]_j^2 dt \leq \sum_{k\epsilon K} \int_T \int_T [r^i(t,s)]_{jk}^2 ds \int_T [f(s)]_k^2 ds$$
$$< \infty \qquad \text{(II.1)}$$

where the final inequality follows from Assumption A6.3.1 and the fact that $[f]_k\epsilon L_2[T]$. It now follows from the definition of $\overline{L}_2[T]$ that $[g]\epsilon\overline{L}_2[T]$.

Let us now prove that R_i is self-adjoint. If $[f],[g]\epsilon\overline{L}_2[T]$, then

$$\begin{aligned}(R_i([f]),[g]) &= \sum_{j\epsilon K} \int_T [g(t)]_j [R_i([f])]_j dt \\ &= \sum_{j,k\epsilon K} \int_T\int_T [r^i(t,s)]_{jk}[g(t)]_j[f(s)]_k \, dtds \\ &= ([f],R_i([g])) . \qquad \text{(II.2)}\end{aligned}$$

To prove that R_i is positive definite, note that R_o automatically has this property due to Assumption A6.3.1 and that, for $i\varepsilon\{1,2,\ldots\}$,

$$[r^i(t,s)] = [r^o(t,s)] + [r^{(i)}(t,s)] \qquad \text{(II.3)}$$

where $[r^o(t,s)]$ is the noise covariance function and $[r^{(i)}(t,s)]$ is the covariance function of the signal process $[S^i(t)]$ whose angular location is α_i (see Assumption A6.3.1). As a result, if $R_{(i)}$ is the integral operator defined on $\overline{L}_2[T]$ with kernel $[r^{(i)}(t,s)]$, it is sufficient to prove $R_{(i)}, i\varepsilon\{1,2,\ldots\}$, is positive since $R_i = R_o+R_{(i)}$. For any $[f]\varepsilon\overline{L}_2[T]$, we have

$$(R_{(i)}([f]),[f]) = \int_T \int_T \int_\Omega U(t)U(s)(\overline{P}_{(i)}xtxs)dtds \qquad \text{(II.4)}$$

where $U(t) = [\overline{S}^i(t)]^T[f(t)]$ and $[\overline{S}^i]$ is a measurable, zero mean Gaussian process with $[r^{(i)}(t,s)]$ as its covariance function. (The process $[\overline{S}^i]$ exists due to Theorem 7.9.) But Eq. (II.4) reduces to

$$\int_\Omega \{\int_T U(t)dt\}^2 d\overline{P}_{(i)} \geq 0 \qquad \text{(II.5)}$$

if the order of integration can be reversed. This order can be reversed, by Theorem 7.11, if $U(t)U(s)$ or equivalently, by Theorem 7.4, if $|U(t)U(s)|$ is integrable on the product space $\Omega xTxT$. Due to Theorem 7.6,

$$\left[\int_{\Omega xTxT} |U(t)U(s)|d\lambda(\overline{P}_{(i)}xtsx)\right]^2 \leq \int_{\Omega xTxT} (U(t))^2 d\lambda \int_{\Omega xTxT} (U(s))^2 d\lambda$$

where λ is the product measure induced on $\Omega xTxT$. But, by Theorem 7.10,

$$\int_{\Omega xTxT} (U(t))^2 d\lambda = \int_T \int_T \int_\Omega (U(t))^2 d\overline{P}_{(i)} dtds$$

$$= T_f \sum_{j,k\varepsilon K} \int_T [r^{(i)}(t,t)]_{jk}[f(t)]_j[f(t)]_k dt \ . \qquad \text{(II.6)}$$

Moreover, the right hand integral of (II.6) is finite since

$$\int_T |[r^{(i)}(t,t)]_{jk}[f(t)]_j[f(t)]_k|dt \leq \{\int_T [[r^{(i)}(t,t)]_{jk}[f(t)]_j]^2 dt\}^{1/2}$$

$$(x)\ \{\int_T [f(t)]_k^2 dt\}^{1/2} \leq b^i\{\int_T [f(t)]_j^2 dt\}^{1/2}\{\int_T [f(t)]_k^2 dt\}^{1/2},$$

again by Theorem 7.6, where b^i is the bound on the function $[r^{(i)}(t,s)]_{jk}$ discussed in Assumption A6.3.1. Thus $R_{(i)}$ is positive and $R_i, i\varepsilon M_o'$, is positive definite.

