71-4697

NUSPL, Peter Paul, 1940-
MODELS FOR NONSTATIONARY STOCHASTIC PROCESSES WITH APPLICATIONS TO UNDERWATER ACOUSTICS.

The University of Michigan, Ph.D., 1970
Engineering, electrical

University Microfilms, Inc., Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

MODELS FOR NONSTATIONARY STOCHASTIC PROCESSES WITH APPLICATIONS TO UNDERWATER ACOUSTICS

by
Peter Paul Nuspl

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Computer, Information & Control Engineering)
in The University of Michigan
1970

Doctoral Committee:

Professor Theodore G. Birdsall, Chairman
Professor Louis F. Kazda
Professor Lawrence L. Rauch
Assistant Professor Louis P. Solomon
Doctor Marlin P. Ristenbatt, Cooley Electronics Laboratory

ABSTRACT

MODELS FOR NONSTATIONARY STOCHASTIC PROCESSES WITH APPLICATIONS TO UNDERWATER ACOUSTICS

by Peter Paul Nuspl

Chairman: Theodore G. Birdsall

Some models for nonstationary stochastic processes are presented and applied to real data processes. The nonstationary processes studied are those with time-varying autocorrelation matrices. Since probability distributions are complete statistical descriptions of a random variable, the fundamental problems are to find conditional distributions for the process x given observations y in a time interval.

The Markov condition and the Gaussian distribution are assumed for the process x. It is called Gauss-Markov and the model is a set of linear, discrete-time difference equations, which are conveniently expressed in state-space notation. Also presuming knowledge of the model parameters, the solutions have been stated as the Kalman-Bucy equations. The conditional mean and conditional covariance are computed recursively based on past observations.

The model parameters consist of the state-transition matrix, the driving noise covariance matrix and the observation noise

covariance matrix. They are often not known, so a set of consistent estimators is given and then modified to become "tracking" estimators.

A class of linear filters and digital processors is presented which have simple computer implementation. The data-acquisition consists of simultaneous processing of hydrophone outputs to produce signal power and noise power measurements of underwater acoustic receptions. The study shows that the signal power process is apparently second-order stochastic and that the second-order model is adequate as a smoothing and prediction filter. The noise process has shipping and biological noise which may be of interest in themselves, but which make effective modeling difficult; by selective elimination of such observations a study of ambient ocean noise is possible. We conclude that the models are useful in studying these nonstationary stochastic processes.

To my wife, Maria

Tina and Tony

ACKNOWLEDGEMENTS

The author wishes to thank the members of his doctoral committee for their guidance and suggestions during the course of this study. Special thanks are extended to the Chairman, Professor T. G. Birdsall, for his generous availability for discussions and for his critical review of all phases of this research. Thanks are due to the author's colleagues at the University of Michigan, in particular to K. Metzger and G. N. Cederquist for their development of the Cooley Programming System (CPS).

The research reported in this dissertation was supported by the Office of Naval Research, Department of the Navy, under Contract No. Nonr-1224(36), titled "Acoustic Signal Processing."

The author is grateful to Mrs. Ann Fulmer for her excellent typing of the dissertation and to Mrs. Betty Willsher for her careful preparation of many illustrations.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF ILLUSTRATIONS	viii
LIST OF SYMBOLS	xi
CHAPTER 1: INTRODUCTION	1
1.1 Nature of the Problem	2
1.1.1 Stochastic Processes	2
1.1.2 Forms of Stationarity and Nonstationarity	3
1.1.3 Models	4
1.1.4 Objectives	5
1.1.5 Acoustic Signal and Noise Processes	6
1.2 Historical Background	7
1.3 Plan of the Thesis	8
1.4 Notation and Format	9
CHAPTER 2: STOCHASTIC PROCESSES AND MODEL EQUATIONS	11
2.1 Introduction	12
2.2 Stochastic Processes	12
2.2.1 Markov Property	14
2.2.2 Gaussian Process	15
2.2.3 Ergodicity	16
2.3 State-Variables and Matrix Notation	19
2.3.1 Linear Vector Equation	19
2.3.2 State Transition Matrix	22
2.3.3 An Example	23
2.4 Model Equations	26
CHAPTER 3: KALMAN-BUCY FILTER	28
3.1 Introduction	29
3.2 Kalman Filter	30
3.2.1 Bayes Estimation	30
3.2.2 Recursive Equations for $\hat{x}_1$, $\hat{x}_2$	31
3.2.3 The Solutions	36

TABLE OF CONTENTS (Cont.)

	Page
3.3 Continuous-Time Filter	39
3.4 Initial Conditions	40
3.5 Performance Indices	43
CHAPTER 4: MODEL PARAMETER ESTIMATION	45
4.1 Introduction	46
4.2 Consistent Estimates for the Stationary Model	48
4.3 Tracking Estimates for the Nonstationary Model	52
4.4 Filtering of Field Data	57
4.5 Initial Conditions on Statistics	60
4.6 Summary of Recursive Equations of Model C	62
CHAPTER 5: A CLASS OF LINEAR FILTERS	65
5.1 Introduction	66
5.2 A Standard Form of a Class of Analog Filters	66
5.2.1 Impulse Response	66
5.2.2 Transfer Function	69
5.2.3 The τ-Diagram	70
5.2.4 Special Cases: $r = \pm 1$, $r = \pm j$	72
5.2.5 General Case; Complex r	79
5.3 A Standard Form of Digital Processors	84
5.3.1 Sampling	84
5.3.2 Equivalent Digital Processor	85
5.3.3 MPR Block Diagram	87
5.3.4 Examples	88
5.4 Combined Operation	89
5.5 Time-Domain Viewpoint	91
CHAPTER 6: DATA PROCESSING	93
6.1 Introduction	94
6.2 Computer Program Flow Diagram	97
6.3 Monte-Carlo Tests	
6.3.1 Gauss-Markov Sample Functions	102
6.3.2 Testing a Stationary Model	102

TABLE OF CONTENTS (Cont.)

	Page
6.4 Source of Field Data	104
6.4.1 Underwater Acoustics	104
6.4.2 Dual Comb Data Processing	104
6.4.3 Processes T, N	107
6.5 Experimental Procedure	108
CHAPTER 7: MODEL APPLIED TO THE T PROCESS	112
7.1 Introduction	113
7.2 T Process Sample Functions	113
7.3 One-State Model, n = 1	113
7.3.1 Coefficients in the Estimators	118
7.3.2 Initial Conditions	124
7.3.3 Filtered Process	124
7.3.4 Performance Results	126
7.4 Modified Two-State Model, n = 2	130
7.4.1 Minimum Mean-Square Fit to Obtain Vector y	131
7.4.2 Coefficients in the Estimators	134
7.4.3 Filtered Process	135
7.4.4 Performance Results	135
7.5 Comments on the T Process	142
CHAPTER 8: MODEL APPLIED TO THE N PROCESS	145
8.1 Introduction	146
8.2 N Process Sample Functions	146
8.3 One-State Model, n = 1	149
8.3.1 Coefficients in the Estimators	149
8.3.2 Initial Conditions	151
8.3.3 Filtered Process	152
8.3.4 Performance Results	152
8.4 Comments on the N Process	152
CHAPTER 9: CONCLUSION	159
9.1 Summary of Results	160
9.1.1 Theory and Methods	160
9.1.2 Data Processing	162
9.1.3 Achieved Objectives	164
9.1.4 Contributions	165

TABLE OF CONTENTS (Cont.)

	Page
9.2 Future Work	166
APPENDIX A: A TECHNICAL GLOSSARY	169
APPENDIX B: EXPECTATIONS OF PRODUCTS WITH Φ_k, Φ_{k-1} AS FACTORS	181
APPENDIX C: EXPECTATIONS OF QUARTIC TERMS FOR THE VECTOR CASE	186
APPENDIX D: EXPRESSIONS FOR THE OBSERVATION VECTOR	189
APPENDIX E: THE STEADY-STATE COVARIANCE Λ	191
REFERENCES	193

LIST OF ILLUSTRATIONS

Figure	Title	Page
4.1	Model C	60
4.2	a) Estimation of Φ_k, Q_k, R_k	64
	b) Kalman filter block diagram, H = I, n x n	64
5.1	A delay line of T units	68
5.2	A tapped delay line filter	68
5.3	τ-diagrams	71
5.4	Amplitude spectra $\|H(f)\|$ for r = 1	74
5.5	a) Power spectra $\|H(f)\|^2$ for r = .90	82
	b) Power spectra $\|H(f)\|^2$ for r = 1.0	82
	c) Power spectra $\|H(f)\|^2$ for r = .5	83
5.6	$\|H(0)\|^2$ versus r	83
5.7	Sampled filter	85
5.8	Sampled circulating average	86
5.9	Simple circulating average	86
5.10	P-position circulating average	87
5.11	MPR processor	88
5.12	The complex block average	89
5.13	A cascade of block averages	90
5.14	A cascade of circulating average and block average	90
6.1	Preprocessing of data	96

LIST OF ILLUSTRATIONS (Cont.)

Figure	Title	Page
6.2	Program flow chart	99
6.3	Flow chart for ICIN	100
6.4	Flow chart for OUTPUT	101
6.5	Gauss-Markov sequence sample function	103
6.6	Data-acquisition processing	106
7.1	a) T process sample function b) T process sample function c) T process sample function	114 115 116
7.2	T process sample function, one of every 5 data points	117
7.3	T process: estimate $\hat{x}_{k\mid k}$ for n = 1	125
7.4	T process: forecast $\hat{x}_{k+j\mid k}$ for n = 1	127
7.5	Estimates $\hat{\Phi}_k$, $\hat{Q}_k$, $\hat{R}_k$ for n = 1	128
7.6	P_k, $J0_k$ for n = 1	129
7.7	Line fit for vector y_k	133
7.8	T process: estimate $\hat{x}_{k\mid k}$ for n = 2	136
7.9	T process: estimate $\hat{x}_{k\mid k}$ for n = 2	137
7.10	T process: estimate $\hat{x}_{k\mid k}$ for n = 2	138
7.11	Estimates $\hat{\Phi}_k$, $\hat{Q}_k$, $\hat{R}_k$ for n = 2	139
7.12	Error covariance P_k and performance index $J0_k$	141
7.13	T process: forecasts $\hat{x}_{k+j\mid k}$ for n = 2	143

LIST OF ILLUSTRATIONS (Cont.)

Figure	Title	Page
8.1	N process sample function	147
8.2	Output of the nonlinear processor	150
8.3	N process: estimate $\hat{x}_{k\|k}$ for $n = 1$, time-invariant model	153
8.4	N process: estimate $\hat{x}_{k\|k}$ for $n = 1$, time-varying model	154
8.5	Model parameters $\hat{\Phi}_k$, $\hat{Q}_k$, $\hat{R}_k$	155
8.6	Model parameters $\hat{\Phi}_k$, $\hat{Q}_k$ with $R_k = 1.0$	156
8.7	N process: forecast $\hat{x}_{k+j\|k}$ for $n = 1$	157
A.1	A limiter	174

LIST OF SYMBOLS

Ω	abstract space
ω	element in Ω
ϵ	indicates "member in"
$r(\cdot)$, $x(\cdot)$	random variables
$\mathcal{A}$	indexing set
s, t, t_k	time $t \epsilon \mathcal{A}$
$r(\cdot, t)$, $r(t)$	stochastic processes
$x(\cdot, t)$, $x(t)$, x_k	stochastic processes
$y(\cdot, t)$, $y(t)$, y_k	stochastic processes
E	expectation
$R(\tau)$, $R(\tau, t)$	autocorrelation function
$\mathcal{F}$	σ-field of sets in Ω
A, B	sets in Ω (in context)
$P(\cdot)$	probability measure
$A \mid B$	A given condition B holds
$p(\cdot)$	probability density function
k	discrete time
x_k, y_k	column vector random variable at time k
Λ_x, Λ_k	covariance matrix of x
n	dimension of x_k
m	dimension of y_k

LIST OF SYMBOLS (Cont.)

α, β	density parameters
$\equiv$	is identically equal to
$\triangleq$	is defined to be
T	a transformation, in Theorem 2.1
T	the process studied in Chapter 7
$\mathcal{I}$	σ-field of invariant sets
μ_a	time average over (-a, a)
$R_a(\lambda)$	time average over (-a, a)
F(t)	coefficient matrix
w(t)	vector driving function
$\Phi(t, t_0)$	state transition matrix
I	identity matrix
$\Phi_k \equiv$	$\Phi(k, k-1)$
u_k	process equation driving term
Q_k	covariance matrix of u_k
H_k	observation matrix
v_k	observation equation driving term
R_k	covariance matrix of v_k
$\delta_{pk} \triangleq 1$ $p = k$	(Kronecker delta)
0 otherwise	
R, R_{ms}	risk defined in Chapter 3
$\hat{x}(\cdot)$	estimate of x

LIST OF SYMBOLS (Cont.)

$\hat{x}_{k\,|\,k}$	optimal estimate given past observations
$\tilde{x}_{k\,|\,k}$	error in the estimate
P_k	prediction error covariance matrix
$\overline{P}_k$	filter error covariance matrix
A_k	optimal gain in the Kalman filter
$\hat{x}_{k+j\,|\,k}$	prediction of x_{k+j} given observations up to time k
JN_k	performance index, Eq. 3.54
$\hat{\Phi}_k$, $\hat{Q}_k$, $\hat{R}_k$	estimators for model parameters
c_{k-j}, d_{k-j}, e_{k-j}	coefficients in statistics
r_ϕ, r_0, r_1	radices of coefficients in statistics
$\rho(\Phi)$	spectral radius of Φ
A_0, $A_{0,k}$, A_1, $A_{1,k}$	statistics in Model A
B_0, $B_{0,k}$, B_1, $B_{1,k}$	statistics in Model B
S^{-1}	inverse of matrix S
S^+	pseudo-inverse of matrix S
$\mathscr{L}$	the L-sphere about $\overline{r}$
$h(\tau)$	impulse response of a filter

LIST OF SYMBOLS (Cont.)

$H(f)$	Fourier transform of $h(\tau)$
T, Δ	discrete time interval lengths
M	number of observations averaged
P	number of positions in MPR filter
r	radix of coefficients in MPR filter
BAV	block average
CIRAV	circulating average
MPR	multi-positional recirculator
$\delta_\epsilon(t)$	a sequence of delta functions, ϵ spaced
N	the process studied in Chapter 8
σ^2	scalar variance
χ^2	chi-squared statistic

CHAPTER 1

INTRODUCTION

Synopsis

1.1 Nature of the Problem

Stochastic processes

Forms of stationarity and nonstationarity

Models

Objectives

Acoustic signal and noise processes

1.2 Historical Background

1.3 Plan of the Thesis

Introduction

Model development

Data processing

Conclusion

1.4 Notation and Format

1.1 Nature of the Problem

This thesis is a study of models for nonstationary stochastic processes. The key words in the title are introduced and the nature of the research problem will be described in this section. The theory and the methods of dealing with this problem, and the data processing and results are the main part of this study.

1.1.1 Stochastic Processes. Beginning with the fundamental concepts in probability theory, some common terms in stochastic process theory are defined. A random variable* is a measurable function defined on the abstract space

$$\Omega = \{\omega\} \ .$$

A notation often used is $r(\cdot)$, and $r(\omega)$ is a particular value called a sample value. Where there is no possibility of confusion, the notation is often abbreviated to r for a random variable and the underlying space and probability structure are understood.

A stochastic process is a family of random variables indexed by a set $\mathcal{A}$. The usual notation is $r(\cdot, t)$ for $t \in \mathcal{A}$, and $r(\omega, t)$ is an indexed set of values called a sample function; there is such a sample function for each $\omega \in \Omega$. When there is no confusion, the notation $r(t)$ is commonly used. Likewise, for each fixed t, $r(\cdot, t)$ is a random variable. The use of the letter t in the indexing set

---

*Underlined terms are also defined in Appendix A.

suggests that "time" is the most common indexing parameter. Furthermore, in later chapters, the indexing set $\mathcal{A}$ will be the positive integers, which corresponds to discrete time. A discussion of the necessary topics in stochastic process theory is given in Chapter 2.

1.1.2 Forms of Stationarity and Nonstationarity. Nonstationarity is a negative concept so it is defined as various forms of negations of stationarity. A strict-sense stationary process (SSSP) has all its statistics not affected by a shift in the time origin. A wide-sense stationary process (WSSP) has an expected value which is constant and an autocorrelation which depends only on the time difference.

$$E\{r(\omega, t)\} = \mu = \text{constant} \tag{1.1}$$

$$E\{r(\omega, t + \tau)\, r(\omega, t)\} = R(\tau) \tag{1.2}$$

In later chapters, only Gaussian processes will be used, so that an important remark should be made. For a Gaussian process, since the mean and autocorrelation uniquely determine all statistics, strict-sense and wide-sense properties are the same.

Nonstationary processes are studied in this thesis. The particular type of nonstationarity allowed is that described by the phrase not wide-sense stationary. Another statement of this condition is that the mean and autocorrelation are allowed to be time-varying.

$$E\{r(\omega, t)\} = \mu(t) \tag{1.3}$$

$$E\{r(\omega, t + \tau)\, r(\omega, t)\} = R(\tau, t) \ . \tag{1.4}$$

Without loss in generality, a further stipulation is usually made that the (unconditional) expectation be zero for all t. Thus, there remains to specify only the second-order statistic $R(\tau, t)$.

1.1.3 Models. The first and most significant word in the title is the word "models"; in mind, of course, are mathematical models. A natural question which comes up in such work is, "Why are models developed?" A brief discussion of this fundamental point seems apropos and several reasons for modeling are given.

Often, a model is the only tractable description of the physical process being studied; indeed, the real world is very complex and many factors are involved, so that judgement and simplification are necessary to describe the process.

In many human endeavors, there is often an ignorance (used in the dictionary sense, and not as derogatory) of the physical laws, or at best we have sparse information on governing laws. Modeling, then, can be a technique to describe our observations.

A third reason for modeling might be an inability to solve physical laws even though they are readily described. Such solutions might be unavailable due to the high dimensions involved or due to non-computability, or they might not be economically feasible.

These are some of the reasons for modeling. The reader will easily find examples of these points in his own field of studies. The application of such models in this work is to acoustic signal processes.

1.1.4 Objectives. The main objective is to study a class of stochastic processes and a stated purpose is to develop models for such processes. To be more specific, two classes of general problems will be treated: the first is called the filtering problem and the second is called the prediction problem.