To prove that R_i is Hilbert-Schmidt, we shall first show that it is completely continuous. But Assumption A6.3.1 together with Theorem 7.18 and Theorem 7.19 imply this statement. Then, since R_i is positive definite and completely continuous, Theorem 7.20 says that there exists an orthonormal basis, $\{[\psi_n^i]\}$, for $\overline{L}_2[T]$ whose elements are eigenfunctions of R_i and an associated sequence of eigenvalues $\{\lambda_n^i\}$. Let us now consider

$$\int_T \int_T \sum_{j,k\varepsilon K} [r^i(t,s)]_{jk}^2 dsdt \qquad (II.7)$$

which equals

$$\text{Tr}\left\{\int_T\int_T [r^i(t,s)]\cdot[r^i(t,s)]^T dsdt\right\} \qquad (II.8)$$

when Tr{ } denotes the trace of the given matrix. Then, applying Theorem 7.22 and Assumption A6.3.1 which say that

$$[r^i(t,s)] = \sum_{n=1}^{\infty} \lambda_n^i[\psi_n^i(t)]\cdot[\psi_n^i(s)]^T, \qquad (II.9)$$

(II.8) becomes

$$\text{Tr}\left\{\sum_{m,n=1}^{\infty} \lambda_m^i\lambda_n^i \int_T\int_T [\psi_m^i(t)]\cdot[\psi_m^i(s)]^T\cdot[\psi_n^i(s)]\cdot[\psi_n^i(t)]^T dsdt\right\} \qquad (II.10)$$

when the order of summation and integration is interchanged which is possible due to the uniform convergence of the series. Performing the s integration, (II.10) reduces to

$$\mathrm{Tr}\left\{\sum_{m=1}^{\infty} (\lambda_m^i)^2 \int_T [\psi_m^i(t)]\cdot[\psi_m^i(t)]^T dt\right\}$$

which equals

$$\sum_{j\varepsilon K} \sum_{m=1}^{\infty} (\lambda_m^i)^2 (a_{mj}) \qquad \text{(II.11)}$$

when we let

$$a_{mj} = \int_T [\psi_m^i(t)]_j^2 dt \quad . \qquad \text{(II.12)}$$

(Note that a_{mj} is positive.) Then, interchanging the order* of summation in (II.11), this sum reduces to

$$\sum_{m=1}^{\infty} (\lambda_m^i)^2 \qquad \text{(II.13)}$$

since

$$\sum_{j\varepsilon K} a_{mj} = \int_T [\psi_m^i(t)]^T \cdot [\psi_m^i(t)] dt = 1.$$

But (II.13), which equals (II.7), is also equal to the square of the Hilbert-Schmidt norm of R_i, i.e.,

$$N^2(R_i) = \sum_{m=1}^{\infty} ||R_i[\psi_m^i]||^2 = \sum_{m=1}^{\infty} (\lambda_m^i)^2 \qquad \text{(II.14)}$$

and is finite as a result of Assumption A6.3.1.

Q.E.D.

*See Ref. 59, p. 161.

APPENDIX III

A PROPERTY OF THE PROCESSES $[Z^i]$ AND $[Z^o]$

If $[Z^i]$ and $[Z^o]$ are the K-dimensional, zero mean, measurable Gaussian processes discussed following Theorem 7.25 and $\{[g_j]\}$ is any sequence of elements from $\overline{L}_2[T]$, then the sequence of random variables $\{\theta^i_j\}$ and $\{\theta^o_j\}$ defined for all $[f]\varepsilon\Omega$ by

$$\theta^i_j([f]) = \int_T [Z^i(t,[f])]^T[g_j(t)]dt$$

and

$$\theta^o_j([f]) = \int_T [Z^o(t,[f])]^T[g_j(t)]dt$$

are jointly Gaussian on $(\Omega,\mathcal{B},P_i)$ and $(\Omega,\mathcal{B},P_o)$, respectively.

Proof: We shall only prove the random variables $\theta^i_1,\theta^i_2,\ldots$ are jointly Gaussian since the proof that the random variables $\theta^o_1,\theta^o_2,\ldots$ are jointly Gaussian is identical. But, to prove $\theta^i_1,\theta^i_2,\ldots$ are jointly Gaussian, it suffices to prove $\theta^i_1,\theta^i_2,\ldots,\theta^i_n$ are jointly Gaussian when n is an arbitrary integer since the sequence $\{\theta^i_j\}$ can be reordered without loss of generality.

Let us first verify the fact that $\theta^i_j, j\varepsilon\{1,2,\ldots\}$, is a random variable. Since $[Z^i]$ and $[g_j]$ are measurable on ΩxT, $[Z^i]^T[g_j]$ is measurable on ΩxT and

$$\int_{\Omega xT} |[Z^i]^T[g_j]|d(P_i xt) \leq \sum_{j\varepsilon k} \int_{\Omega xT} |[Z^i]^T_k[g_j]_k|d(P_i xt) \qquad (III.1)$$

where $[Z^i]_k$ and $[g_j]_k$ are the kth components of $[Z^i]$ and $[g_j]$, respectively, and $P_i xt$ is the product measure on the produce space ΩxT. Then, due to Theorem 7.6 and Theorem 7.10, the right hand side of Eq. (III.1) is less than

$$\sum_{k\epsilon T} \left\{ \int_T [g_j]_k^2 dt \right\}^{1/2} \left\{ \int_T \int_\Omega [Z^i]_k^2 dP_i dt \right\}^{1/2} .$$

But

$$\int_T \int_\Omega [Z^i]_k^2 dP_i dt = \int_T [r^i(t,t)]_{kk} dt$$

$$\leq b^i T_f$$

where $[r^i(t,s)]_{kk}$ is the kkth element of the matrix covariance function of $[Z^i]$ and b^i is the bound on this function as discussed in Assumption A6.3.1. Therefore, the left hand side of (III.1) is finite and $|[Z^i]^T[g_j]|$ is integrable on ΩxT by Theorem 7.4. Finally,

$$\int_T [Z^i(t,[f])]^T[g_j(t)]dt$$

is (measurable) and integrable on Ω by Theorem 7.10.

Let us now show $\theta_1^i, \theta_2^i, \ldots, \theta_n^i$ are jointly Gaussian. Since the continuous real valued functions* are dense in $L_2[T]$, for every $[g_j]$, $j\epsilon\{1,2,\ldots,n\}$, we can find a sequence of continuous functions $\{[g_{jk}]\}_k$ for which

$$\lim_{k\to\infty} ||[g_{jk}] - [g_j]|| = 0. \qquad \text{(III.2)}$$

Moreover, for any k, the random variables $\{\theta_{jk}^i; j=1,2,\ldots,n\}$ defined by

$$\theta_{jk}^i = \int_T [Z^i(t,[f])]^T[g_{jk}(t)dt$$

are jointly Gaussian by Theorem 7.12. Also, we find

*See Ref. 39, p. 251.

$$\lim_{k\to\infty} E\{\theta^i_{jk}\theta^i_{jk}\} = \lim_{k\to\infty} E\{\theta^i_{jk}\theta^i_j\}$$

$$= E\{\theta^i_j\theta^i_j\}$$

when the order of integration is interchanged and Eq. (III.2) is applied. This implies, for each $j\epsilon\{1,2,\ldots,n\}$, the sequence $\{\theta^i_{jk}\}_k$ converges in the mean and, hence, in probability to θ_j. As a result, the characteristic function of the Gaussian random variables $\{\theta^i_{jk};j=1,2,\ldots,n\}$, call it Φ_k, converges* to the characteristic function of $\{\theta^i_j;j=1,2,\ldots,n\}$, call it Φ. The general form of Gaussian characteristic functions then shows that the limiting characteristic function is also Gaussian.

Q.E.D.

*See Ref. 28, p. 169.

APPENDIX IV

A CONVERGENCE THEOREM

Let $[Z^o]$ and $\{[Z_n^o]\}$ be as in Theorem 7.28 and define $H_i([Z^o]) = [W^o]$ and $H_i([Z_n^o]) = [W_n^o]$ with H_i also as in Theorem 7.28. Then,

$$\int_T [Z(t,[f])]^T[W^o(t,[f])]dt = \lim_{n\to\infty} \int_T [Z_n^o(t,[f])]^T[W_n^o(t,[f])]dt \quad m(\overline{P}_o)$$

where $m(\overline{P}_o)$ denotes convergence in measure with respect to $\overline{P}_o$.

Proof: With the help of Theorem 7.6 and Theorem 7.10, we can write

$$\begin{aligned}\int_\Omega \left| \int_T \{[Z^o]^T[W^o] - [Z_n^o]^T[W_n^o]\}dt \right| dP_o &\leq \int_T\int_\Omega |[Z_n^o]^T\{[W^o] - [W_n^o]\}|dP_o dt \\ &+ \int_T\int_\Omega |[W^o]^T\{[Z^o] - [Z_n^o]\}|dP_o dt \\ &\leq \varepsilon_n^1 \int_T\int_\Omega [Z_n^o]^T[Z_n^o]dP_o dt \\ &+ \varepsilon_n^2 \int_T\int_\Omega [W^o]^T[W^o]dP_o dt \quad \text{(IV.1)}\end{aligned}$$

where

$$\varepsilon_n^1 = \int_T\int_\Omega \{[W^o] - [W_n^o]\}^T\{[W^o] - [W_n^o]\}dP_o dt$$

and

$$\varepsilon_n^2 = \int_T\int_\Omega \{[Z^o] - [Z^o{}_n]\}^T\{[Z^o] - [Z_n^o]\}dP_o dt.$$