The Filtering Problem

Given the observed values of a stochastic process over some interval of time, find the conditional probabilities of all values of another related stochastic process.

The Prediction Problem

Given the observed values of a stochastic process over some interval of time, find the conditional probabilities of all future values of the stochastic process or some related stochastic process.

There is another objective. Since nonstationary stochastic processes are being studied and since the models will have time-varying parameters, a purpose of the experimental part of this thesis is to accumulate information on such statistics and parameters. Such information may lead to further understanding of the physical laws or

to more sophisticated models.

1.1.5 Acoustic Signal and Noise Processes. Oceanographic acoustic signals and noise processes have already been mentioned as the specific area of application of the theory and models. While a deterministic model has many advantages and may be practical under a controlled set of circumstances, a model with stochastic inputs seems entirely appropriate for the solutions sought. In the opinion of the writer, a stochastic model, in which the parameters are themselves unknown and stochastic parts of such a model, is a reasonable approach to the problems. This necessity for a stochastic study is elaborated upon below.

Consider the physical system under study: an ocean or a large part of it. It is a large body of water which is under constant change; it is affected by daily, lunar and seasonal tides, by weather conditions. Ambient ocean noise is affected by the temperature and salinity of the water, and is contributed to by earth tremors, by biological noise and by man-made sounds from inside and outside. Propagation conditions can bring sound from several hundred miles and yet "mask out" nearby sources. How can we account for the vagaries of a school of fish, the path of a rain storm or ship?

Of course, this study is not of such detail. The interest is in a gross description of the net effect of all these causes and not the causes themselves. The approach is to observe the process, model it and then process or filter.

This leads to a philosophy of stochastic models which is briefly described. Under strictly ergodic assumptions, a "measurement of the system" can be made and if this can be done often enough, perhaps the result will be a smooth picture of the actual process. However, such an approach is not acceptable due to the inherent "steps" in the model or the need for much processing. Rather, the philosophy adopted by a model-builder is that reasonable and available structure must be built into the model at the outset. This admits a time-varying model and a recursive solution which is smoothed and economical when the required conditions are met. This is a personal viewpoint and is subject to critical review.

1.2 Historical Background

A brief historical background will include the major contributions to the topics mentioned above. The foundations of probability theory were laid down by Kolmogorov (1933) and within the last 25 years, stochastic process theory has become a branch of mathematics (Ref. 5). Probability had early application in such areas as detection theory and communication theory, but stochastic process theory has had engineering application only within the last 15 years. A celebrated stochastic process is the Gauss-Markov process, which found application in the development of the now-classical Kalman-Bucy filter (Ref. 11). A survey of the literature indicates extensive use of these concepts (Ref. 4), though publications in the open literature

are still sparse.

There are other areas of background interest to this report. Parameter estimation is well-treated in writings on statistics and yet it is an area of present-day research, with conditions such as non-stationarity and dimensionality to investigate. At the foundation of the processes to be modeled, sit extensive studies on acoustics and oceanography (Refs. 18 and 21). These disciplines have traditionally taken a physical approach and they seek laws and cause-effect relations. The stochastic modeling approach, on the other hand, depends on physics for guidance and direction but attempts to obtain statistical descriptions directly from observations.

Through this sketch of the background of the problem, the interested reader will find additional material to complete the picture according to his needs. This introduction to the problem is intended to be general and indicative of the methods used in the following chapters. A plan of the thesis follows.

1.3 Plan of the Thesis

This chapter is introductory and states the problems and methods of solution. The body of the thesis can be thought of in two parts: one is model development, Chapters 2, 3, 4 and the other is data processing, Chapters 5, 6, 7, and 8.

Chapter 2 states results of stochastic process theory which are essential to the development and presents the state vector and

matrix representation of the model to be used. Chapter 3 contains a derivation of the recursive equations and states the formal theorem which is the result known as the Kalman-Bucy filter. In Chapter 4, two classes of parameter estimation schemes are discussed. For a stationary process and for other stated conditions, a consistent set of estimators of the model parameters have been found. For a nonstationary process with certain other properties, a set of estimators is proposed and evaluated.

Chapter 5 is a self-contained discussion of a class of linear filters and digital processors which are used in the data processing of Chapters 6, 7 and 8. Chapter 6 is a general description of the required data processing, while Chapter 7 is an application to a process which is a signal power measurement and Chapter 8 is an application of the model to a process which is a noise power measurement. The last chapter summarizes some general observations and concludes the thesis.

1.4 Notation and Format

A special comment on notation seems desirable. The notation in the literature on such diverse topics is, understandably, diverse. Wherever possible, the notation in this thesis has been made the same as that in the literature on Kalman-Bucy filters; and in particular, the notation in Bucy and Joseph (Ref. 4) is followed with but a few exceptions. Other notation is consistent with this policy and

the present practice of journals. To assist the reader, a List of Symbols and explanations is included.

The format of the thesis is unusual in two respects. A synopsis of each chapter is given at the beginning of the chapter; hopefully, this synoptic presentation will assist the reader. Also, definitions of many technical terms have been collected in the first appendix, which is a technical glossary. Readers from different background will find this an assistance.

CHAPTER 2

STOCHASTIC PROCESSES AND MODEL EQUATIONS

Synopsis

2.1 Introduction

2.2 Stochastic Processes

- Markov property
- Gaussian process
- Ergodicity

2.3 State Variables and Matrix Notation

- Definitions
- Linear vector model
- State transition matrix
- Example

2.4 Model Equations

- Conditions of the model
- Process equation
- Observation equation
- Initial conditions

2.1 Introduction

This chapter presents the elements of stochastic process theory which are needed. A review of conditional probability and conditional expectation is given, followed by a discussion of some properties of a Markov process. The processes studied will be shown to be Gaussian, so that they are usually called Gauss-Markov (GMP). A brief discussion of ergodicity is relevant at that point.

Another section contains some features of state variable theory, including the matrix differential and difference equation representations of systems of equations. Some of the advantages of this theory are stated and an example will illustrate these concepts.

With the preliminary technical background completed, the model is presented, accompanied by general discussions on conditions and validity of such a model.

2.2 Stochastic Processes

Stochastic process theory is a modern and growing branch of probability theory. A presentation such as this can give only the essential elements needed in this development. Fundamental concepts in probability are conditional probability and conditional expectation. Their general definitions are given in modern terminology in Appendix A, and their more usual definitions are given here.

Conditional probability: If the random variable y takes values in a set B and if $P(y \in B) > 0$, then the conditional probability of A

given $y \in B$ is defined by

$$P(A \mid y \in B) = \frac{P(A,\ y \in B)}{P(y \in B)} \tag{2.1}$$

Conditional expectation: When the conditional probability $P(A \mid y \in B)$ exists, the conditional expectation of x given $y \in B$ is defined by

$$E\{x \mid y \in B\} = \int_A x \, dP(x \in A \mid y \in B) \tag{2.2}$$

It is significant to note that both of these quantities are random variables and that they each enjoy the properties of probabilities and expectations, respectively. The modern writings on probability put considerable emphasis on the generalizations of these concepts and their results; for an understandable treatment, the reader should consult a text such as that by Breiman (Ref. 3).

When the probability characterization can be expressed as probability density functions, as is the case in this thesis, the conditional probability density and conditional expectation take the following form.

$$p(x \mid y_1, y_2, \ \cdots, \ y_k) = \frac{p(x, y_1, y_2, \ \cdots, \ y_k)}{p(y_1, y_2, \ \cdots, \ y_k)} \tag{2.3}$$

$$E(x \mid y_1, y_2, \ \cdots, \ y_k) = \frac{\int_A x \, p(x, y_1, y_2, \ \cdots, \ y_k) \, dx}{p(y_1, y_2, \ \cdots, \ y_k)} \tag{2.4}$$

2.2.1 Markov Property. The definitions of a random variable and a stochastic process were given in Chapter 1. A special class of processes is now defined: a Markov process is a stochastic process with the following Markov property.

$$P(x(t_{k+1}) < \lambda | x(t_k), \ldots, x(t_1)) = P(x(t_{k+1}) < \lambda | x(t_k)) \quad \text{a.s.} \tag{2.5}$$

for any $t_{k+1} \geq t_k \geq \ldots \geq t_1 \geq 0$ and for any λ. (Almost surely, a.s. or "with probability 1" are technical terms in probability.) This condition is more than adequate for this presentation, yet a more general definition is given in Appendix A. Strictly speaking, the Markov property is a statement about the conditional probability at the one instant t_{k+1} in the future. However, this extends to a general statement about the future, given the present and past.

Proposition: Let A be a set of future values of the Markov process x(t); then

$$P(A | x(s), s \leq t) = P(A | x(t)) \quad \text{a.s.} \tag{2.6}$$

Speaking intuitively and imprecisely, the statement is often made that, for a Markov process, the future is independent of the past when the present is given.

So far, the random variable x(t) (or x_k) has been considered to be a scalar and such a process is often called Markov-1.

The extension to the vector case is straightforward. Henceforth, let x denote a column vector of n elements.

$$x = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{bmatrix} \tag{2.7}$$

Its m'th component is $x^{(m)}$ and will seldom be needed as notation, and x_k denotes the vector random variable at the (discrete) time k. Similar Markov properties are given to the vector x and the process is then called Markov-n or vector Markov-1.

2.2.2 Gaussian Process. A further condition placed on the process is that every finite set of the x_k's must be jointly Gaussian; it is a Gaussian process. This is a very important practical constraint, because only for this case have extensive results been found. The Gaussian probability density function is familiar to the reader in the form

$$p_x(X) = \frac{1}{(2\pi)^{n/2} \, |\Lambda|^{\frac{1}{2}}} \exp[-\tfrac{1}{2}(X-\mu_x)' \, \Lambda_x^{-1} (X-\mu_x)] \tag{2.8}$$

where μ_x is the vector mean: $\mu_x = E\{x\}$

Λ_x is the covariance matrix: $\Lambda_x = E\{(x-\mu_x)(x-\mu_x)'\}$.

Familiarity with the properties of this distribution is presumed.

The actual processes to be studied in later chapters are acoustic signal power measurements expressed in db (decibels). A brief note (Ref. 12) in reference to a recent article (Ref. 15) shows that such processes are Gaussian; that proof is adequate.

Proposition: Let $z(t)$ be a stationary Gauss-Markov process with the continuous covariance function $R(t,u) = \exp[-\alpha|t-u|]$. Define the statistic

$$y = \log\left[\frac{1}{a}\int_{-\frac{a}{2}}^{\frac{a}{2}} z^2(t)\,dt\right] \tag{2.9}$$

and let $\beta = \frac{\alpha a}{2}$. Then the probability density for y is approximately

$$p_y(Y) = \frac{1}{(2\pi/\beta)^{\frac{1}{2}}} \exp[\beta - \beta \cosh Y - Y/2] \tag{2.10}$$

and for $\beta > 4$, this is nearly Gaussian with mean $\frac{-1}{2\beta}$ and variance $\frac{1}{\beta}$.

The significance of this result is that it is the only definite assurance that the processes under study are Gaussian, which is an essential ingredient of the recipe. The other necessary condition is the Markov property, so that Gauss-Markov theory can be applied.

2.2.3 Ergodicity. Another fundamental concept is ergodicity, which concerns the statistics of a stochastic process. The intuitive

expression of the ergodic property is that time averages equal ensemble averages and this presumes that relevant statistics of a process can be determined from a single sample function $x(\omega, t)$. The formal definitions of an ergodic process and an ergodic transformation require several concepts from measure theory, so that these definitions are given in Appendix A. For a discussion of ergodicity, the reader should consult the book by Breiman (Ref. 3). The ergodic property is usually stated as the following ergodic theorem.

Theorem 2.1. Let T be measure-preserving on $(\Omega, \mathscr{F}, P)$. Then for x any random variable such that $E|x| < \infty$,

$$\lim \frac{1}{n} \sum_{k=0}^{n-1} x(T^k \omega) = E\{x|\mathscr{I}\} \qquad \text{a.s.} \qquad (2.11)$$

Such ergodic properties have considerable significance in the experimental study of physical systems. Whenever a measurement is made, an average of observations is taken or any other filtering is accomplished, some form of ergodicity is being used. Specifically, local ergodicity is often required; the ergodic property is assumed to hold over an interval $(-a, a)$ but nothing is said outside this interval.

Consider the time average

$$\mu_a = \frac{1}{2a} \int_{-a}^{a} x(t)\, dt \quad . \qquad (2.12)$$

Since x is a random variable, so is μ_a, but its expectation

$$E\{\mu_a\} = E\{x(t)\} \tag{2.13}$$

is a constant if x is stationary, or a known function of t if x is nonstationary. Furthermore, if the variance of μ_a tends to zero as the interval becomes large, the process x is said to be ergodic in the mean. Conditions for this and similar properties are discussed by Papoulis (Ref. 13) and some examples are given there.

A form of ergodicity used in Chapter 4 is called ergodicity of the autocorrelation; this means that the time average

$$R_a(\lambda) = \frac{1}{2a} \int_{-a}^{a} x(t+\lambda)\, x(t)\, dt \tag{2.14}$$

has a limiting value $\lim_{a \to \infty} R_a(\lambda) = R(\lambda)$, which is the autocorrelation of x, and this occurs if and only if

$$\lim_{a \to \infty} \frac{1}{a} \int_{0}^{2a} \left(1 - \frac{\tau}{2a}\right) (R_{\phi\phi}(\tau) - R^2(\lambda))\, d\tau = 0 \tag{2.15}$$

where

$$\phi(t) = x(t+\lambda)\, x(t) \tag{2.16}$$

$$R_{\phi\phi}(\tau) = E\{x(t+\lambda+\tau)\, x(t+\tau)\, x(t+\lambda)\, x(t)\} \quad . \tag{2.17}$$

Such an ergodic property for second order statistics seems necessary, so it is assumed henceforth and used by the estimators of Chapter 4.

2.3 State-Variables and Matrix Notation

In the study of dynamical systems, the classical technique used to describe such systems is the formulation of differential and difference equations. Such theory is then concerned with the existence and uniqueness of solutions under various boundary conditions. A modern formulation of the last 15 years, called state-variable theory, casts these same equations in matrix differential-difference form. The result is a time-domain formulation which reverses the trend to time-invariant transform theory. Some of the major ideas in this theory will be presented, for the model will be defined in this manner.

2.3.1 Linear Vector Equation. This presentation is limited to the linear case, though the theory is not. Linearity is very important in many applications and makes possible the pleasing results of the next chapter. A linear set of differential equations is often an easy and direct way to describe a physical system; the equations model the physical system, sometimes adequately but usually for only general characteristics. In general, a system of order n is studied but all the examples below are for $n = 1$ or $n = 2$.

$$\frac{d^2 z(t)}{dt^2} + a(t) \frac{d z(t)}{dt} + b(t)\, z(t) = c(t) \tag{2.18}$$

$$z(0) = z_0 \tag{2.19}$$

$$\frac{d z(0)}{dt} = v_0 \tag{2.20}$$

This equation is a second-order linear differential equation in its familiar form; in general, the coefficients a and b are known functions of time and c(t) is a known "driving function." To specify the solution of an n'th-order equation, the values $z(t), \ldots \frac{d^{(n-1)} z(t)}{dt^{n-1}}$ at t_0 are needed and n variables are required to define the solution at any time. For $n = 2$, define

$$x^{(1)}(t) = z(t) \qquad \text{and} \qquad x(t) = \begin{bmatrix} x^{(1)}(t) \\ x^{(2)}(t) \end{bmatrix} \tag{2.21}$$

$$x^{(2)}(t) = \frac{d z(t)}{dt}$$

x(t) is called the <u>state vector</u> of the system because it describes the <u>state of the system</u>, which is the minimal amount of information at t_0 about the effects of past inputs necessary to describe completely the output for $t \geq t_0$. The set of all possible states is the <u>state space</u>.

A simple derivation of the state equation for Eq. 2.18 is given. Observe that

$$\frac{d\, x^{(1)}(t)}{dt} = x^{(2)}(t) \tag{2.22}$$

$$\frac{d\, x^{(2)}(t)}{dt} = \frac{d^2\, z(t)}{dt^2} = c(t) - a(t)\, x^{(2)}(t) - b(t)\, x^{(1)}(t) \tag{2.23}$$

so that the state equation is the vector-matrix first-order differential equation

$$\frac{d\, x(t)}{dt} = F(t)\, x(t) + w(t) \tag{2.24}$$

where

$$F(t) = \begin{bmatrix} 0 & 1 \\ -b(t) & -a(t) \end{bmatrix} , \tag{2.25}$$

$$w(t) = \begin{bmatrix} 0 \\ c(t) \end{bmatrix} \quad \text{and} \quad x(0) = \begin{bmatrix} z_0 \\ v_0 \end{bmatrix} \tag{2.26}$$

The matrix F is called the coefficient matrix and may be time-varying and $w(t)$ is the vector driving term.

The solution of Eq. 2.24 and conditions for existence and uniqueness are well-known. The solution consists of a homogeneous part and a particular part in the form

$$x(t) = \Phi(t, t_0)\, x(t_0) + \int_{t_0}^{t} \Phi(t, \tau_0)\, w(\tau)\, d\tau \tag{2.27}$$

where $\Phi(t, t_0)$ is discussed next.

2.3.2 State Transition Matrix. $\Phi(t, t_0)$ is the state transition matrix and is defined to be the solution of the differential equation

$$\frac{d\,\Phi(t, t_0)}{dt} = F(t)\,\Phi(t, t_0) \tag{2.28}$$

with initial condition $\Phi(t_0, t_0) = I$. This matrix has many interesting properties, two of which will be needed in following chapters. It is positive definite and always has an inverse $\Phi^{-1}(t, t_0)$. For the time-invariant case, F is constant and Φ is a function of the difference $t - t_0$ only. A simple analytical solution for Φ is then always possible. But for the time-varying case, no such easy solution has been found and it is usually most expedient to integrate the differential equation (Eq. 2.28). There lies the crux of the problem in dealing with time-varying systems or nonstationary processes. For small dimension, such computations are feasible and practical application has been made using $n = 9$; but large dimensionality precludes the use of the full theory.

For a comprehensive treatment of linear system theory, the book by Zadeh and Desoer (Ref. 23) is an excellent source. There are short sections on this topic in many of the references cited in later chapters.

Some of the advantages of the state-variable formulation should

be pointed out. Stochastic, nonstationary processes and finite time intervals are easily included. Second, the solution is in a form such that it can be implemented on a computer. Also, there is a strong duality between estimation and control theory which this theory can exploit. A fourth advantage is that nonlinear systems can be studied by this approach.

A list of advantages should be tempered by some disadvantages. Closed-form or intuitively pleasing results are often difficult to obtain and also, some problems are not so readily solved by this approach.

2.3.3 An Example. A simple example will illustrate the theory and prepare for a later application. Consider the time-invariant equation (Eq. 2.24) with $n = 2$ and

$$F = \begin{bmatrix} 0 & 1 \\ -\lambda^2 & 0 \end{bmatrix} \tag{2.29}$$

This is recognized as having a homogeneous solution which is undamped oscillatory. The solution for Φ from Eq. 2.28 is readily obtained as

$$\Phi(t, t_0) = \begin{bmatrix} \cos \lambda(t - t_0) & \frac{1}{\lambda} \sin \lambda(t - t_0) \\ -\lambda \sin \lambda(t - t_0) & \cos \lambda(t - t_0) \end{bmatrix} \tag{2.30}$$

and so from Eq. 2.27

$$x(t) = \begin{bmatrix} x_0 \cos \lambda(t - t_0) + \frac{v_0}{\lambda} \sin \lambda(t - t_0) \\ \\ -\frac{x_0}{\lambda} \sin \lambda(t - t_0) + v_0 \cos \lambda(t - t_0) \end{bmatrix} + u(t) \qquad (2.31)$$

The first is the homogeneous part and the second is the particular part,

$$u(t) = \int_{t_0}^{t} \Phi(t, \tau)\, w(\tau)\, d\tau \qquad (2.32)$$

The change to discrete time and difference equations will now be made. Let t take on only integer values $t = k$ and $t_0 = 0$. Then

$$\Phi(t, t_0) = \Phi(k, 0) \qquad (2.33)$$

and by the time-invariance of this solution for Φ,

$$\Phi(k, k-1) = \Phi(1, 0) \qquad \text{for all } k \ . \qquad (2.34)$$

In general, for the time-varying case, Φ is a function of both variables and not just their difference. Because of the iterative nature of the following formulation of the model and its subsequent use, there is interest only in $\Phi(k, k-1)$, which will be denoted by the abbreviation

$$\Phi_k \equiv \Phi(k, k-1) \qquad (2.35)$$

and called the one-step state transition matrix. Similarly, the

particular part will have the notation

$$u_k = \int_{k-1}^{k} \Phi(k, \tau)\, w(\tau)\, d\tau \tag{2.36}$$

and $\{u_k\}$ is a sequence of inputs. The discrete form of the state equation is the vector equation

$$x_k = \Phi_k x_{k-1} + u_k \qquad k \geq 1 \ . \tag{2.37}$$

An approximation will be used in Chapter 7, which requires the following development. The one-step transition matrix Φ_k of the example is

$$\Phi_k = \begin{bmatrix} \cos\lambda & \frac{1}{\lambda}\sin\lambda \\ -\lambda\sin\lambda & \cos\lambda \end{bmatrix} \tag{2.38}$$

In the applications, the system natural frequency λ will be small; $\lambda << 1$, so a reasonable approximation is

$$\Phi_k = \begin{bmatrix} 1 - \frac{\lambda^2}{2} & 1 - \frac{\lambda^2}{6} \\ -\lambda^2 & 1 - \frac{\lambda^2}{2} \end{bmatrix} \tag{2.39}$$

$$= \begin{bmatrix} \phi_k & 1 - (1-\phi_k)/3 \\ -2(1-\phi_k) & \phi_k \end{bmatrix}$$

where $\phi_k = 1 - \frac{\lambda^2}{2}$ and could be time-varying.

2.4 Model Equations

The stochastic sequence will be modeled by a set of two vector equations; one is called the process equation and the other is called the observation equation; together, they will be called the model equations.

Process: $$x_k = \Phi_k x_{k-1} + u_k \tag{2.40}$$

Observation: $$y_k = H_k x_k + v_k \tag{2.41}$$

For the vector case, x_k and u_k are n-dimensional vectors and y_k and v_k are m-dimensional vectors. The sequences $\{u_k\}$ and $\{v_k\}$ are assumed to be Gaussian white noise (GWN) sequences with zero mean; their properties are completely characterized by their covariances. The variance Q_k for u_k and R_k for v_k are defined as shown.

$$E\{u_p u_k'\} = \delta_{pk} Q_k \tag{2.42}$$

$$E\{v_p v_k'\} = \delta_{pk} R_k \tag{2.43}$$

where δ_{pk} is the Kronecker δ. The $n \times n$ matrix Φ_k has already been discussed as the state transition matrix for the process x_k. The observation matrix H_k is $m \times n$ and is the link between the

observation and the process. v_k is the observation noise and is part of the reason why the observation does not contain full information about x. The other part of the reason is that H is $m \times n$ dimensional, $m \leq n$ and so for $m < n$ not all of x is observable. $\{u_k\}$ is called the driving stochastic sequence (by analogy to the case when u_k is deterministic) and is the reason why $\{x_k\}$ is a random sequence.

Since the driving variables u_k and v_k, are Gaussian and the equations are linear, the process and observation random variables, x_k and y_k, are Gaussian. Furthermore, by the zero-mean properties of u_k and v_k, x_k has zero-mean a priori and so has y_k. But given observations, the conditional means will not be zero. The conditional mean of the process x_k given all past observations will be shown to be the required solution for the filtering problem.

The Markov property of x_k is evident from the process equation. The recursive relation establishes this, and any Markov sequence can be put into this form. The dimension of x is n and , determines the order of the Markov process.

The model has considerable intuitive appeal. An n-th order linear dynamical system is represented by the matrices Φ and H. This system has the input u_k, a stochastic sequence which is well-described, and observation "noise" v_k, also a well-defined stochastic process. We ask for the probability characterization of the output process x, given the observations y.

CHAPTER 3

KALMAN-BUCY FILTER

Synopsis

3.1 Introduction

3.2 Kalman Filter

Bayes estimation

Recursive equations for $\hat{x}_1$, $\hat{x}_2$

The solutions -- filtering

-- prediction

3.3 Continuous-Time Filter

Estimator equation

Gain equation

Variance equation

3.4 Initial Conditions

Transient conditions

Steady-state conditions

3.5 Performance Indices

3.1 Introduction

The previous chapter has presented the necessary topics in stochastic process theory, which have been used to build a general model for stochastic processes. The concepts and notations of state-variable theory and matrix representations of differential equations have been introduced. These are very useful in the precise statements and solutions of the filtering and prediction problems.

The stochastic sequence we are studying is Gauss-Markov. These conditions of Gaussian statistics and the Markov property are central to the tractable solution of the problems. Their implications will be expressly stated in what follows.

The formulation and solution of the filtering and prediction problems for the discrete case were first presented by Kalman (Refs. 9 and 10). Subsequent work by Kalman and Bucy developed corresponding equations for the continuous case (Ref. 11). The whole topic became known as "Kalman-Bucy filters," although not all the equations represent filters. Recent treatments of the subject are to be found in the books by Van Trees (Section 6.3.2) (Ref. 22) and by Bucy and Joseph (Ref. 4).

This theory is well-documented, so the purpose is merely to present the highlights of the development and establish the notation of the solutions.

3.2 Kalman Filter

The prediction and filtering problems have already been discussed in general terms. Under the condition that x and y are Gauss-Markov, the solutions are quite simple in principle. The conditional probability distribution of a Gaussian process is completely described by its conditional mean and conditional covariance. From the Markov property, the mean and variance need to be known only for the latest discrete value of time.

3.2.1 Bayes Estimation. In estimation problems of the Bayes type, an a priori density $p_x(x)$ is assumed known. The cost function $C(\tilde{x})$ may be taken to be a function of a single variable $\tilde{x} = x - \hat{x}$, where $\hat{x} = \hat{x}(y)$ is the estimate, x is the true value and $\tilde{x}$ is the error. y is the observed variable. Then the risk is defined as

$$\mathcal{R} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C[X - \hat{x}(Y)]\ p_{x,y}(X, Y)\, dX\, dY \qquad (3.1)$$

and the Bayes estimate minimizes this risk.

When C is quadratic in $\tilde{x}$,

$$C(\tilde{x}) = \tilde{x}'\,\tilde{x}\ , \qquad (3.2)$$

the solution is called the minimum mean square error (MMSE) estimate. Since $p_{x,y} = p_y\, p_{x|y}$, the risk becomes

$$\mathcal{R}_{ms} = \int_{-\infty}^{\infty} p_y(Y) \int_{-\infty}^{\infty} [X - \hat{x}(Y)]' [X - \hat{x}(Y)] \; p_{x|y}(X|Y) \, dX \, dY \quad (3.3)$$

These are m and n-fold integrals. $\mathcal{R}_{ms}$ is minimized by minimizing the inner integral; differentiating with respect to $\hat{x}(Y)$ and setting the result to zero produces the unique minimum

$$\hat{x}(Y) = \int_{-\infty}^{\infty} X \, p_{x,y}(X|Y) \, dX \quad (3.4)$$

which is the conditional mean (the mean of the a posteriori density).

Actually, this estimate is optimum for a broader class of problems. Under Gaussian conditions, it is optimum even for other cost functions. These additional features will not be needed in what follows.

3.2.2 Recursive Equations for $\hat{x}_1$, $\hat{x}_2$. Recursive equations for the conditional mean will now be developed; some of the details are included.

Let the initial condition on x by x_0; x_0 may be a random variable. Let it have the known variance matrix $E\{x_0 x_0'\} = \Lambda_0$ and zero mean. The estimates $\hat{x}_1$ and $\hat{x}_2$ will be found, then the general solution for the filtering and prediction problems will be given.

The conditional probability $p_{x_1|y_1}$ is needed.

$$p_{x_1|y_1}(X|Y) = \frac{p_{y_1|x_1}(Y|X)\, p_{x_1}(X)}{p_{y_1}(Y)} \tag{3.5}$$

$$p_{y_1|x_1}(Y|X) = \text{constant} \cdot \exp[-\tfrac{1}{2}(Y - HX)'\, R^{-1}(Y - HX)] \tag{3.6}$$

$$p_{x_1}(X) = \text{constant} \cdot \exp[-\tfrac{1}{2}(X - \Phi x_0)'\, \Lambda_0^{-1}(X - \Phi x_0)] \tag{3.7}$$

Therefore,

$$p_{x_1|y_1}(X|Y) = \text{constant} \cdot \exp[-\tfrac{1}{2}(Y - HX)'\, R^{-1}(Y - HX) - \tfrac{1}{2}(X - \Phi x_0)'\, \Lambda_0^{-1}(X - \Phi x_0)] \tag{3.8}$$

This conditional distribution of x_1 given y_1 is Gaussian. Thus there are $\hat{x}(Y)$ and Λ_1 such that

$$p_{x_1|y_1}(X|Y) = \text{constant} \cdot \exp[-\tfrac{1}{2}(X - \hat{x}(Y))'\, \Lambda_1^{-1}(X - \hat{x}(Y))] \,. \tag{3.9}$$

The conditional mean $\hat{x}(Y)$ and conditional covariance Λ_1 can be found by completing the quadratic form of Eq. 3.8. The quadratic term, $X'[\Lambda_0^{-1} + H'\, R^{-1}\, H]\, X$ identifies

$$\Lambda_1^{-1} = \Lambda_0^{-1} + H'\, R^{-1}\, H \,. \tag{3.10}$$

The linear terms in X eventually yield

$$\Lambda_1^{-1}\,\hat{x}_1(y_1) = \Lambda_0^{-1}\,\Phi\,\bar{x}_0 + H'\,R^{-1}\,y_1 \ , \tag{3.11}$$

$$\hat{x}_1(y_1) = \Phi\,x_0 + \Lambda_1\,H'\,R^{-1}\,(y_1 - H\,\Phi\,x_0) \ . \tag{3.12}$$

Now the filtering error covariance matrix $\bar{P}_1$ is defined by

$$\bar{P}_k \triangleq E\,\{\tilde{x}_{k|k}\,\tilde{x}_{k|k}'\} \qquad k = 1, 2, \ \ldots \tag{3.13}$$

where $\tilde{x}_{k|k} \triangleq x_k - \hat{x}_{k|k}$.

This is the error in the estimate for x_k given observations up to and including time k. Since the estimate is the conditional mean, this is a convenient notation for the conditional expectation.

$$\hat{x}_{k|k} \triangleq E\,\{x_k | y_k, y_{k-1}, \ \ldots, \ y_1, \ x_0\} \qquad k \geq 1 \tag{3.14}$$

$\tilde{x}_{k|k}$ has zero mean so $\bar{P}_1 = \Lambda_1 = [\Lambda_0^{-1} + H'\,R^{-1}\,H]^{-1}$.

The estimate for x_2 is found as follows. The observation equation $y_k = H_k x_k + v_k$ with independent v_k implies that x_2 dominates y_1 in regard to the distribution of y_2 ; in simple words, given x_2, y_1 does not provide any more information about y_2. This is expressed as

$$p_{y_2|x_2, y_1} = p_{y_2|x_2} \tag{3.15}$$

and is where the Markov condition is used.

Consider the joint probability of x_2 and y_2 conditional to y_1.

$$p_{y_2, x_2|y_1} = p_{y_2|x_2, y_1}\, p_{x_2|y_1} = p_{x_2|y_2, y_1}\, p_{y_2|y_1} \tag{3.16}$$

Substituting Eq. 3.15 in Eq. 3.16,

$$p_{y_2|x_2}\, p_{x_2|y_1} = p_{x_2|y_2, y_1}\, p_{y_2|y_1} \tag{3.17}$$

or

$$p_{x_2|y_2, y_1} = \frac{p_{y_2|x_2}\, p_{x_2|y_1}}{p_{y_2|y_1}} \,. \tag{3.18}$$

Let $$\hat{x}_{k|k-1} \triangleq \Phi\, \hat{x}_{k-1|k-1} \tag{3.19}$$

and $$\tilde{x}_{k|k-1} \triangleq x_k - \hat{x}_{k|k-1} \,. \tag{3.20}$$

$$P_1 \triangleq \Lambda_0 \tag{3.21}$$

The prediction error covariance matrix P_k is defined by

$$P_k \triangleq E\{\tilde{x}_{k|k-1}\, \tilde{x}'_{k|k-1}\} \qquad k = 2, 3, \ldots \tag{3.22}$$

In Eq. 3.18, $p_{y_2|x_2}$ is normal with mean Hx_2, variance R, by virtue of the observation equation

$$y_2 = H x_2 + v_2 \,. \tag{3.23}$$

The density $p_{x_2|y_1}$ is obtained from the process equation

$$x_2 = \Phi x_1 + u_2 \quad . \tag{3.24}$$

For given y_1, $\hat{x}_1(y_1)$ is known from Eq. 3.12, and

$$x_1 = \hat{x}_1(y_1) + \tilde{x}_1 \tag{3.25}$$

where $\tilde{x}_1$ is normal with mean 0 and variance $\overline{P}_1$. So $p_{x_2|y_1}$ is normal with mean $\Phi \hat{x}_1$ and variance

$$P_2 = \Phi \overline{P}_1 \Phi' + Q \; . \tag{3.26}$$

It has just been established that the two factors in the numerator of Eq. 3.18 are quadratic exponentials in $(y_2 - H x_2)$ and $(x_2 - \Phi \hat{x}_1(y_1))$ respectively. Since the denominator of Eq. 3.18 does not depend on x_2, $p_{x_2|y_2, y_1}$ depends on x_2 only through these quadratic exponentials and hence must be normal. The following results are obtained by completing the quadratic form as before.

$$\overline{P}_2^{-1} = P_2^{-1} + H' R^{-1} H \tag{3.27}$$

$$\hat{x}_2(y_2) = \Phi \hat{x}_1 + P_2 H' (H P_2 H' + R)^{-1} (y_2 - H \Phi \hat{x}_1) \tag{3.28}$$

The following theorem summarizes these results for the recursion from time k-1 to time k. The matrices Φ, Q, R are

presumed known, and they may be time-varying. In the above notation, $\hat{x}_1(y_1) = \hat{x}_{1|1}$ and $\hat{x}_2(y_2) = \hat{x}_{2|2}$; $\hat{x}_{1|0} = \Phi x_0$, $x_{0|0} = x_0$.

3.2.3 The Solutions. The filtering problem has the following solution; it is called the Kalman filter or the Kalman-Bucy equations.

Theorem

The model with known matrices Φ_k and H_k and known noise covariances Q_k and R_k, has the following linear form:

$$x_k = \Phi_k x_{k-1} + u_k \tag{3.29}$$

x_0 = random variable,
initial condition on x

$$y_k = H_k x_k + v_k \tag{3.30}$$

$$E\{u_j u_k'\} = \delta_{jk} Q_k \tag{3.31}$$

$$E\{v_j v_k'\} = \delta_{jk} R_k \tag{3.32}$$

where δ_{jk} is the Kronecker δ. Independence between u_k and v_k for all k is assumed and u, v are independent of x, y. Suppose $R_k \geq aI$ for $a > 0$, then $\hat{x}_{k\,k}$ is the solution of the difference equations

$$\begin{aligned} \hat{x}_{0|0} &= x_0 \\ \hat{x}_{k|k-1} &= \Phi_k \hat{x}_{k-1|k-1} \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + A_k(y_k - H_k \hat{x}_{k|k-1}) \end{aligned} \tag{3.33}$$

for $k = 1, 2, \ldots$, where

$$A_k = P_k H_k' (H_k P_k H_k' + R_k)^{-1} \qquad k \geq 1 \tag{3.34}$$

$$P_k = \Phi_k \overline{P}_{k-1} \Phi_k' + Q_k \qquad k > 1 \tag{3.35}$$

$$P_1 = E\{x_0 x_0'\}$$

$$\overline{P}_k = (I - A_k H_k) P_k (I - A_k H_k)' + A_k R_k A_k' \tag{3.36}$$

Observe that P_k must be computed first, then A_k. $\overline{P}_k$ is then computed for the next iteration. The formal statement of this theorem and its proof are given by Bucy and Joseph (Ref. 4). Their method of proof is different than the recursive derivation of Section 2.2.

Some general comments can be made. The equation for $\hat{x}_{k|k}$ is a linear combination of the previous estimate and the latest observation. Since it minimizes the risk, it is called the optimal estimate. In the form given, $\hat{x}_{k|k-1}$ is computed from the estimate $\hat{x}_{k-1|k-1}$; $H_k \hat{x}_{k|k-1}$ can also be computed before y_k is observed.

Only one matrix multiplication is required after y_k is observed.