Then, from Eqs. (7.118) and (7.120), $[Z^o] = \underset{n\to\infty}{\text{l.i.m.}}[Z_n^o](t\times P_o)$ and $[W^o] = \underset{n\to\infty}{\text{l.i.m.}}[W_n^o](t\times P_o)$ so that ε_n^1 and ε_n^2 approach zero as $n\to\infty$. Furthermore,

$$\begin{aligned}\int_T\int_\Omega [W^o]^T[W^o]dP_o dt &= \int_\Omega\int_T [W^o]^T[W^o]dt\,dP_o \\ &\leq \|H_i\|^2 \int_\Omega\int_T [Z^o]^T[Z^o]dt\,dP_o\end{aligned}$$

and is, therefore, finite. Also, from the definition $[Z_n^o]$, we obtain

$$\int_T \int_\Omega [Z_n^o]^T [Z_n^o] dP_o dt = \sum_{j=1}^{n} \{(R_o^{1/2}[\phi_j], R_o^{1/2}[\phi_j]) + (R_o^{1/2}[\overline{\phi}_j], R_o^{1/2}[\overline{\phi}_j])\} \tag{IV.2}$$

with $\{[\phi_j]\}$ and $\{[\overline{\phi}_j]\}$ as in Lemma 7.9. But, as shown in Lemma 7.9, the summation on the right of (IV.2) converges to Tr $\{\int_T [r^o(t,t)]dt\}$ which is finite by Assumption A6.3.1. It can now be seen that the right hand side of Eq. (IV.1) approaches zero as $n \to \infty$.

Now let

$$A_n = \{[f]: |\int_T [Z^o]^T [W^o] - [Z_n^o]^T [W_n^o] dt| > \varepsilon\}$$

where ϵ is any real number greater than zero. Then, if n_o is chosen such that, for $n \geq n_o$, the left hand side of (IV.1) is less than ϵ^2, it follows that

$$P_o\{A_n\} < \varepsilon$$

when $n \geq n_o$ and

$$\int_T [Z^o(t,[f])]^T [W^o(t,[f])] dt = \lim_{n\to\infty} \int_T [Z_n^o(t,[f])]^T [W_n^o(t,[f])] dt \quad m(P_o)$$

which implies

$$\int_T [Z^o(t,[f])]^T [W^o(t,[f])] dt = \lim_{n\to\infty} \int_T [Z_n^o(t,[f])]^T [W_n^o(t,[f])] dt \quad m(\overline{P}_o) .$$

Q.E.D.

APPENDIX V

COMPARISON OF "MINIMUM PROBABILITY OF ERROR" AND "MAXIMUM A POSTERIORI PROBABILITY" PARTITIONS

In this appendix we shall discuss the relationship that exists between the "Minimum Probability of Error" and the "Maximum A Posteriori Probability" partitions of the observation space. In particular, we shall find, defining the "A Posteriori Probability" in a special way, these two partitions are identical. It should be pointed out, however, this discussion is not intended to be a rigorous proof of this relationship but rather an addition argument illustrating the "optimality" of this partition.

In Theorem 7.24, it was found that the partition of the observation space that minimizes the "Probability of Error" classifies the observation [y] in the subset Λ_i if i is the least index for which

$$a_i \frac{d\bar{P}_i}{d\bar{P}_o}([y]) \geq a_j \frac{d\bar{P}_i}{d\bar{P}_o}([y]) \qquad (V.1)$$

for all $j \neq i$. (The set for which $a_i \frac{d\bar{P}_i}{d\bar{P}_o}([y])$ achieves $\max_i a_i \frac{d\bar{P}_i}{d\bar{P}_o}([y])$ for more than one index i has zero probability as shown in Theorem 7.25 and, therefore, need not be considered.) The constants $\{a_i\}$ represent the a priori probabilities of the various angles of arrival.