The term A_k is the optimal "gain" and has been directly related to the information that the observation y_k has about the process x_k. Letting $H_k = I$, an intuitive interpretation is readily available; when the observation covariance R_k is much higher than the prediction error covariance P_k, the gain A_k is very small and the correction term is not weighted heavily; on the other hand whenever R_k is small, A_k is nearly the identity matrix I, so that the past estimate is not nearly as important as the present observation.

The error covariance matrices P_k and $\overline{P}_k$ have already been identified; P_k is the one-step prediction error covariance; $\overline{P}_k$ is the filter error covariance. Together their equations are the discrete form of the Riccati equation and limit to this equation as the time increment tends to zero.

The significant observation should be made that P_k, $\overline{P}_k$ and A_k do not depend on the values of the conditioning variables, the observations $y_k, y_{k-1}, \ldots, y_1$. This is a special property of the multivariate Gaussian distribution and has the important practical result that the structure of the optimal filter can be determined before and independent of the observations; it will not be "adaptive". In special cases, the gain A_k might even be computed well in advance and then used by the optimal estimate equation as needed. Usually, memory restrictions preclude such massive storage and the method

is to solve the Riccati equation as needed. The structure and memory requirements are clearly evident from this formulation. They have the advantageous property that they are fixed.

The prediction problem has a solution which is a corollary to the theorem.

Corollary

Under the conditions of the theorem, and since u_k has zero mean, the prediction $\hat{x}_{k+j|k}$ is given by

$$\hat{x}_{k+1|k} = \Phi_{k+1} \hat{x}_{k|k} \tag{3.37}$$

$$\hat{x}_{k+2|k} = \Phi_{k+2} \hat{x}_{k+1|k} = \Phi_{k+2} \Phi_{k+1} \hat{x}_{k|k} \tag{3.38}$$

$$\vdots$$

$$\hat{x}_{k+j|k} = \Phi_{k+j} \cdots \Phi_{k+1} \hat{x}_{k|k} . \tag{3.39}$$

This result follows by noting that the conditional estimate is still optimum. With no more observations, the one-step prediction is just iterated. This has the physical meaning of running the autonomous (no input) system starting from the initial condition of $\hat{x}_{k|k}$ at time k to get state values at future times $k+j$.

3.3 Continuous-Time Filter

For comparative purposes, the equations of the continuous-time filter will be stated.

Estimator equation:

$$\frac{d\,\hat{x}(t)}{dt} = F(t)\,\hat{x}(t) + h_0(t,t)\,[y(t) - \hat{x}(t)] \qquad (3.40)$$

Gain equation:

$$h_0(t,t) = P(t)\,H'(t)\,R^{-1}(t) \qquad (3.41)$$

Variance equation:

$$\frac{d\,P(t)}{dt} = R(t)\,P(t) + P(t)\,F'(t)$$

$$- P(t)\,H'(t)\,R^{-1}(t)\,H(t)\,P(t) + Q(t) \qquad (3.42)$$

F(t) is the system coefficient matrix, as defined in Chapter 2. These two forms are quite similar, with the usual difference equation and differential equation distinctions and parallelism. Many of the properties of the discrete equations carry over to the continuous equations. It is possible to derive one set from the other.

These continuous equations are presented by Van Trees (Ref. 22). Since our applications in later chapters are for the discrete case and the computations will be on a digital computer, these equations will not be discussed further.

3.4 Initial Conditions

As in any differential equation or difference equation problem,

initial conditions are required to completely specify the solution; furthermore, the state-variable method is in the time-domain, so actual values of the initial conditions are needed. Much of the material written on Markov processes deals with the transient condition, where some initial distribution is given and the problem is to find suitable transition probabilities and conditional probabilities. Such a study is not the purpose of this section, but rather initial conditions are sought which satisfy steady state constraints (temporally constant). Furthermore, such initial conditions should be expressed in terms of the model parameters; the state transition matrix Φ, the observation matrix H, the driving noise covariance Q and the observation noise covariance R.

The covariance of the process x is denoted by

$$\Lambda_k \triangleq E\{x_k x_k'\} \qquad k \geq 0 \tag{3.43}$$

and its initial condition Λ_0 will be found. The reader will recall that, in the derivation of the recursive equations for $\hat{x}_{1|1}$ and $\hat{x}_{2|2}$, Λ_0 was presumed known. The emphasis here is to give it that value as determined by using Φ and Q; from the process equation, Λ_k satisfies

$$\Lambda_k = \Phi_k \Lambda_{k-1} \Phi_k' + Q_k \tag{3.44}$$

since x_k and u_k are uncorrelated. So for a steady condition,

$$\Lambda_0 = \Phi_0 \Lambda_0 \Phi_0' + Q_0 \,. \tag{3.45}$$

Appendix E shows the conditions for a solution to this equation and that a reasonable approximation can be readily computed. This matrix is needed in the next chapter to establish initial values for some statistics.

The recursive formulation of the Kalman-Bucy filter assumes $P_1 = \Lambda_0$, which is in keeping with the condition that no observations have been taken. Observe that this value of P_1 is independent of R and that this represents a situation of the transient type. This condition is not admitted, for a steady-state value for P_1 is needed, as though the recursion had been running for some time. $P_1 = P_0$ must satisfy the conditions obtained from Eqs. 3.34, 3.35, 3.36.

$$P_0 = \Phi_0 \overline{P}_0 \Phi_0' + Q_0 \tag{3.46}$$

$$\overline{P}_0 = (I - A_0 H_0) P_0 (I - A_0 H_0)' + A_0 R_0 A_0' \tag{3.47}$$

$$A_0 = P_0 H_0' (H_0 P_0 H_0' + R_0)^{-1} \tag{3.48}$$

These equations can be simplified somewhat, because $H_0 = H = I$ and another form of $\overline{P}$ can be written:

$$\overline{P}_0 = (I - A_0 H_0) P_0 \tag{3.49}$$

However, no closed-form solution for P_0 is evident, for this is the

matrix Riccati equation. This difficulty is not insurmountable; these equations are part of the Kalman filter and a solution can be obtained by iterating the equations with Φ_0, Q_0 and R_0 as given. $A_0 = .9\ I$ or some previous value will start the iterations and often only a few are needed.

3.5 Performance Indices

The Kalman-Bucy filter is a filter in the usual sense of the word; this one has an input vector y_k and an output vector $\hat{x}_{k|k}$, which is the minimum mean-square estimate of the process x_k. Also, the "filter" is capable of producing predictions or forecasts of the state variable, $\hat{x}_{k+j|k}$. It is natural to define indices of performance in many studies and in this case particularly, as sums of differences squared.

P_k and $\overline{P}_k$ are computed by the filter itself. P_k is the one-step prediction error covariance matrix for the process x_k, as defined in Eq. 3.22. $\overline{P}_k$ is the closely-related filter error covariance matrix, defined in Eq. 3.13. Some additional performance indices will be defined. Such statistics are time-averages of observable processes; since x_k is not observable, but y_k is, define $\hat{y}_{k|k}$ analogously to $\hat{x}_{k|k}$ and find that

$$\hat{y}_{k|k} = H_k \hat{x}_{k|k} \tag{3.50}$$

$$\hat{y}_{k+j|k} = H_{k+j} \hat{x}_{k+j|k} \ . \tag{3.51}$$

The performance indices are therefore defined as the time-averages of the mean-square error of the N-th forecast.

$$J0_k = \frac{1}{M} \sum_{j=k-1-M}^{k} (y_j - \hat{y}_{j|j})^2 \tag{3.52}$$

$$J1_k = \frac{1}{M} \sum_{j=k-1-M}^{k} (y_{j=1} - y_{j+1|j})^2 \tag{3.53}$$

$$JN_k = \frac{1}{M} \sum_{j=k-1-M}^{k} (y_{j+N} - \hat{y}_{j+N|j})^2 \tag{3.54}$$

N should be kept small, where the forecasts have meaning. The point should also be made that these statistics are not used in the model, so they are viable as comparisons of different models and initial conditions.

CHAPTER 4

MODEL PARAMETER ESTIMATION

Synopsis

4.1 Introduction

4.2 Consistent Estimates for the Stationary Model

Model A: Φ, Q, R

4.3 Tracking Estimates for the Nonstationary Model

Model B: Φ, Q, R

Coefficients

4.4 Filtering of Field Data

Model C

Preprocessing

4.5 Initial Conditions on Statistics

$E\{y_3 y_1'\}$

$E\{y_2 y_1'\}$

$B_{0,0}$ and $B_{1,0}$

4.6 Summary of Recursive Equations of Model C

Block diagram

4.1 Introduction

This chapter contains a derivation of the equations to estimate the parameters of the model. These estimates are then used in the model to determine the conditional mean and conditional covariance of the process. Consistent estimates in the stationary case are modified to become tracking estimates of the slowly-varying parameters. In addition, some initial "nonlinear filtering" and block-averaging are introduced to smooth the data. A complete block diagram is given for the recursive computation of the model parameters.

The model equations, given in Chapter 2, are summarized here for convenience.

$$\text{Process:} \qquad x_k = \Phi_k x_{k-1} + u_k \tag{4.1}$$

$$\text{Observation:} \quad y_k = H_k x_k + v_k \tag{4.2}$$

where

u_k is distributed as $N(0, Q_k)$; u_k independent, and

v_k is distributed as $N(0, R_k)$; v_k independent.

Both the u, v statistics may be time-varying and it is assumed that u and v are statistically independent. The parameters Φ_k, Q_k and R_k are unknown and will be estimated.

This is a good place to discuss the "estimate-and-plug"

technique. Generally speaking, a large school of thought contends that when a parameter is unknown, it should be estimated in some reasonable manner and then used (with caution) as though the parameter were known precisely. On the other hand, the Bayesian school holds that parameters are either precisely known, or if unknown, the parameter is a random variable or random process. Hence the parameter has a distribution which a Bayesian must formulate and use. The parameter is "learned" or "estimated" in the sense that observations concerning the parameter are used to sharpen the knowledge through probability functions of smaller and smaller variance. Prior information in the a priori distribution is updated with a new observation to obtain a new distribution called the a posteriori distribution. Such studies of recent origin include the works of Spooner (Ref. 16), Birdsall (Ref. 2) and Jaarsma (Ref. 8).

The Bayesian philosophy is a tough task-master. It demands a complete statistical description of the a priori information (though the results may not be very sensitive to the a priori distribution) and it requires the use of "Bayes' Law"* to obtain the a posteriori pdf. Such manipulations are often not mathematically tractable and, in the author's opinion, usually not practically feasible. Often, the required observations cannot be made or the computations are too complex or

---

*Bayes' Law is the historical name for the accepted definition of conditional probability, given in Chapter 2.

too time-consuming. The required computations have not been developed for lack of sufficiently well-defined stochastic models. For these and other reasons which will be given in this chapter, the approach adopted here is to estimate the parameters in a simple and direct manner. Some properties of these estimators will be studied. Later chapters will demonstrate their use with field data and randomly-generated data as input.

4.2 Consistent Estimates for the Stationary Model

For the special case of a stationary model (in which $\Phi_k = \Phi$, $Q_k = Q$ and $R_k = R$ do not depend on time k), the work of Anderson, Kleindorfer, Kleindorfer and Woodroffe (Ref. 1) (referred to as AKKW in the following text) has produced some interesting estimators. They are consistent* estimators when the system is stable and stationary. The algorithms are easily implemented as a special purpose computer or on a programmed computer. Their development follows.

Model A

Consider the proposition:

$$E\{y_k\, y'_{k-2}\} = \Phi\, E\{y_{k-1}\, y'_{k-2}\} \qquad k \geq 3 \tag{4.3}$$

---

*See Appendix A for definitions of underlined terms.

From the model equations, for the completely observable case $H = I$,

$$E\{y_k\, y'_{k-2}\} = E\{(x_k + v_k)\,(x'_{k-2} + v'_{k-2})\} \tag{4.4}$$

$$= E\{x_k\, x'_{k-2}\} + E\{x_k\, v'_{k-2}\}$$

$$+ E\{v_k\, x'_{k-2}\} + E\{v_k\, v'_{k-2}\}$$

Under the conditions of the model, x_i and v_j are orthogonal for every pair i, j, so the second and third terms are 0. Also, v_k and v_{k-2} are orthogonal.

$$E\{y_k\, y'_{k-2}\} = E\{x_k\, x'_{k-2}\} \tag{4.5}$$

$$\Phi\, E\{y_{k-1}\, y'_{k-2}\} = \Phi\, E\{(x_{k-1} + v_{k-1})\,(x'_{k-2} + v'_{k-2})\} \tag{4.6}$$

$$= E\{(\Phi\, x_{k-1} + \Phi\, v_{k-1})\,(x'_{k-2} + v'_{k-2})\} \tag{4.7}$$

$$= E\{(x_k - u_{k-1} + \Phi\, v_{k-1})\,(x'_{k-2} + v'_{k-2})\} \tag{4.8}$$

$$= E\{x_k\, x'_{k-2}\} + E\{x_k\, v'_{k-2}\}$$

$$- E\{u_{k-1}\, x'_{k-2}\} - E\{u_k\, v'_{k-2}\}$$

$$+ E\{\Phi\, v_{k-1}\, x'_{k-2}\} + E\{\Phi\, v_{k-1}\, v'_{k-2}\}$$

All terms but the first are 0 by orthogonality.

$$\Phi\, E\{y_{k-1}\, y'_{k-2}\} = E\{x_k\, x'_{k-2}\} \tag{4.9}$$

At this point the AKKW study estimated the terms in Eq. 4.3 as follows.

$$E\{y_k\, y'_{k-2}\} \doteq g_k \sum_{j=3}^{k} y_j\, y'_{j-2} \tag{4.10}$$

$$E\{y_{k-1}\, y'_{k-2}\} \doteq g_k \sum_{j=3}^{k} y_{j-1}\, y'_{j-2} \tag{4.11}$$

(The scalar function g_k is not relevant at this point.) Then the estimate for Φ after k observations is denoted by $\hat{\Phi}_k$ and is given by

$$\hat{\Phi}_k = \left(\sum_{j=3}^{k} y_j\, y'_{j-2}\right)\left(\sum_{j=3}^{k} y_{j-1}\, y'_{j-2}\right)^{+} \triangleq S_{02}\, S_{12}^{+} \qquad k \geq 3 \tag{4.12}$$

where + indicates the pseudo-inverse.*

Let the spectral radius of Φ be denoted by $\rho(\Phi)$. The following theorem is given without proof.

Theorem 4.1. If $\rho(\Phi) < 1$ and if Φ and Q are nonsingular, then $\hat{\Phi}_k$ is a strongly consistent estimate of Φ.

To develop estimates for Q and R for the same model, consider

---

*See Appendix A for definitions of underlined terms.

$$A_0 = E\{(y_k - \Phi y_{k-1})(y_k - \Phi y_{k-1})'\} \tag{4.13}$$

and

$$A_1 = E\{(y_k - \Phi y_{k-1})(y_{k-1} - \Phi y_{k-2})'\} \tag{4.14}$$

In the manner illustrated above, it can be shown that

$$A_0 = Q + R + \Phi R \Phi' \tag{4.15}$$

$$A_1 = -\Phi R . \tag{4.16}$$

Since Φ is non-singular, R and Q are uniquely determined.

$$R = -\Phi^{-1} A_1 \tag{4.17}$$

$$Q = A_0 - R - \Phi R \Phi' \tag{4.18a}$$

$$= A_0 + \Phi^{-1} A_1 + A_1 \Phi' \tag{4.18b}$$

Observe that A_0 is positive definite from Eq. 4.13 and that A_1 should be negative definite from Eq. 4.17.

Theorem 4.2. If Φ and Q are nonsingular and if $\hat{\Phi}_k$ is given by Eq. 4.12, then

$$A_{0,k} = \frac{1}{k} \sum_{j=3}^{k} (y_j - \hat{\Phi}_j y_{j-1})(y_j - \hat{\Phi}_j y_{j-1})' \tag{4.19}$$

$$A_{1,k} = \frac{1}{k} \sum_{j=3}^{k} (y_j - \hat{\Phi}_j y_{j-1})(y_{j-1} - \hat{\Phi}_j y_{j-2})' \tag{4.20}$$

are strongly consistent estimates of A_0, A_1 respectively. Observe that in Eqs. 4.19 and 4.20 only known values of $\hat{\Phi}_j$ are required at time k.

No proof will be given here. A sufficiently detailed proof is given in AKWW (Ref. 1). From these two theorems, these estimates for Φ, Q, R are strongly consistent.

$$\hat{\Phi}_k = \left(\sum_{j=3}^{k} y_j \, y'_{j-2} \right) \left(\sum_{j=3}^{k} y_{j-1} \, y'_{j-2} \right)^{+} \tag{4.12}$$

$$R_k = -\Phi_k^{-1} A_{1,k} \tag{4.21}$$

$$Q_k = A_{0,k} - R_k - \Phi_k R_k \Phi_k' \tag{4.22a}$$

$$= A_{0,k} + \Phi_k^{-1} A_{1,k} + A_{1,k} \Phi_k' \tag{4.22b}$$

4.3 Tracking Estimates for the Nonstationary Model

The preceding section shows the general approach to the problem, but in this section the condition of stationarity is removed. However, a general principle adopted throughout the remainder of the chapter is that time-variations of the model parameters Φ, Q, R are slow relative to the observation rate of the process. Hence, estimators similar to the stationary case will have some applicability. The forms of these estimators and the reasoning behind them will be given below.

Model B

Consider the proposition:

$$\text{a)} \qquad E(y_k\, y'_{k-2}) = E(\Phi_k\, y_{k-1}\, y'_{k-2}) \tag{4.23}$$

$$\text{b)} \qquad E(\Phi_k\, y_{k-1}\, y'_{k-2}) \doteq \Phi_k\, E(y_{k-1}\, y'_{k-2}) \tag{4.24}$$

For the stationary model, these equations reduce to Eq. 4.3. To prove Eq. 4.23, the previous proof of Eq. 4.3 with $\hat{\Phi}_k$ for Φ is valid. To justify Eq. 4.24, if Φ_k is a known function, the result is evident and exact. When Φ_k is a random process, the approximation is "good" if Φ_k and $y_{k-1}\, y'_{k-2}$ are independent and an unbiased estimate of Φ_k is used. Both of these conditions will be satisfied by the "tracking" estimator for Φ_k.