The question now arises, "To what are we referring when we specify the a posteriori probability of the angle of arrival given a received observation [y]?" To answer this question, let us first restrict ourselves to the discrete case where the noise process [N(t)] and the possible signal processes $\{[S^i(t)]\}$ take on only a finite set of possible values. Then, if [y] is an observation of one of the processes

$$[Y^i(t)] = [N(t)] + [S^i(t)] , \tag{V.2}$$

Bayes* rules gives us

$$Pr(i/[y]) = \frac{a_i Pr_i([y])}{\sum_i a_i Pr_i([y])} \tag{V.3}$$

where $\{a_i\}$ represents the a priori probabilities of the measures on the processes $\{[S^i(t)]\}$, $Pr_i([y])$ is the probability of [y] given $[S^i(t)]$ is the transmitted process and $Pr(i/[y])$ is the conditional or "a posteriori" probability of i being the index of the transmitted process given [y] is the received observation. The "Maximum A Posterior Probability" of the index is then the least index i for which

$$Pr(i/[y]) \geq Pr(j/[y]) \tag{V.4}$$

for all $j \neq i$. (In the general case where the noise and signal processes are not limited to a discrete set, the measure of the set for which more than one index can achieve this value has zero measure.) Moreover, if we divide the numerator and denominator of (V.3) by $Pr_o([y])$ which denotes the probability of [y] given noise alone is present, we find the index i which satisfies (V.4) also satisfies

$$a_i \frac{Pr_i([y])}{Pr_o([y])} \geq a_j \frac{Pr_j([y])}{Pr_o([y])} \tag{V.5}$$

for all $j \neq i$ since $Pr_o([y])$ and $\sum_i a_i Pr_i([y])$ are independent of the index i.

*See Ref. 43, p. 18.

Suppose, now, the noise and signal processes are no longer limited to a finite set of possible values. In this case, we see that both $Pr_i([y])$ and $Pr_o([y])$ go to zero so, to generalize (V.5), it is reasonable to consider some type of limit operation of the form

$$\lim_{\varepsilon\to\infty} \frac{Pr_i([y]\varepsilon A_\varepsilon)}{Pr_o([y]\varepsilon A_\varepsilon)} \qquad (V.6)$$

where $\{A_\varepsilon\}$ is a monotone decreasing sequence of measurable sets containing [y] which, in some sense,approaches [y]. To obtain an expression for this limit, let us denote by $\bar{P}_o$ and $\bar{P}_i$ the measures induced on the observation space by Pr_o and Pr_i, respectively. Then, if $\bar{P}_o$ and $\bar{P}_i$ are equivalent, due to Theorem 7.14 we can write

$$\bar{P}_i(A_\varepsilon) = \int_{A_\varepsilon} \frac{d\bar{P}_i}{d\bar{P}_o}\, d\bar{P}_o \qquad (V.7)$$

where $d\bar{P}_i/d\bar{P}_o$ is the Radon-Nikodym derivative of $\bar{P}_i$ with respect to $\bar{P}_o$. It can now be seen, if $d\bar{P}_i/d\bar{P}_o$ approaches a constant value on A_ε as $\varepsilon\to 0$, the limit expression of (V.6) becomes

$$\lim_{\varepsilon\to 0} \frac{\bar{P}_i([y]\varepsilon A_\varepsilon)}{\bar{P}_o([y]\varepsilon A_\varepsilon)} = \frac{d\bar{P}_i}{d\bar{P}_o} . \qquad (V.8)$$

But, according to Theorem 7.28, $\bar{P}_i$ and $\bar{P}_o$ are equivalent if there exists a bounded self-adjoint operator H_i which satisfies Eq. (7.112) and, if such an operator exists, then

$$\frac{d\bar{P}_i}{d\bar{P}_o} = D_i^{-1}\exp\{\tfrac{1}{2}(H_i([y]),[y])\} . \qquad (V.9)$$

Thus, if such an H_i exists and if we define

$$A_\varepsilon = \{[f]\varepsilon\Omega : ||[f] - [y]|| < \varepsilon\} , \qquad (V.10)$$

we see that, indeed, $d\bar{P}_i/d\bar{P}_o$ approaches a constant value on A_ε as $\varepsilon \to 0$ since H_i is a bounded operator.

Returning now to (V.5) and replacing $Pr_i([y])/Pr_o([y])$ by the limit expression of (V.6) which equals $d\bar{P}_i/d\bar{P}_o$ due to Eq. (V.8), we see that a reasonable "Maximum A Posteriori Probability" partition of the observation space places [y] in Λ_i if

$$a_i \frac{d\bar{P}_i}{d\bar{P}_o} \geq a_j \frac{d\bar{P}_j}{d\bar{P}_o}$$

for all $j \neq i$. But this is exactly the "Minimum Probability of Error" partition as shown in Eq. (V.1).

This completes the discussion of this appendix which, as stated in the opening paragraph, has been somewhat less than rigorous. Hopefully, however, it has given some additional insight into the degree of "optimality" enjoyed by this particular partition.