The following modifications to Eqs. 4.10 and 4.11 seem justified in the case when the expectations are time-dependent.

$$E\{y_k\, y'_{k-2}\} \doteq g_k \sum_{j=3}^{k} c_{k-j}\, y_j\, y'_{j-2} \tag{4.25}$$

$$E\{y_{k-1}\, y'_{k-2}\} \doteq g_k \sum_{j=3}^{k} c_{k-j}\, y_{j-1}\, y'_{j-2} \tag{4.26}$$

where

$$g_k = \left(\sum_{j=3}^{k} c_{k-j} \right)^{-1}$$

The sum is required to converge for $k \to \infty$, so $c_{k-j} \to 0$ as $j \to 3$. For example, let $r_\phi < 1$ and $c_{k-j} = r_\phi^{k-j}$. With a proper choice of r_ϕ, the estimates in Eqs. 4.25 and 4.26 will follow time-variations in the statistics. Therefore,

$$\hat{\Phi}_k = \left(\sum_{j=3}^{k} c_{k-j}\, y_j\, y'_{j-2} \right) \left(\sum_{j=3}^{k} c_{k-j}\, y_{j-1}\, y'_{j-2} \right)^{+} \tag{4.27}$$

will be called a tracking estimator; compare to the estimate in Eq. 4.12.

Similar modifications of A_0, A_1 in Eqs. 4.13 and 4.14 and their estimators will also be made. Since Φ is hypothesized as being a slowly-varying random process, consider

$$B_0 = E\{(y_k - \hat{\Phi}_k\, y_{k-1})\,(y_k - \hat{\Phi}_k\, y_{k-1})'\} \tag{4.28}$$

$$B_1 = E\{(y_k - \hat{\Phi}_k\, y_{k-1})\,(y_{k-1} - \hat{\Phi}_{k-1}\, y_{k-2})'\} \tag{4.29}$$

Using the model Eqs. 4.1 and 4.2 with $H = I$,

$$y_k - \hat{\Phi}_k\, y_{k-1} = u_k + v_k - \hat{\Phi}_k\, v_{k-1} \tag{4.30}$$

$$y_{k-1} - \hat{\Phi}_{k-1}\, y_{k-2} = u_{k-1} + v_{k-1} - \hat{\Phi}_{k-1}\, v_{k-2} \tag{4.31}$$

Consequently, because all cross-terms are approximately 0,

(Appendix B for terms with Φ)

$$B_0 = Q_k + R_k + E\{\hat{\Phi}_k \, v_{k-1} \, v'_{k-1} \, \hat{\Phi}_k'\}$$

$$\doteq Q_k + R_k + \hat{\Phi}_k \, R_{k-1} \, \hat{\Phi}_k' \qquad (4.32)$$

where the approximate equality holds when $\hat{\Phi}_k$ is a good estimate, unbiased, independent of v_k. Similarly,

$$B_1 = E\{-\Phi_k \, v_{k-1} \, v'_{k-1}\} \doteq -\hat{\Phi}_k \, R_{k-1} \, . \qquad (4.33)$$

The dependence of B_0, B_1 on time is explicitly displayed to contrast with Eqs. 4.15 and 4.16 for A_0 and A_1.

Tracking estimates for B_0, B_1 are now given.

$$B_{0,k} \doteq b_{0,k} \sum_{j=3}^{k} d_{k-j} (y_j - \hat{\Phi}_j \, y_{j-1}) (y_j - \hat{\Phi}_j \, y_{j-1})' \qquad (4.34)$$

$$B_{1,k} \doteq b_{1,k} \sum_{j=3}^{k} e_{k-j} (y_j - \hat{\Phi}_j \, y_{j-1}) (y_{j-1} - \hat{\Phi}_{j-1} \, y_{j-2})' \qquad (4.35)$$

where $b_{0,k}$, $b_{1,k}$, d_{k-j} and e_{k-j} are scalars and where $b_{0,k}^{-1} = \sum_{j=3}^{k} d_{k-j}$ and $b_{1,k}^{-1} = \sum_{j=3}^{k} e_{k-j}$ are required to converge. Required conditions are that $d_{k-j} \to 0$ and $e_{k-j} \to 0$ as $j \to 3$. For example, as above, let $r_0 < 1$, $d_{k-j} = r_0^{k-j}$ and $r_1 < 1$, $e_{k-j} = r_1^{k-j}$.

Since $\hat{\Phi}_k$ is an approximation to a state transition matrix, it should be nonsingular, and $\hat{R}_{k-1}$ is uniquely determined.

$$\hat{R}_{k-1} = -\hat{\Phi}_k^{-1} \, B_{1,k} \qquad (4.36)$$

One more assumption is necessary to get $\hat{Q}_k$. Suppose the matrix $\hat{R}_k$ varies slowly with k; as a first approximation, let $\hat{R}_k = \hat{R}_{k-1}$. Then

$$\hat{Q}_k = B_{0,k} - \hat{R}_{k-1} - \hat{\Phi}_k \hat{R}_{k-1} \hat{\Phi}_k' \tag{4.37a}$$

$$\hat{Q}_k = B_{0,k} + \hat{\Phi}_k^{-1} B_{1,k} + B_{1,k} \hat{\Phi}_k' \tag{4.37b}$$

To conclude this section, a brief discussion of these tracking estimators seems in order. No rigorous proofs of the following properties are offered but their plausibility will be illustrated. For the stationary random process, the estimators given by AKKW (Ref. 1) are strongly consistent; in general terms, when enough observations are taken, the estimates get close to their true value. This desirable feature occurs at a practical value of the number of observations; several hundred observations were enough in the AKKW study. Chapter 7 will have results for the time-varying case. When the model is allowed to be nonstationary through time-varying Φ, Q and R, the above changes in the estimators make it possible for these parameters to be tracked. The effect of the coefficients $\{c_{k-j}\}$, $\{d_{k-j}\}$ and $\{e_{k-j}\}$ is to deemphasize past data by giving it less weight in the sums.

These tracking estimators were tested by using field data. Since such field data will not satisfy <u>all</u> the conditions of the model <u>all</u> the time (discussion below), additional processing is introduced.

4.4 Filtering of Field Data

The process y_k as observed from a physical source perhaps does not satisfy the conditions of the models. For example, for some short observation interval, the following conditions might prevail.

a) Multiplicative noise is present.

$$x_{k+1} = \Phi_{k+1} x_k m_k + u_{k+1} \tag{4.38}$$

where m_k is hypothesized to have a mean of 1, small variance but unknown distribution.

b) Additive impulse noise is present.

$$x_{k+1} = \Phi_{k+1} x_k + u_{k+1} + p_{k+1} \tag{4.39}$$

where p_{k+1} is impulse noise which is unknown but significant. The effect may be an apparent rapid change in Q, which the models cannot follow.

Under such conditions, the models developed above will not be true. However, by their nature, the models should follow gross characteristics of the process and should be expected to smooth out such transients. Such transient conditions were not intended to be modeled here; to help eliminate them, pre-processing is introduced in the form of a fixed linear filter and a limiting-type, (possibly) time-varying nonlinear processor.

A known, fixed linear filter is acceptable at this point in the development of the model. The output process of such a filter will be modeled by the models of this chapter; by linearity, the statistics of the input process can be determined from the statistics of the output.

The form of the filter is chosen for ease of implementation on a digital computer. The parameters of the filter will be selected by experiment with the field data and then fixed.

The particular filter is the so-called Block Average filter (BAV).

$$\frac{1}{M} \sum_{j=1+k-M}^{k} r_j \qquad k \geq M$$

where r_j is the input process (received, observed). The output process y_k is usually sampled at some rate $1/S$, or one output value for every S input values. Hence

$$y_k = \frac{1}{M} \sum_{j=kS+1-M}^{kS} r_j \qquad \begin{matrix} kS \geq M \\ k = 1, 2, \ldots \end{matrix} \qquad (4.40)$$

j = discrete time parameter of the observed process r_j

k = discrete time parameter of model

y_k = observed process for the model.

The frequency-domain characterization of the BAV filter is

given in Chapter 5. The important facts are that it is a low-pass filter with a 3-db bandwidth proportional to $\frac{1}{M}$. The selection of the parameters M and S is left for Chapter 5.

The need for a nonlinear processor of the limiting type is pointed out by an analysis of real data, which is done in Chapter 7. The general form of the processor is given here to include it with the development of the model. The form is dictated by the condition b) above: additive impulse noise. To reduce the effect of such observations, a limiter was first considered. However, preliminary tests showed that the model then depended strongly on the choice of the limiting value, which is undesirable. The following processor was considered next.

$$r_j = \begin{cases} \hat{r}_j & \hat{r}_j \in \{\text{L-sphere about } \bar{r}\} = \mathcal{A} \\ \Phi_L \hat{r}_{j-1} & \text{otherwise} \end{cases} \tag{4.41}$$

where

$\hat{r}_j$ = observed process

$\bar{r}$ = mean of observed process

Φ_L = fixed state transition matrix

L = limiting value.

Φ_L and L must be selected jointly. The region $\mathcal{A}$ is the L-sphere about $\bar{r}$; by selecting L properly and fixing it, "bad" observations

can be eliminated. If the process can be modeled by Model B, all but a few observations will be in $\mathcal{A}$; the probability measure of the set $\mathcal{A}$ is very close to 1. The parameters $\bar{r}$, Φ_L and L of the fixed nonlinear processor are selected by experiment in Chapter 7. The effect of the processor on data is observed by counting the instances that $\hat{r}_j \notin R$ to verify that the postulated conditions are satisfied.

A modification of this processor was also studied. $\hat{x}_{k|k}$ is the MMSE (minimum mean square estimate) of x_k, which is the output of the Kalman-Bucy filter. Take

$$r_j = \hat{x}_{k|k} \quad \text{for} \quad \hat{r}_j \notin \mathcal{A} \qquad k < j \leq k+1 \qquad (4.41)$$

The combined processing of the nonlinear processor, the BAV filter and Model B will be called Model C. The above notations suggest the following order of operations.



Fig. 4.1. Model C

4.5 Initial Conditions on Statistics

Initial estimators for Φ, Q and R should reflect any prior information about the parameters at the starting time of the observations.

Since the subsequent estimates are unbiased, it seems reasonable to start with (a priori) expected values for initial conditions on the four summations used by the estimators. These initial conditions can be calculated by using the stationary condition. From Eq. 4.5,

$$\begin{aligned} E\{y_k\, y'_{k-2}\} &= E\{x_k\, x'_{k-2}\} \\ &= E\{\Phi_k\, \Phi_{k-1}\, x_{k-2}\, x'_{k-2}\} + E\{(\Phi_k\, u_{k-1} + u_k)\, x'_{k-2}\} \\ &= \Phi_k\, \Phi_{k-1}\, \Lambda_{k-2} \end{aligned} \tag{4.42}$$

where Λ is the variance of the process x. This result follows from the condition that past values of x_k are not correlated with future values of u_k. Then, initially (at $k = 2$ in this notation) and with $\Phi_2 = \Phi_1 = \Phi_0$, and $\Lambda_2 = \Lambda_1 = \Lambda_0$,

$$E\{y_2\, y_0'\} = \Phi_0\, \Phi_0\, \Lambda_0 . \tag{4.43}$$

Repeating Eq. 4.9,

$$\Phi\, E\{y_{k-1}\, y'_{k-2}\} = E\{x_k\, x'_{k-2}\} . \tag{4.9}$$

Therefore, initially

$$E\{y_1\, y_0'\} = \Phi_0\, \Lambda_0 . \tag{4.44}$$

These are the initial conditions for the sums in the Φ estimate. The initial values for B_0 and B_1 are given by Eqs. 4.32 and 4.33

with $k = 2$ and $R_1 = R_2$.

$$B_{0,2} = Q_2 + R_2 + \Phi_0 R_2 \Phi_0' \tag{4.45}$$

$$B_{1,2} = -\Phi_0 R_2 \tag{4.46}$$

These are the initial conditions for the sums in the estimation of Q and R. Altogether, the following initial values are needed:

$$\Lambda_0 \quad \text{and} \quad \Phi_2, Q_2, R_2$$

The initial value Λ_0 for the process covariance was discussed in Chapter 3, Section 3.4. The initial conditions on the other statistics have been expressed in terms of Φ_2, Q_2, R_2, which represent the initial model. These three matrices must be given starting values by using reasonable physical conditions.

Again, similar comments apply to these initial conditions as were made regarding the model initial conditions. As k becomes large, the sensitivity to these initial values is decreased. For this reason, in AKKW no such initial conditions were assumed in their Monte Carlo studies, yet convergence was moderately rapid. They assumed values of 0 for all the statistics at $k = 2$.

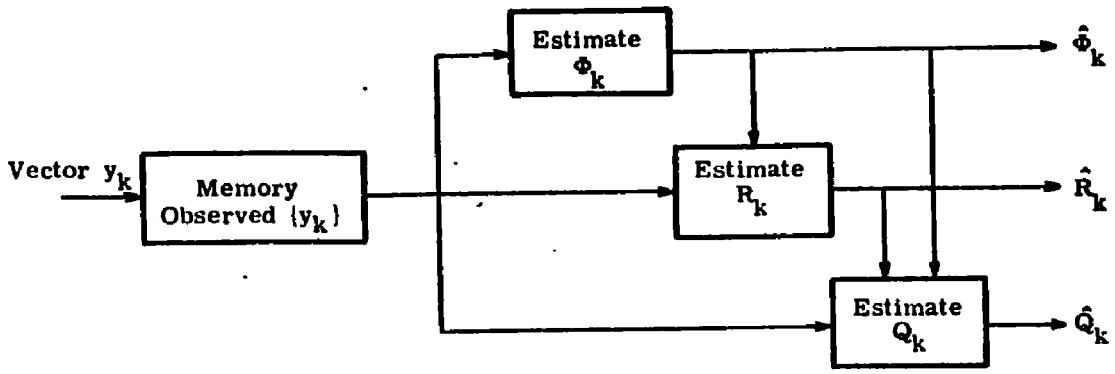
4.6 Summary of Recursive Equations of Model C

This chapter is summarized by a block diagram of the parameter estimation for the model. Model C consists of the

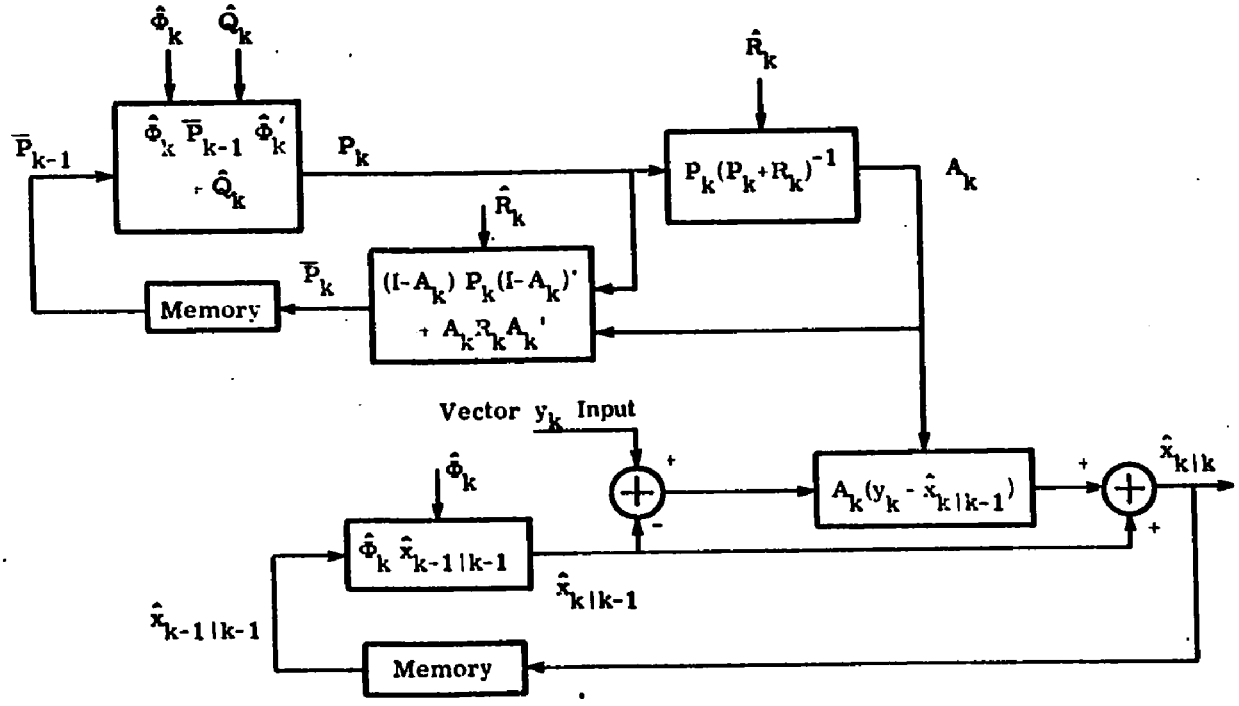
preprocessor which is nonlinear, the linear Block Average and the estimators of Model B. The nonlinear processor has the purpose of eliminating abnormal data and is characterized by Eqs. 4.40 and 4.41. The block averaging performs a slight amount of smoothing of the data and is expressed by Eq. 4.39.

Recall that $H = I$ in this model. $\hat{\Phi}_k$ is a tracking estimator for Φ_k, the state transition matrix of the model; it is estimated by Eq. 4.27. The driving noise covariance Q_k is estimated by $\hat{Q}_k$ in Eqs. 4.37a and 4.37b. The observation noise covariance R_k is estimated by $\hat{R}_{k-1}$ in Eq. 4.36.

The estimators and the Kalman-Bucy equations are shown in block diagram form in Fig. 4.2.



a)



b)

Fig. 4.2. a) Estimation of Φ_k, Q_k, R_k
b) Kalman filter block diagram, H = I, n x n

CHAPTER 5

A CLASS OF LINEAR FILTERS

Synopsis

5.1 Introduction

Analog and digital filters

5.2 A Standard Form of a Class of Analog Filters

Impulse response

Transfer function

τ-diagram

Special cases

General case: complex coefficients

5.3 A Standard Form of a Class of Digital Processors

Sampling

Equivalent digital processor

Block diagram

Examples

5.4 Combined Operation

5.5 Time-Domain Viewpoint

5.1 Introduction

This chapter will discuss and illustrate a standard form of filtering or processing of data. Generally speaking, we talk of filtering when analog signals are being used and we say processing when digital signals are involved; however, there is no universally accepted demarkation. There is a common nature in the several forms of digital processors in use, so that a single analysis will include them all.

The procedure will be to study a general form of the analog filter by describing its impulse response and amplitude spectrum. We will demonstrate that this form has a simple digital processor equivalent. Some specific types of processors will be illustrated and a cascade of such processors will have meaning.