REFERENCES

1. H. Hertz, Electric Waves, Macmillan Co., 1893.

2. G. Marconi, "Wireless Telegraphy," Journal of the Institute of Electrical Engineers, 28 (1899).

3. J. Zenneck, Wireless Telegraphy, McGraw-Hill Book Co., 1915.

4. S. G. Brown, "Directed Wireless Telegraphy," Electrician, 57 (1906).

5. R. Keen, Wireless Direction Finding, Iliffe and Sons Limited (1938).

6. E. Bellini and A. Tosi, "The Range and Advantages of Directive Aerials and the Bellini-Tosi Radiogoniometer," La Lumière Electrique, 5 (1909).

7. R. S. Smith, "Electronic Observers for Radio Direction Finding," RRL Publ. No. 273, Radiolocation Research Laboratory, Univ. of Ill., Urbana, Ill., Feb. 1965.

8. W. J. Lindsay and D. S. Heim, "Design for Spinning Goniometer Automatic Direction Finding," J. of Res. of the National Bureau of Standards—D. Radio Propagation, 65D, No. 3, 237-243(May-June 1961).

9. C. E. Lindahl and B. F. Barton, "Analysis of a Detection Technique for a Spinning-Goniometer Direction-Finding System," IEEE Trans. on Aerospace and Navigational Electronics, 124-127 (June 1963).

10. J. L. Allen, "Array Antennas: New Applications for an Old Technique," IEEE Spectrum, 115-130 (Nov. 1964).

11. R. C. Hansen, editor of Microwave Scanning Antennas, Vols. 2-3, Academic Press, 1966.

12. S. A. Schelkunoff and H. T. Friis, Antennas, Theory and Practice, John Wiley and Sons, Inc., 1952.

13. K. J. Astrom, R. W. Koepcke, and F. Tung, "On the Control of Linear Discrete Dynamic Systems with Quadratic Loss," Res. Report RJ-222, IBM, San Jose Research Lab., San Jose, Calif., Sept. 1962.

14. A. A. Fel'dbaum, "The Theory of Dual Control, I-IV," Avtomatika i Telemekhanika, 21, Nos. 9, 11 (1960); 22, Nos. 1,2 (1961).

15. E. P. Maslov, "Markov Object Parameter Estimation," Avtomatika i Telemekhanika, 25, No. 1 (1964).

16. V. P. Zhivoglyadov, "Optimal Dual Control of Plants with Pure Delay," Avtomatika i Telemekhanika, 25, No. 1 (1964).

REFERENCES (Continued)

17. R. L. Stratonovich, "Certain Extremal Problems in Mathematical Statistics and Conditional Markov Processes," Teoriya Veroyatn i Primeneniya, t. 7, v. 2, 1962.

18. R. E. Mortensen, "Optimal Control of Continuous-Time Stochastic Systems," Ph.D. Thesis, Univ. of Calif., Berkeley, 1965.

19. W. M. Wonham, "Stochastic Problems in Optimal Control," 1963 IEEE International Convention Record, Part 2, pp. 114-124.

20. A. M. Legendre, Nouvelles methods pour la dètermination des orbites des comètes, Paris, 1806.

21. K. F. Gauss, Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections, Dover Pub. Inc., 1963 (reprint).

22. K. Pearson, "Contributions to the Mathematical Theory of Evolution," Philos. Trans. Roy. Soc. London, 185 (1894).

23. R. A. Fisher, "On an Absolute Criterion for Filtering Frequency Curves," Mess. of Math., 41 (1912).

24. R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," Philos. Trans. Roy. Soc. London, 222 (1922).

25. R. A. Fisher, "Theory of Statistical Estimation," Proc. Cambridge Philos. Soc., 22 (1925).

26. N. Wiener, The Extrapolation, Interpolation, and Smoothing of Stationary Time Series, John Wiley and Sons, Inc., 1949.

27. A. N. Kolmogorov, "Interpolation and Extrapolation von Stationarew Zufalligen Folgen," Bull. Acad. Sci. (USSR), Ser. Math., 5 (1941).

28. M. Loève, Probability Theory, D. Van Nostrand Co., Inc., 1963.

29. J. L. Doob, Stochastic Processes, John Wiley and Sons, Inc., 1953.

30. P. K. Halmos, Measure Theory, D. Van Nostrand Co., Inc., 1950.

31. F. W. Lehan and R. J. Parks, "Optimum Demodulation," 1953 IRE National Convention Record, pp. 101-103.

32. D. C. Youla, "The Use of Maximum Likelihood in Estimating Continuously Modulated Intelligence which has been Corrupted by Noise," IRE Trans. on Information Theory, IT-3, 90-105 (March 1954).