A time-domain analysis will be offered for the standard form of the digital processor. We comment on the distinguishing and complementary features of the time-domain and frequency-domain analysis.

Historically speaking, such filters have been applied recently to the processing of acoustic signals (Refs. 7, 19 and 20). The derivations of the filter characteristics have taken many forms, often repeated and with unnecessary assumptions. A goal of the following material is a unified presentation.

5.2 A Standard Form of a Class of Analog Filters

5.2.1 Impulse Response. A particular class of analog filters

will be studied, namely that having the impulse response

$$h(\tau) = \frac{1}{M} \sum_{k=0}^{M-1} r^{M-1-k} \delta(\tau - kP\Delta) . \qquad (5.1)$$

A brief discussion of the origin of this function will be useful. The concept and physical implementation of a delay line* are quite familiar to an engineer. In the continuous time-domain, such pure delay is often modeled by the linear filter with the impulse response

$$h_1(\tau) = \delta(\tau - T) , \qquad (5.2a)$$

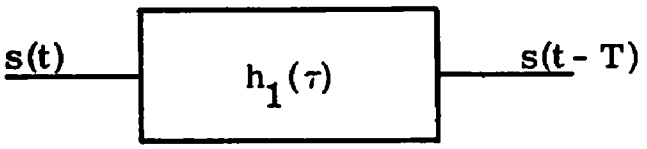
where T is the delay. For any input signal s(t), the output is the convolution

$$s(t) * \delta(t - T) = s(t - T) \qquad (5.2b)$$

The convolution operator * is well-known in the literature. The frequency-domain description of the pure delay is $H_1(f) = e^{-j2\pi fT}$, which has modulus 1 and causes a linear phase change of -T radians/sec. Figure 5.1 summarizes these properties.

The next step is to use delay lines of lengths which are multiples of T. Obviously, we can consider this to be a tapped delay line, and let the number of sections be M-1. It is important to recognize that the final result may not be in this form, but that the

*Underlined terms are defined in Appendix A.

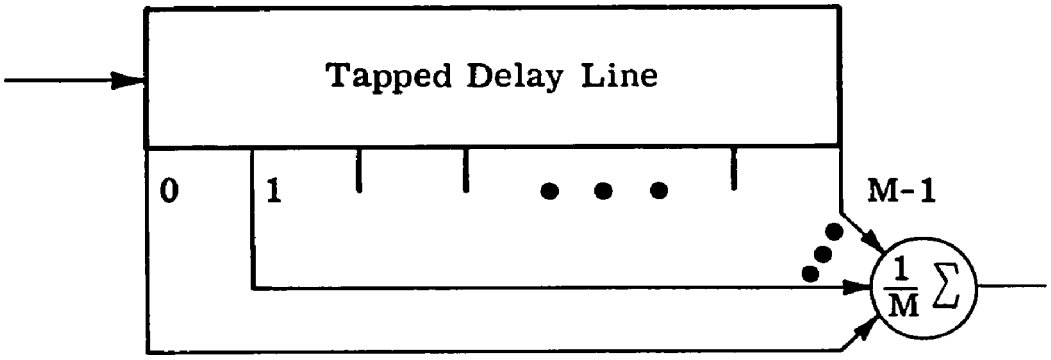


$$h_1(\tau) = \delta(\tau - T)$$

$$\updownarrow$$

$$H_1(f) = e^{-j2\pi fT}$$

Fig. 5.1. A delay line of T units



$$h_2(\tau) = \frac{1}{M} \sum_{k=0}^{M-1} \delta(\tau - kT)$$

Fig. 5.2. A tapped delay line filter

analysis is linear and is simplified by this approach. The model is illustrated in Fig. 5.2. The input point is one of the taps so that one of the sequence of inputs has zero delay. Two aspects of this filter should be noted. It is input-output stable because the latest input (at the left) enters as the earliest input exits and is forgotten (at the right). Secondly, it is linear, since the delay operator and summation are linear. The impulse response is evident.

$$h_2(\tau) = \frac{1}{M} \sum_{k=0}^{M-1} \delta(\tau - kT) \tag{5.3}$$

The generalization to Eq. 5.1 results by allowing the coefficients r^{M-1-k} and letting $T = P\Delta$; r will determine the particular filter and the discussion of $P\Delta$ will be deferred until later.

5.2.2 Transfer Function. The Fourier transform of $h(\tau)$ is readily computed.

$$H(f) = \frac{1}{M} \sum_{k=0}^{M-1} r^{M-1-k} \mathcal{F}\{\delta(\tau - kP\Delta)\}$$

$$= \frac{r^{M-1}}{M} \cdot \sum_{k=0}^{M-1} r^{-k} e^{-j2\pi fkP\Delta} \tag{5.4}$$

A geometric progression to M terms has the sum

$$1 + \lambda + \lambda^2 + \ldots + \lambda^{M-1} = \begin{cases} \dfrac{1 - \lambda^M}{1 - \lambda} & \lambda \neq 1 \\ M & \lambda = 1 \end{cases} \tag{5.5}$$

λ may be complex; here $\lambda = r^{-1} e^{-j2\pi fT}$.

$$H(f) = \frac{r^{M-1}}{M} \frac{\left(1 - r^{-M} e^{-j2M\pi fT}\right)}{\left(1 - r^{-1} e^{-j2\pi fT}\right)} \tag{5.6}$$

Four general cases can now be treated; $r = \pm 1$, $r = \pm j$. The case $r = \pm 1$ is the simplest case, for which the results are well known. When these four cases have been treated, an extension to the whole complex r plane will be evident and presented.

5.2.3 The τ-Diagram. The impulse response and its Fourier Transform are the usual representations of linear filters. For the class of Eq. 5.1, a third useful representation has been introduced. Here it will be called the τ-diagram and is the impulse train representation on the τ-time axis of the impulse response; it is the time-reverse of the operation of the filter on incoming signals. The general form for any r has the form shown in Fig. 5.3 with the label r. Several cases for specific values of r are also shown.

A merit of the τ-diagram is its pictorial display of the filter. $h(\tau)$ is just the sum of these components. The diagram also shows that when r is complex, a complex filter results; it should be regarded as a 2-channel device producing two real outputs for one real

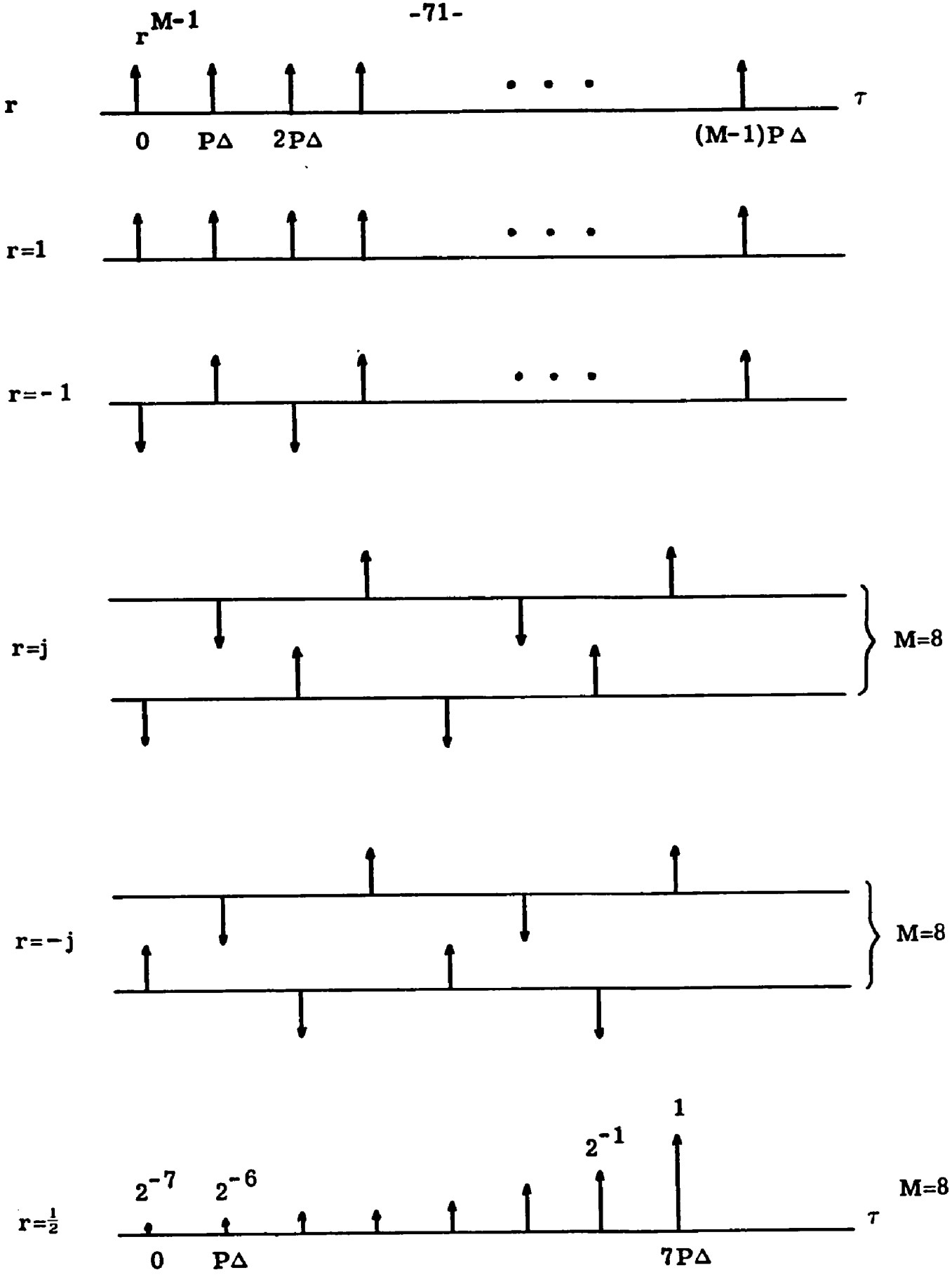


Fig. 5.3 τ - diagrams

input. The input is the real signal $s(t)$; the output is the complex signal

$$z(t) = x(t) + j\, y(t)$$

$$= s(t) * h_R(t) + j\, s(t) * h_I(t) \qquad (5.7)$$

Another point made by the τ-diagram is that, for $r = -1$, M must be even and for imaginary r, M must be divisible by 4. Otherwise, the results are apparently not useful.*

5.2.4 Special Cases: $r = \pm 1$, $r = \pm j$. For this case, a general result can be maintained because $r = e^{j2\pi\alpha}$ for $\alpha = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$.

$$H(f) = \frac{e^{j2\pi(M-1)\alpha}}{M} \frac{\left(1 - e^{-j2\pi M\alpha}\, e^{-j2M\pi fT}\right)}{\left(1 - e^{-j2\pi\alpha}\, e^{-j2\pi fT}\right)}$$

$$= \frac{e^{j\pi M\alpha - j2\pi\alpha}}{M\, e^{-j\pi\alpha}} \cdot \frac{e^{-jM\pi fT}}{e^{-j\pi fT}} \frac{\sin(M\pi fT + M\pi\alpha)}{\sin(\pi fT + \pi\alpha)}$$

$$= e^{j\pi(M-1)\alpha - j\left(\frac{M-1}{2}\right)2\pi fT} \operatorname{sinc}\left(MT\left(f + \frac{\alpha}{T}\right); M\right) \qquad (5.8)$$

Some of the steps should be explained; first factor out the terms in α; then take out each of the second factors; the sine terms result.

The sinc function is defined by

$$\operatorname{sinc} x = \frac{\sin \pi x}{\pi x} \qquad (5.9)$$

---

*This is an oversimplification; it will be clarified below.

A function which has been called the periodic analog of sinc x has also been defined.

$$\text{sinc}(x;M) = \frac{\sin(\pi x)}{M \sin\left(\frac{\pi x}{M}\right)} \tag{5.10}$$

We identify the phase shift term in H(f), which corresponds to an average delay of $\frac{(M-1)}{2}$ T and a phase shift of $(M-1)\pi\alpha$. The sinc function with parameter M is real and is the gain term.

$$|H(f)| = \left|\text{sinc}\left(T\left(f + \frac{\alpha}{T}\right);M\right)\right| \tag{5.11}$$

In this form, the frequency shift is evident for different α. The amplitude spectrum will be found for the case $\alpha = 0$. Then the four possible filters will be identified.

Amplitude Spectrum

The amplitude spectrum is the plot of |H(f)| against frequency f. The properties of the periodic sinc function follow from the zeros of the two sine functions involved. A value of 1 occurs at f values for which both terms have zeros; at $f = 0, \pm \frac{1}{P\Delta} \ldots$. Zeros occur at $f = \pm\frac{1}{MP\Delta}, \pm\frac{2}{MP\Delta}, \ldots \pm\frac{m}{MP\Delta}$ and in general for m not a multiple of M.

With this general approach, a single plot of |H(f)| with the parameters M, P, r and Δ will suffice for all the filters considered below. Figure 5.4 illustrates this function for different scales

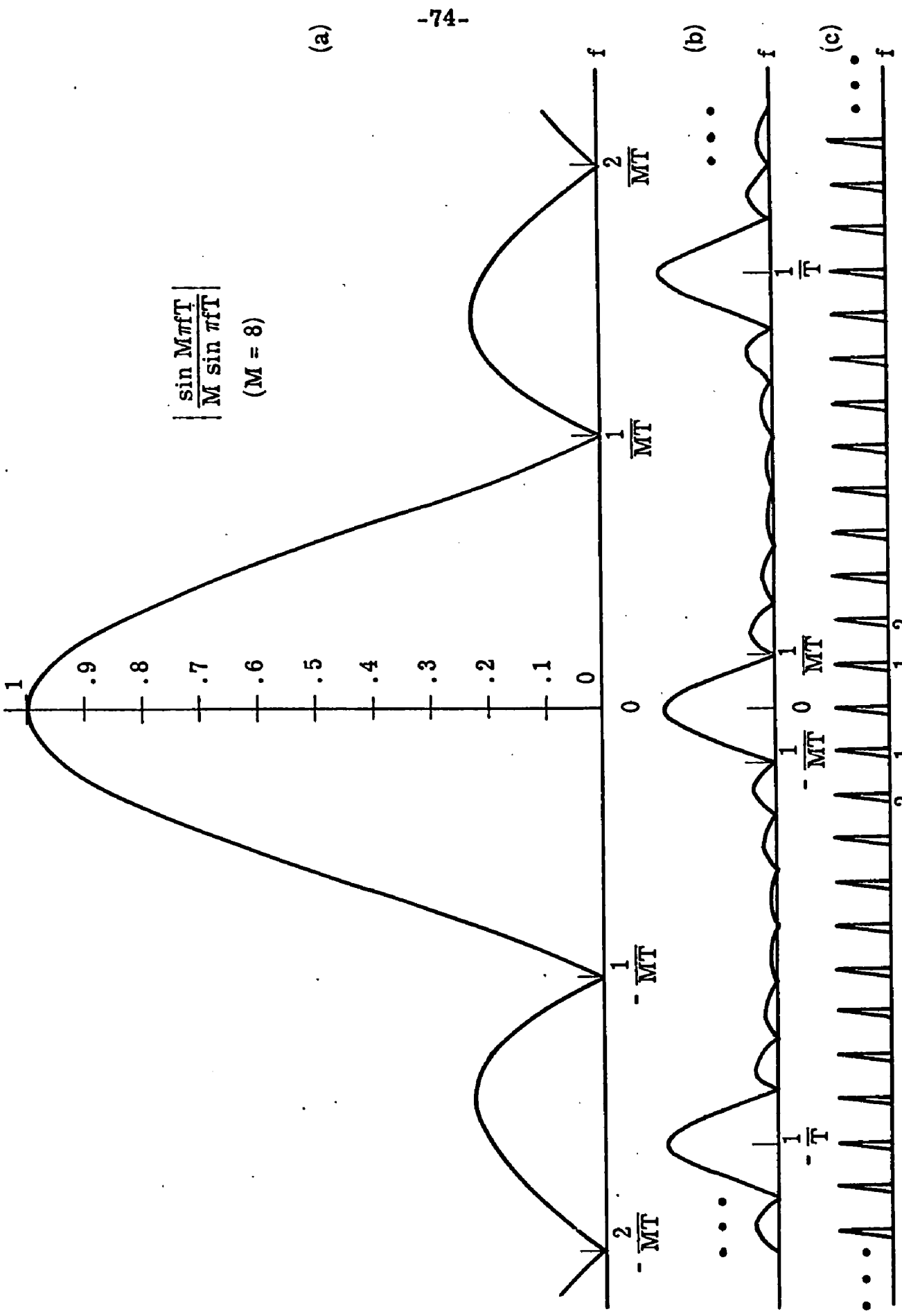


Fig. 5.4. Amplitude spectra $|H(f)|$ for $r = 1$

on the f-axis; recall that this is the standard form with $r = 1$ or $\alpha = 0$.

Observe that the maximum between the zeros at $\frac{1}{MP\Delta}$ and at $\frac{2}{MP\Delta}$ is much smaller than 1. Just for the record, this maximum occurs at approximately $f = \frac{1.43}{MP\Delta}$ for all $M > 4$. The maximum has the value $\frac{-\sin \pi 1.43}{\pi 1.43} = .218$, approximately for all $M > 4$. These results follow directly from

$$\lim_{M \to \infty} \operatorname{sinc}(x;M) = \operatorname{sinc} x \qquad (5.12)$$

and are significant in that they demonstrate that the relative shape of the response is about the same for all $M > 4$ for this case $|r| = 1$.

The power spectrum $|H(f)|^2$ has the same peak and zero locations of course, but the shape of the peaks is that of $\operatorname{sinc}^2(x;M)$. The first maximum is down 13 db relative to the peak.

Comb Filter or CIRAV

The case $r = 1$ or $\alpha = 0$ is studied above. From Fig. 5.4 c) observe that the spectrum looks like a comb extended infinitely in both directions; this filter has been given the name comb filter. Continuing the analogy, the peaks will be called teeth. Its spectrum can be summarized in the following list.

Comb filter $r = 1$

Tooth Center Frequencies $f_n = \frac{n}{P\Delta}$; $n = 0, \pm 1, \ldots$

Tooth Bandwidth $= \frac{1}{MP\Delta}$

Quality $Q = \frac{f_{n+1} - f_n}{\text{bandwidth}} = M$

The case $r = -1$ or $\alpha = \frac{1}{2}$ is readily available from the above results in Eq. 5.11; there is just a frequency shift of $\frac{1}{2P\Delta}$. Since $\frac{1}{P\Delta}$ is the separation of the teeth, the case $r = -1$ is the filter with teeth shifted half this distance.