33. D. Slepian, "Estimation of Signal Parameters in the Presence of Noise," IRE Trans. on Information Theory, IT-3, 68-89 (March 1964).

REFERENCES (Continued)

34. J. B. Thomas and E. Wong, "On the Statistical Theory of Optimum Demodulation," IRE Trans. on Information Theory, IT-7, 420-425 (Sept. 1960).

35. H. L. Van Trees, Unpublished notes for an Electrical Engineering course offered by the Mass. Inst. of Tech., Cambridge, Mass.

36. T. T. Kadota, "Optimum Reception of Binary Gaussian Signals," The Bell System Technical Journal, 43, No. 6 (Nov. 1964).

37. T. T. Kadota, "Optimum Reception of Binary Sure and Gaussian Signals," The Bell System Technical Journal, 44, No. 8 (Oct. 1965).

38. T. T. Kadota, "Optimum Reception of M-ary Gaussian Signal in Gaussian Noise," The Bell System Technical Journal, 44, No. 9 (Nov. 1965).

39. T. T. Kadota, "Generalized Maximum Likelihood Test and Minimum Error Probability," IEEE Trans. on Information Theory, IT-12, 65-67 (Jan. 1966).

40. E. J. Kelley, I. S. Reed, and W. L. Root, "The Detection of Radar Echoes in Noise, I," J. Soc. Indust. Appl. Math., 8, No.2, 309-341 (June 1960).

41. E. J. Kelley, I. S. Reed, and W. L. Root, "The Detection of Radar Echoes in Noise, II," J. Soc. Indust. Appl. Math., 8, No. 3, 481-507 (Sept. 1960).

42. R. E. Bellman, Dynamic Programming, Princeton Univ. Press, 1957.

43. W. B. Davenport and W. L. Root, Random Signals and Noise, McGraw-Hill Book, Co., Inc., 1958.

44. A. E. Taylor, Functional Analysis, John Wiley and Sons, Inc., 1958.

45. T. S. Pitcher, "An Integral Expression for the Log Likelihood Ratio of Two Gaussian Processes," J. SIAM, Appl. Math., 14, No. 2, 228-233 (March 1966).

46. B. H. Bharucha, "Minimum Error Probability Detection for Infinitely Many Hypotheses," IEEE Trans. on Information Theory, IT-12, 65 (Jan. 1966).

47. H. R. Pitt, Integration Measure and Probability, Oliver and Boyd, 1963.

48. C. Goffman and G. Pedrick, First Course in Functional Analysis, Prentice-Hall, Inc., 1965.

REFERENCES (Concluded)

49. N. Dunford and J. T. Schwartz, Linear Operators, Vol. 2, Interscience, New York, 1958.

50. J. Neveu, Mathematical Foundations of the Calculus of Probability, Holden-Day, Inc., 1965.

51. W. L. Root, "Singular Gaussian Measures in Detection Theory," Proc. of the Sym. on Time Series Analysis, John Wiley and Sons, Inc., 1963.

52. W. Schmeidler, Linear Operators in Hilbert Space, Academic Press, 1965.

53. A. C. Zaanen, Linear Analysis, North-Holland, Pub. Co., Amsterdam, 1953.

54. R. C. Buck, Studies in Modern Analysis, Vol. 1, Prentice-Hall, Inc., 1962.

55. F. Riesz and B. Sz-Nagy, Functional Analysis, Frederick Ungar Pub. Co., New York, 1955.

56. E. J. Kelley and W. L. Root, "A Representation of Vector-Valued Random Processes," J. of Math. and Phys., 39, No. 3, 211 (Oct. 1960).

57. H. L. Royden, Real Analysis, Macmillian Co., 1963.

58. T. M. Apostol, Mathematical Analysis, Addison-Wesley Pub. Co., Inc., 1957.

59. W. Rudin, Principles of Mathematical Analysis, McGraw-Hill Book Co., 1964.

60. M F. Gardner and J. L. Barnes, Transients in Linear Systems, John Wiley and Sons, Inc., 1958.

61. D. Middleton, Statistical Communication Theory, McGraw-Hill Book Co., Inc., 1960.

62. C.R.C. Standard Mathematical Tables, Chemical Rubber Pub. Co., 1955.

63. A. A. Ksienski, "Very High Resolution Techniques," Report No. 10, Contract No. DA-36-(039)-SC-90772, Hughes Aircraft Co., Culver City, Calif., Dec. 1964.

64. K. S. Miller, Multidimensional Gaussian Distributions, John Wiley and Sons, Inc., 1964.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Systems Engineering Laboratory
The University of Michigan
Ann Arbor, Michigan

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
Optimal Space-Time Signal Processing and Parameter Estimation with Emphasis on Its Application to Direction Finding

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
SEL Technical Report No. 16

5. AUTHOR(S) (Last name, first name, initial)
Mueller, Donald G.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1967	225	64

8a. CONTRACT OR GRANT NO.
DA 28-043-AMC-01870(E)

b. PROJECT NO.
1P021101 A042

c. 01

d. 02

9a. ORIGINATOR'S REPORT NUMBER(S)
ECOM-01870-14-T

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
7695-14

10. AVAILABILITY/LIMITATION NOTICES
Each transmittal of this document outside the Department of Defense must have prior approval of CG, U. S. Army Electronics Command, Fort Monmouth, N.J. Attn: AMSEL-WL-S

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
U. S. Army Electronics Command
Fort Monmouth, New Jersey 07703
Attn: AMSEL-WL-S

13. ABSTRACT

The purpose of this study is to develop techniques for processing signals received by a special array of K-omnidirectional antennas so as to produce optimal direction-of-arrival estimates. Such estimates are useful for navigational purposes in which case the signals are electromagnetic in nature as well as for seismic investigations where the signals are mechanical vibrations transmitted through the earth. Another possible application is an underwater direction finding system which utilizes an array of hydrophones for its receiving antennas.

Two distinct approaches to the problem are pursued which apply differing mathematical disciplines. The first approach models the array as an element of a phased array direction finding system and attempts to apply "Stochastic Optimal Control Theory" to produce optimal control laws for directing the pointing angle and specifying the beam width of the array. Optimal filtering of the signals is also considered. The controls and filter which result in "Minimum Error Variance" are considered optimal. The second approach applies "Estimation Theory" and considers the array only as an information gathering device whose signals are to be processed directly without the "prefiltering" present in phased arrays. The optimal criterion of this approach required the direction-of-arrival estimate to result in a "Minimum Probability of Error" on, what is shown to be equivalent, to have "Maximum a Posteriori Probability."

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Optimal control laws Minimum error variance Estimation theory Functional analysis Radon-Nikodym Dynamic programming Open-loop analysis						

INSTRUCTIONS

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

8*a*. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

Security Classification

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D		
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)		
1. ORIGINATING ACTIVITY (Corporate author) Systems Engineering Laboratory The University of Michigan Ann Arbor, Michigan	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP	
3. REPORT TITLE Optimal Space-Time Signal Processing and Parameter Estimation with Emphasis on Its Application to Direction Finding		
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) SEL Technical Report No. 16		
5. AUTHOR(S) (Last name, first name, initial) Mueller, Donald G.		
6. REPORT DATE October 1967	7a. TOTAL NO. OF PAGES 225	7b. NO. OF REFS 64
8a. CONTRACT OR GRANT NO. DA 28-043-AMC-01870(E) b. PROJECT NO. 1P021101 A042 c. 01 d. 02	9a. ORIGINATOR'S REPORT NUMBER(S) ECOM-01870-14-T 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) 7695-14	
10. AVAILABILITY/LIMITATION NOTICES Each transmittal of this document outside the Department of Defense must have prior approval of CG, U. S. Army Electronics Command, Fort Monmouth, N.J. Attn: AMSEL-WL-S		
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY U. S. Army Electronics Command Fort Monmouth, New Jersey 07703 Attn: AMSEL-WL-S	

13. ABSTRACT

The purpose of this study is to develop techniques for processing signals received by a special array of K-omnidirectional antennas so as to produce optimal direction-of-arrival estimates. Such estimates are useful for navigational purposes in which case the signals are electromagnetic in nature as well as for seismic investigations where the signals are mechanical vibrations transmitted through the earth. Another possible application is an underwater direction finding system which utilizes an array of hydrophones for its receiving antennas.

Two distinct approaches to the problem are pursued which apply differing mathematical disciplines. The first approach models the array as an element of a phased array direction finding system and attempts to apply "Stochastic Optimal Control Theory" to produce optimal control laws for directing the pointing angle and specifying the beam width of the array. Optimal filtering of the signals is also considered. The controls and filter which result in "Minimum Error Variance" are considered optimal. The second approach applies "Estimation Theory" and considers the array only as an information gathering device whose signals are to be processed directly without the "prefiltering" present in phased arrays. The optimal criterion of this approach required the direction-of-arrival estimate to result in a "Minimum Probability of Error" or, what is shown to be equivalent, to have "Maximum a Posteriori Probability."

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Optimal control laws						
Minimum error variance						
Estimation theory						
Functional analysis						
Radon-Nikodym						
Dynamic programming						
Open-loop analysis						