Comb filter $r = -1$

Tooth Center Frequencies $f_n = \frac{n}{P\Delta} - \frac{1}{2P\Delta}$; $n = 0, \pm 1, \ldots$

Tooth Bandwidth $= \frac{1}{MP\Delta}$

Quality $Q = M$

Complex Comb Filter

When $r = \pm j$, complex filters result. Setting $q = 1$ or 3 in the last of Eqs. 5.8 produces $H(f)$. Removing the phase shift term as identified above, there still remains a complex function. Let $r = +j$ or $\alpha = \frac{1}{4}$.

$$H_c(f) = e^{j\left(\frac{M-1}{2}\right)\frac{\pi}{2}} \operatorname{sinc}\left(M P \Delta\left(f + \frac{1}{4 P \Delta}\right) ; M\right)$$

$$= e^{j\frac{M}{4}\pi} \; e^{-j\frac{\pi}{4}} \operatorname{sinc}(- - -)$$

$$= \begin{cases} (-1 + j)\frac{\sqrt{2}}{2} \operatorname{sinc}(- - -) & M/4 \text{ odd} \\ (1 + j)\frac{\sqrt{2}}{2} \operatorname{sinc}(- - -) & M/4 \text{ even} \end{cases}$$

$$= (-1)^{M/4} (1 - j)\frac{\sqrt{2}}{2} \operatorname{sinc}(- - -)$$

$$= H_R(f) + j H_I(f) \qquad (5.13)$$

which defines the real and imaginary parts. To make the gain on each part equal, M should be divisible by 4, as stated above. Note also the frequency shift of $\frac{1}{4 P \Delta}$ for the case $r = +j$.

There is a more instructive procedure.

$$\frac{1}{M} \sum_{k=0}^{M-1} j^{M-1-k} \delta(\tau - k P \Delta) \qquad (5.14)$$

$$= \frac{1}{M} \sum_{k \text{ even}} + \frac{1}{M} \sum_{k \text{ odd}}$$

$$= j \frac{1}{M} \sum_{\ell=0}^{\frac{M}{2}-1} (-1)^{\frac{M}{2}-1-\ell} \delta(\tau - \ell 2 P \Delta)$$

$$+ \frac{1}{M} \sum_{m=0}^{\frac{M}{2}-1} (-1)^{\frac{M}{2}-1-m} \delta(\tau - P\Delta - m2P\Delta)$$

We recognize two (real) comb filters with parameters $\frac{M}{2}$, P, 2Δ, $r = -1$ and the second has an additional delay of $P\Delta$. Therefore,

$$H_c(f) = H_R(f) + j\, H_I(f) \tag{5.15}$$

$$= \left[\frac{1}{2} e^{-j2\pi fP\Delta} \operatorname{sinc}\left(MP\Delta\left(f + \frac{1}{4P\Delta}\right); \frac{M}{2}\right)\right.$$

$$\left. + j\frac{1}{2} \operatorname{sinc}\left(MP\Delta\left(f + \frac{1}{4P\Delta}\right); \frac{M}{2}\right)\right] e^{j2\pi fP\Delta - j\frac{\pi}{4}}$$

Of course, we claim the filters of Eq. 5.13 and Eq. 5.15 are the same. Note the additional delay in the real filter which synchronizes its output with the imaginary filter; indeed the time-domain operation adds to the real output first, then to the imaginary output. (See the τ-diagram.)

Complex comb filter $r = j$; each part

Tooth Center Frequencies $f_n = \frac{n}{2P\Delta} - \frac{1}{4P\Delta}$; $n = 0, \pm 1, \ldots$

Tooth Bandwidth $= \frac{1}{MP\Delta}$

Quality $Q = \frac{M}{2}$

Note that the teeth are half as far apart compared to $r = \pm 1$, the bandwidth is the same and the quality is half. Similar results are

obtained for $r = -j$.

Complex comb filter $r = -j$; each part

Tooth Center Frequencies $f_n = \frac{n}{2P\Delta} - \frac{3}{4P\Delta}$; $n = 0, \pm 1, \ldots$

Tooth Bandwidth $= \frac{1}{MP\Delta}$

Quality $Q = \frac{M}{2}$

5.2.5 General Case; Complex r. The general case has for r any complex number. It is convenient to represent r in the form

$$r = a\, e^{j2\pi\alpha}; \qquad a > 0, \quad \alpha \text{ real} \tag{5.16}$$

$$= e^{\ell n a + j2\pi\alpha}$$

Also, define $$u = \frac{1}{2}\, \ell n\, a, \quad a = e^{2u} \tag{5.17}$$

$$v = f + \frac{\alpha}{T} \quad . \tag{5.18}$$

From Eq. 5.6,

$$H(f) = \frac{a^{M-1}\, e^{j(M-1)2\pi\alpha}}{M} \; \frac{\left(1 - e^{-2M\left(\frac{1}{2}\ell na + j\pi T\,(f + \frac{\alpha}{T})\right)}\right)}{1 - e^{-2\left(\frac{1}{2}\ell na + j\pi T\,(f + \frac{\alpha}{T})\right)}} \tag{5.19}$$

$$H(v) = \frac{e^{2(M-1)u}\, e^{j(M-1)2\pi\alpha}}{M} \; \frac{\left(1 - e^{-2M(u + j\pi Tv)}\right)}{\left(1 - e^{-2(u + j\pi Tv)}\right)}$$

$$= e^{j(M-1)2\pi\alpha - j(M-1)\pi Tv} \; \frac{e^{(M-1)u}}{M} \; \frac{\sinh\,(Mu + j\pi MTv)}{\sinh\,(u + j\pi Tv)} \tag{5.20}$$

The variable v is f frequency shifted by the amount $\frac{\alpha}{T}$. The parameters are M, T, u. The similarity to the result in Eq. 5.8 is evident; as a check, for $r = 1$, we have $a = 1$ and $\alpha = 0$, so $u = 0$ and $v = f$; then use $\sinh(j\lambda) = j \sin \lambda$.

The power spectrum will be computed.

$$|H(v)|^2 = \frac{e^{2(M-1)u}}{M^2} \frac{(\sinh^2 Mu + \sin^2 \pi MTv)}{(\sinh^2 u + \sin^2 \pi Tv)} \qquad (5.21)$$

All arguments are real. With $u = 0$, the zero's and peaks are determined by the zero's of $\sin \pi MTv$ and $\sin \pi Tv$ as already discussed. With the additional u terms, $|H(v)|^2$ will be changed but its characteristics will remain fixed. For example, $|H(v)|^2$ is still periodic in v with period $\frac{1}{T}$, it will exhibit peaks and ripple between them, but their positions will be changed. For no value of v is $H(v) = 0$.

We claim that $|H(v)|^2$ is "smooth and concave" near its peaks. Near the peak at $v = 0$, approximate $\sin\lambda$ by λ and note that $\sinh^2 Mu > \sinh^2 u$. The peak behaves as the function

$$\frac{\frac{1}{2} + v^2}{\frac{1}{4} + v^2} .$$

Several examples are given for illustration. Consider the filter with $M = 4$, $T = 1$ and $r = .90$. Then $a = .90$, $\alpha = 0$; $u = -.05268$ and $v = f$, there being no frequency shift.

$$|H(f)|^2 = \frac{.729}{16} \frac{(\sinh^2 .21072 + \sin^2 4\pi f)}{(\sinh^2 .05268 + \sin^2 \pi f)} \qquad (5.22)$$

This function is illustrated in Fig. 5.5 a). The function is also plotted for several other values of the parameter M. Figure 5.5 b) illustrates curves for $r = 1.0$ and several values of M. Figure 5.5 c) has the coefficient $r = .5$. Of interest is the value of the peak for r values; Fig. 5.6 shows $|H(0)|^2$ and its dependence on r.

$$|H(0)|^2 = \frac{e^{2(M-1)u}}{M^2} \cdot \frac{\sinh^2 Mu}{\sinh^2 u} \; ; \qquad u = \frac{1}{2} \ln |r| \qquad (5.23)$$

To summarize this section, a standard form has been found for the filter with impulse response $h(\tau)$. The filter has three parameters: M, T, r. M is the number of delays, counting one of zero delay; this is the quality Q of the filter and determines the number of ripples in the spectra. T is the amount of each delay and $\frac{1}{T}$ is the spacing between peaks in the spectra. The complex coefficient r determines the type of filter; in the form $r = a\, e^{j2\pi\alpha}$, the product $\frac{M}{2} \ln a$ determines the shape of the spectra and α fixes the position of the peaks.

To this point, there has been deliberate avoidance of any mention of the properties of the input signal, such as low-pass, bandpass or sampled data. The above analysis is for a class of linear filters

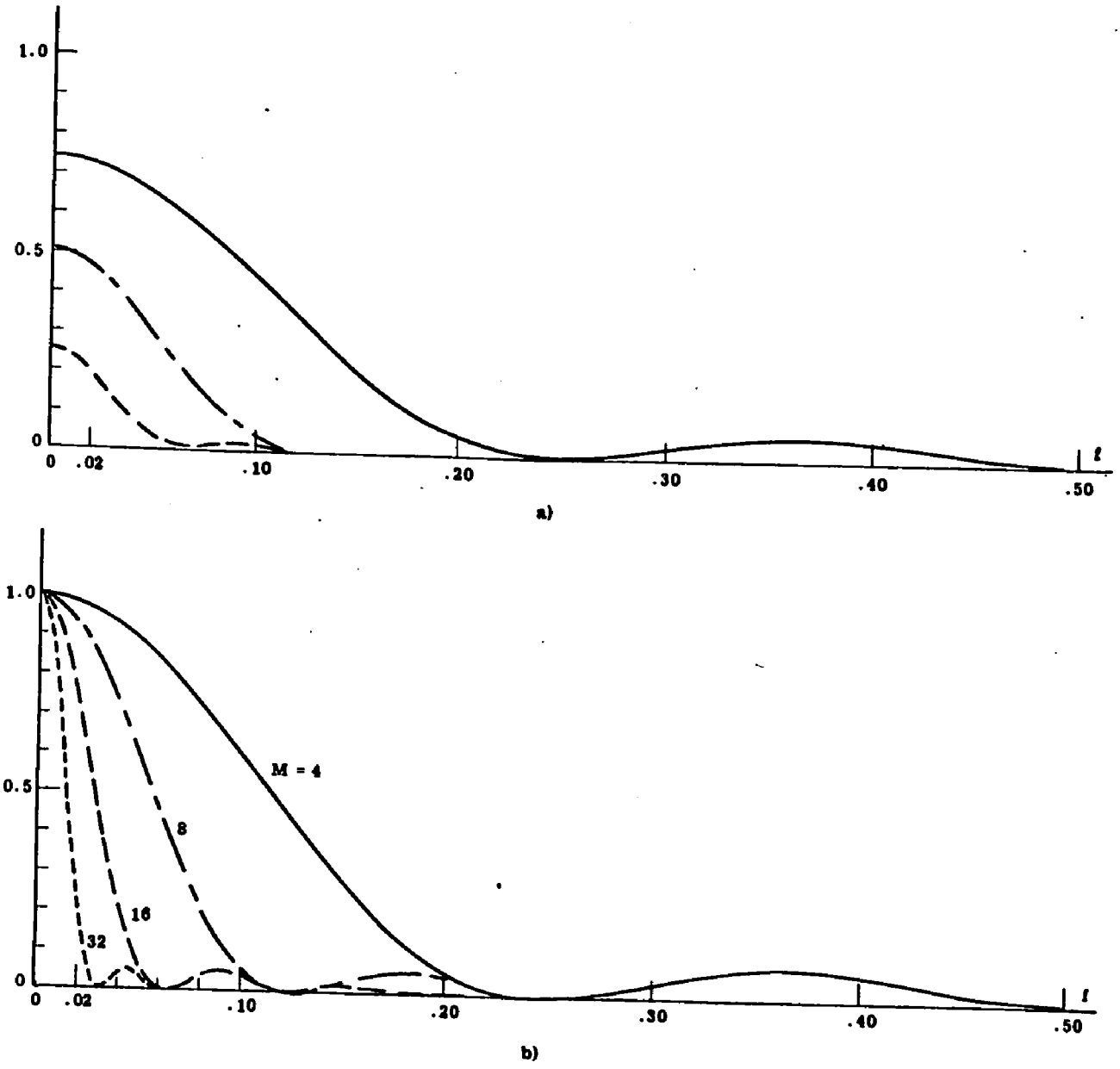


Fig. 5.5. a) Power spectra $|H(f)|^2$ for r = .90
b) Power spectra $|H(f)|^2$ for r = 1.0

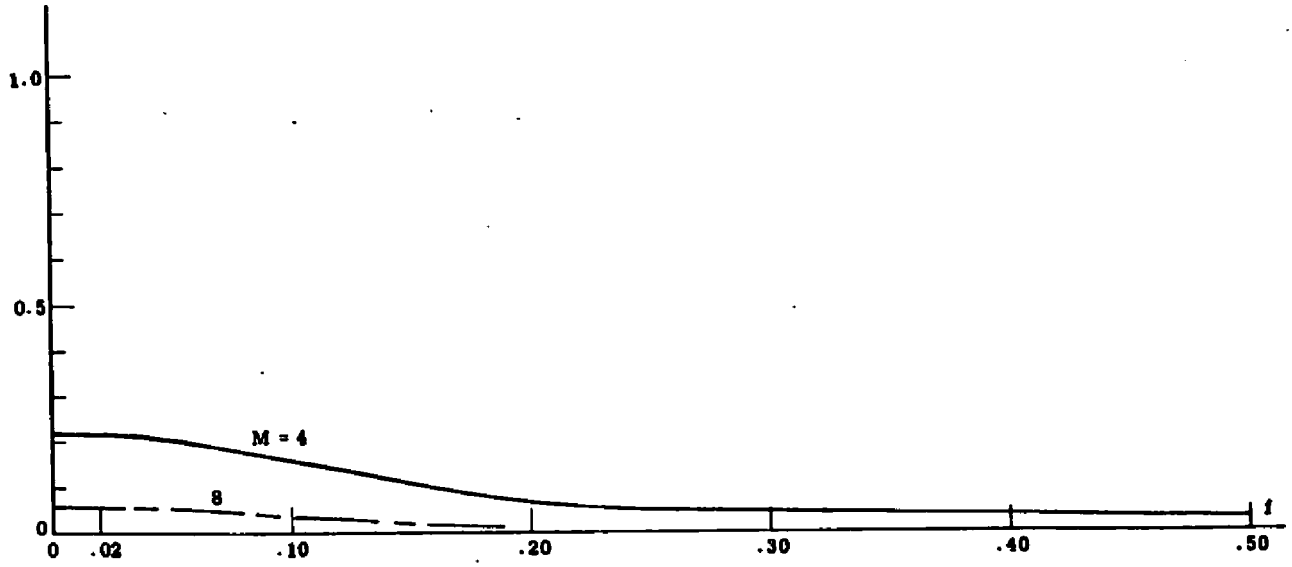


Fig. 5.5. c) Power spectra $|H(f)|^2$ for r = .5

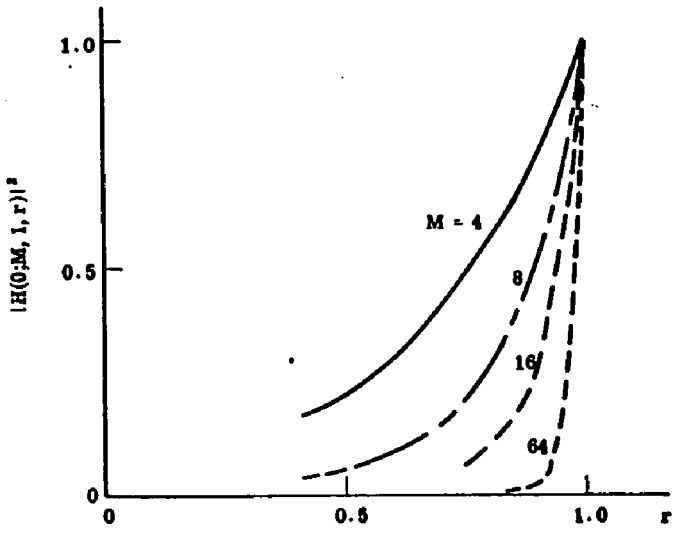


Fig. 5.6. $|H(0)|^2$ versus r

with the specified impulse response of Eq. 5.1. The results are quite general.

5.3 A Standard Form of Digital Processors

The impulse function $\delta(\tau - kT)$ is in the class of generalized functions. We have already used it with other members of this class through the convolution operator $*$. By definition, $s(t) * \delta(t - kT) = s(t - kT)$ and the net operation is a delay (time-displacement) of the function $s(t)$.

5.3.1 Sampling. The operation of sampling a function can also be modeled using the impulse function. The operator $\cdot$ (multiply) is used and sampling is modeled as (by definition)

$$s(t) \cdot \delta(t - \epsilon) = s(\epsilon) \quad . \tag{5.24}$$

An infinite sequence of such delta functions may be represented as

$$\delta_\epsilon(t) = \{\delta(t - k\epsilon)\}_{k=-\infty}^{\infty} \tag{5.25}$$

and is called an impulse train. The sequence of samples of the function is

$$\{s(k\epsilon)\}_{k=-\infty}^{\infty} = \{s(t) \cdot \delta(t - k\epsilon)\}_{k=-\infty}^{\infty} = s(t) \cdot \delta_\epsilon(t) \tag{5.26}$$

With this model of the sampling operation, the digital processor (discrete time) equivalent to $h(\tau)$ can be derived. The big point is

that its spectra are already completely known.

The output of a linear filter may obviously be sampled; this means we are looking for the results only at discrete instants in time. Clearly, the input may also be a sequence of samples (from some other filter) but usually the input and output samples are synchronized. The model for the filter $h(\tau)$ is illustrated in Fig. 5.7.

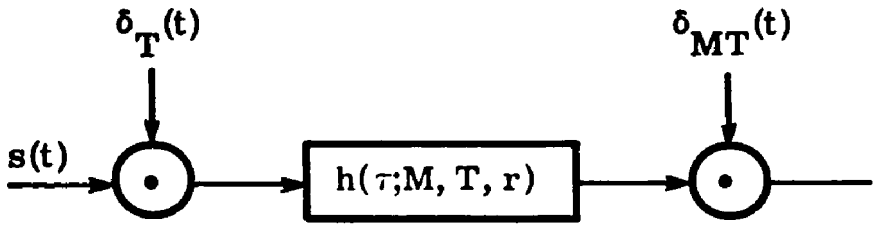


Fig. 5.7. Sampled filter

5.3.2 Equivalent Digital Processor. Now that only samples at discrete times are involved, we recognize that an exact equivalent digital process or (algorithm) can be found for the filters of the class being studied. Since the output is sampled, the tapped delay line realization is not needed. A single delay unit with impulse response $h(\tau - T ; r)$ is used as shown in Fig. 5.8. Such a linear filter is often called a circulating average. In digital devices, the delay operation is just the storing of results for the next recursion. For

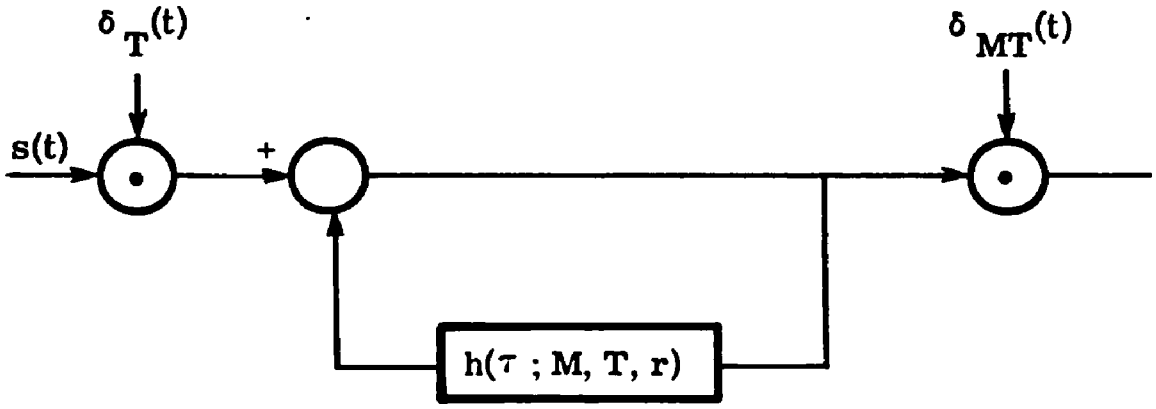


Fig. 5.8. Sampled circulating average

the simplest situation where only one number needs to be stored, the digital processor is shown in Fig. 5.9.

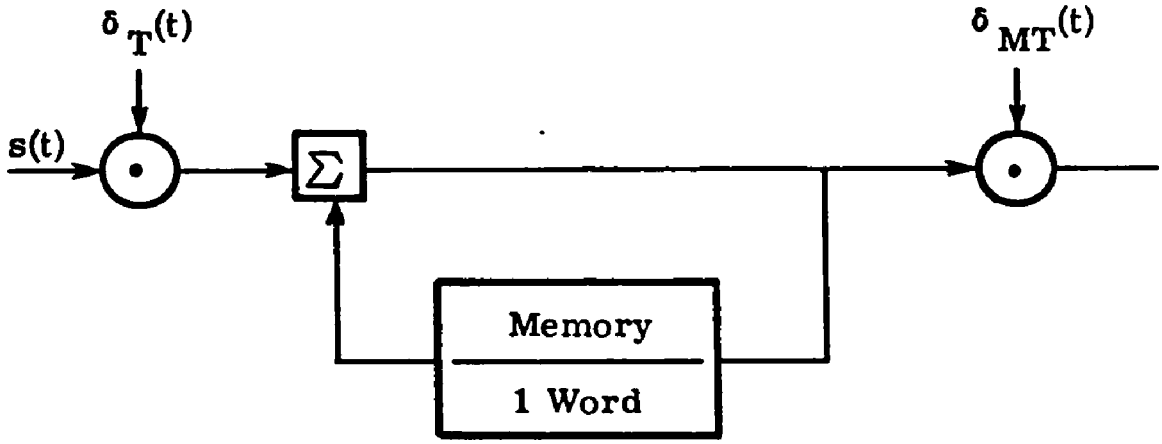


Fig. 5.9. Simple circulating average

The input sampling corresponds to accepting data at discrete times and the output sampling takes the result when the processing is done; these sampling operations are implied in digital processing and will be incorporated with the digital processor.

When a sequence of multiple inputs is provided in one duration T, a multi-position memory is required. Let the number of positions be P and $T = P\Delta$, where Δ is the interval between input samples. See Fig. 5.10.

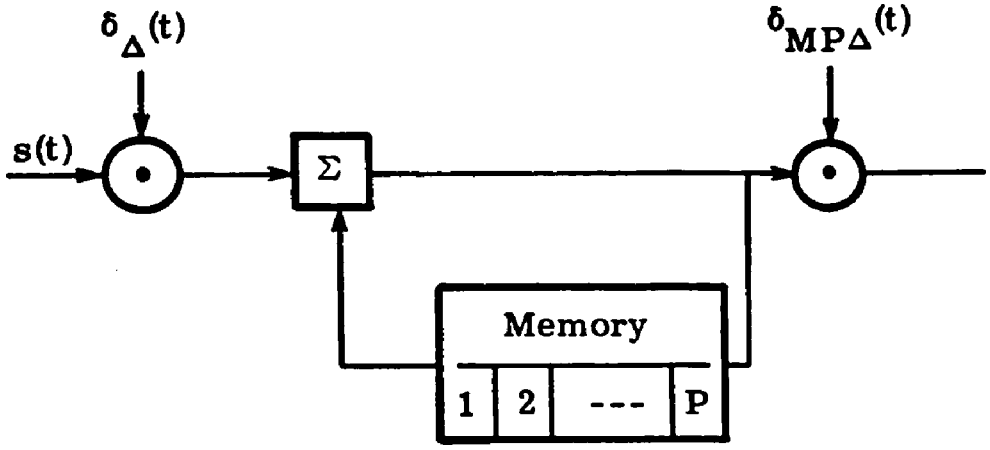


Fig. 5.10. P-position circulating average

5.3.3 MPR Block Diagram. A convenient block diagram description of this standard digital processor is shown in Fig. 5.11. From the parameters involved, this processor will be called an MPR filter and its acronym might have been taken from the apt descriptive phrase, "multi-positional recirculator."

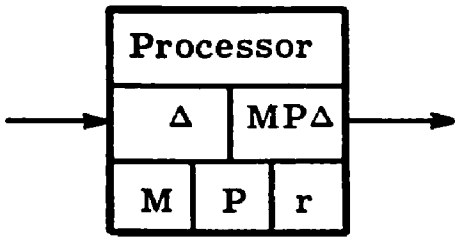


Fig. 5.11. MPR processor

5.3.4 Examples. For the case $P = 1$ and $r = 1$, the processing consists of averaging over M values.

$$\frac{1}{M} \sum_{k=0}^{M} s(k\Delta)$$

so that this is often called Block Averaging (BAV). For $P = 1$ and $r < 1$ with M very large, a weighted block average results which approximates RC filtering.

For $P > 1$, P words of memory are required and P words are contained in the output; we may regard this array of P numbers as the discrete time description of the analog result.

The significance of a complex r has been noted. For $M = 8$, $P = 1$ and $r = j$, a complex filter results with a single complex output. It is equivalent to two parallel real filters; the block diagram is presented in Fig. 5.12. An additional delay of Δ is needed in the model to synchronize the x and y outputs. The MPR standard

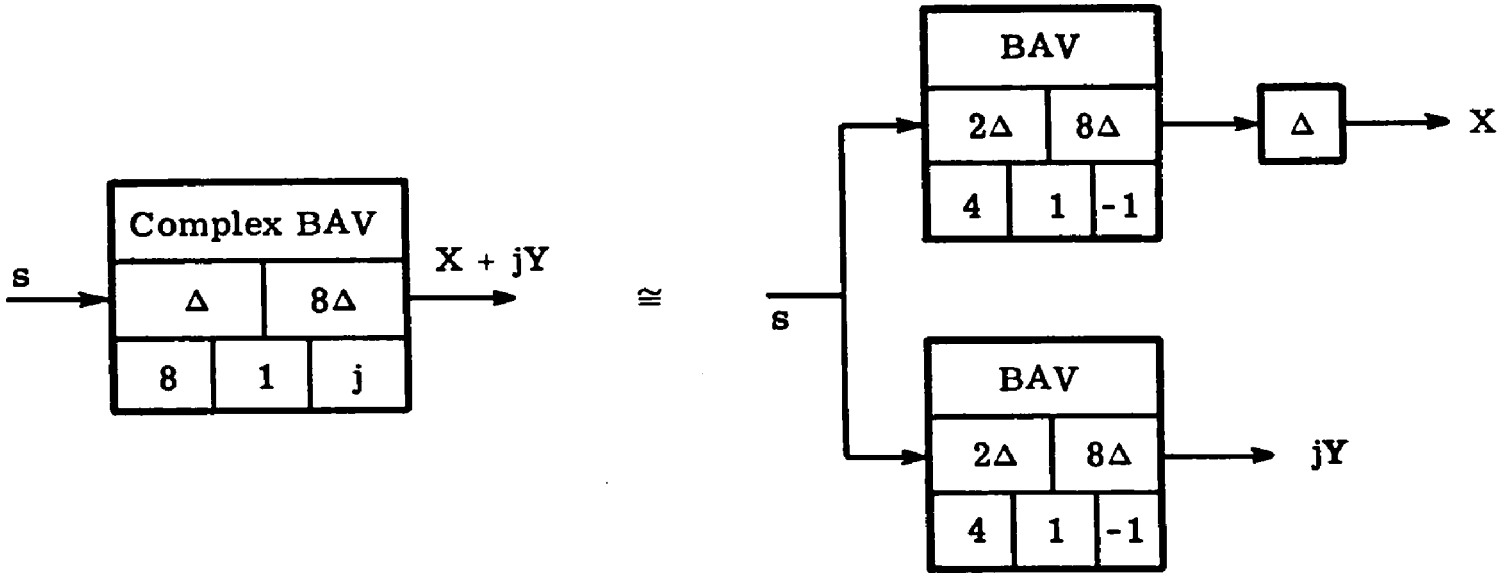


Fig. 5.12. The complex block average

form can not realize a pure delay due to the choice of the form of $h(\tau)$.

5.4 Combined Operation

Just as in the continuous case, these digital processor building blocks may be combined to form more elegant processing. The topics of input-output matching and cascading of filters will be discussed.

A pair of digital processor blocks can be successfully cascaded in real time under certain conditions; their output and input sampling must be coincident. For example, in Fig. 5.13 a cascade

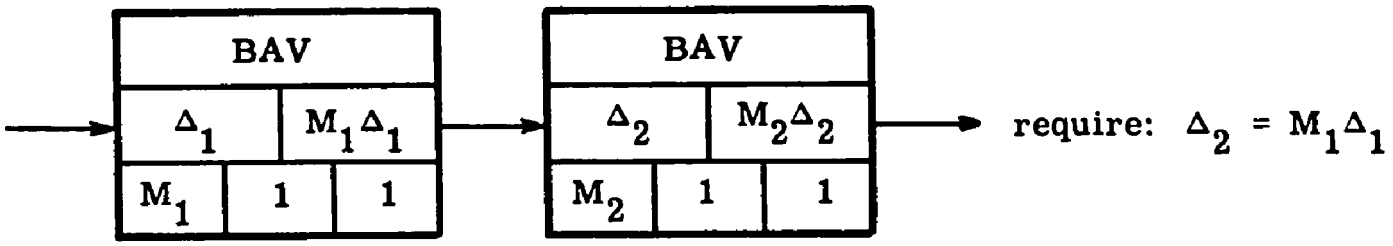


Fig. 5.13. A cascade of block averages

of two BAV blocks is allowed if $\Delta_2 = M_1\Delta_1$. Likewise, in Fig. 5.14 a CIRAV followed by a BAV requires $\Delta_4 = M_3\Delta_3$ and $M_4 = P_3$.

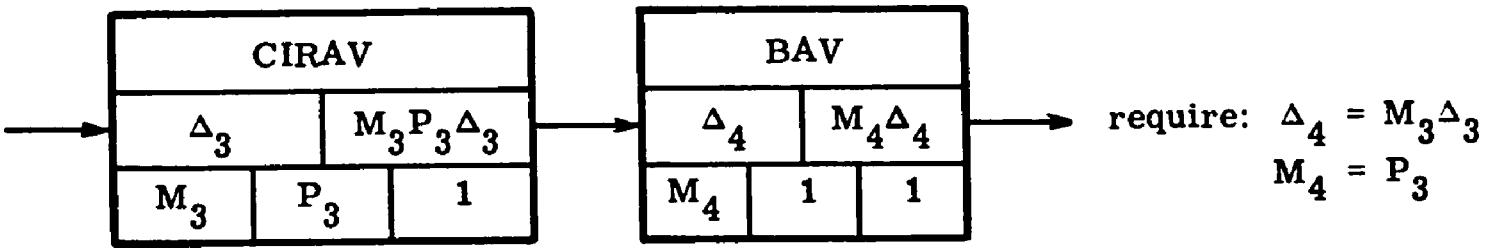


Fig. 5.14. A cascade of circulating average and block average

Furthermore, the memory of the devices can be exploited so that non-real-time processing can be performed. It is sensible to perform a CIRAV in real-time and then to process the P-dimensional array at a faster, slower or later time. This analysis has shown that the resultant spectra will not be affected by such implementation convenience.

5.5 Time-Domain Viewpoint

In the time-domain, the MPR processor is viewed as performing a weighted average of the set of M inputs. For example, if $r = +1$, an average is obtained and a signal is presumably present. If the signal is constant (but unknown), a single accumulation ($P = 1$) is needed. If the signal is periodic, a multi-position accumulator of P words is necessary. In either case, the purpose is to enhance the signal portion and reduce the effect of noise; the signal portion is correlated but the noise is uncorrelated, so the variance of the result is reduced.

On the other hand, if the noise by itself is of interest, the signal portion must be cancelled; this is accomplished with even M and with $r = -1$, so that signs alternate. This type of processing eliminates the signal and noise mean, so that the resultant can be studied for other properties of the noise, such as its covariance and average noise power. Both of these cases, $r = +1$ and -1, and others are used in the data processing described in the next chapter.

Given the statistics of the input process to such processors, the precise description of the output statistics is not simple. In general, an M-fold convolution of probability densities is required. Only for some densities can the integral be evaluated directly, and tabulations of difficult integrals are limited. Fortunately, for the Gaussian case, by linearity, the resultant output is Gaussian and the mean

and variance can be calculated. A detailed statistical analysis of the MPR processor is not required for this study; the processes to be studied are modeled as Gauss-Markov.

There are advantages to the combined viewpoints of the frequency-domain and time-domain. The frequency-domain viewpoint keeps the spectral picture of the signal in the foreground and clarifies from where the signal and noise measurements come. If the signal is periodic and has a known line spectrum, one MPR "comb filter" can be used to "look for" these lines while another measures the noise properties between the spectral lines of the signal. The time-domain viewpoint keeps clear the temporal operation of the MPR filter and is the spring-board for a detailed statistical analysis.

To conclude this chapter, the MPR processor is reviewed. It is a linear, sampled-data processor which has a continuous analog with impulse response of Eq. 5.1. The parameter M is the number of samples used in the averaging. P is the number of positions or words in the result. The parameter r is the radix in the coefficients of the average; r may be complex, in which case one real input sequence produces two real output sequences.

CHAPTER 6

DATA PROCESSING

Synopsis

6.1 Introduction

Model equations

Kalman-Bucy equations

Parameter estimation equations

6.2 Computer Program Flow Diagram

6.3 Monte Carlo Tests

Gauss-Markov sequence sample functions

Testing a stationary model

6.4 Source of Field Data

Underwater acoustics

Dual-comb data processing

Processes T, N

6.5 Experimental Procedure

Inspection of sample functions

Checking initial conditions

Inspection of filtered process

Performance

Time-varying model

6.1 Introduction

In this and following chapters, the Kalman-Bucy equations and the parameter estimation equations will be applied. For convenience, the relevant equations developed in Chapters 2, 3, and 4 are repeated here.

Model Equations

Process: $$x_k = \Phi_k x_{k-1} + u_k \tag{6.1}$$

Observation: $$y_k = H_k x_k + v_k \tag{6.2}$$

Initial Condition: $$x_0 \tag{6.3}$$

Recall that u_k and v_k are Gaussian white noise sequences which are defined in Eqs. 3.23 and 3.24.

Kalman-Bucy Equations

Process estimates: $$\hat{x}_{0|0} = x_0 \tag{6.4}$$

$$\hat{x}_{k|k-1} = \hat{\Phi}_k \hat{x}_{k-1|k-1} \tag{6.5}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + A_k(y_k - H_k \hat{x}_{k|k-1}) \tag{6.6}$$

$$\text{for } k = 1, 2, \ldots$$

where the optimal gain A_k is defined by the computations

$$P_1 = E\{x_0 x_0'\} \tag{6.7}$$

$$P_k = \hat{\Phi}_k \overline{P}_{k-1} \hat{\Phi}_k' + \hat{Q}_k \tag{6.8}$$

$$A_k = P_k H_k'(H_k P_k H_k' + \hat{R}_k)^{-1} \tag{6.9}$$

$$\overline{P}_k = (I - A_k H_k) P_k (I - A_k H_k)' + A_k \hat{R}_k A_k' \tag{6.10}$$

P_k and $\overline{P}_k$ are the error covariances defined in Eqs. 3.19 and 3.13; the required order of computation is Eqs. 6.8, 6.9, 6.10, 6.5 and 6.6.

Parameter Estimation Equations

$$\hat{\Phi}_k = \left(\sum_{j=3}^{k} c_{k-j} y_j y_{j-2}' \right) \left(\sum_{j=3}^{k} c_{k-j} y_{j-1} y_{j-2}' \right)^{+} \tag{6.11}$$

$$\hat{R}_{k-1} = -\hat{\Phi}_k^{-1} B_{1,k} \tag{6.12}$$

$$\hat{Q}_k = B_{0,k} - \hat{R}_{k-1} - \hat{\Phi}_k R_{k-1} \hat{\Phi}_k' \tag{6.13}$$

The statistics $B_{0,k}$ and $B_{1,k}$ are defined in Eqs. 4.34 and 4.35. Figure 6.1 shows the processing of the data before its use in the model. Some of the operations are described in the next chapter.

The application of these equations is toward a study of two acoustic processes. These acoustic processes will be defined in

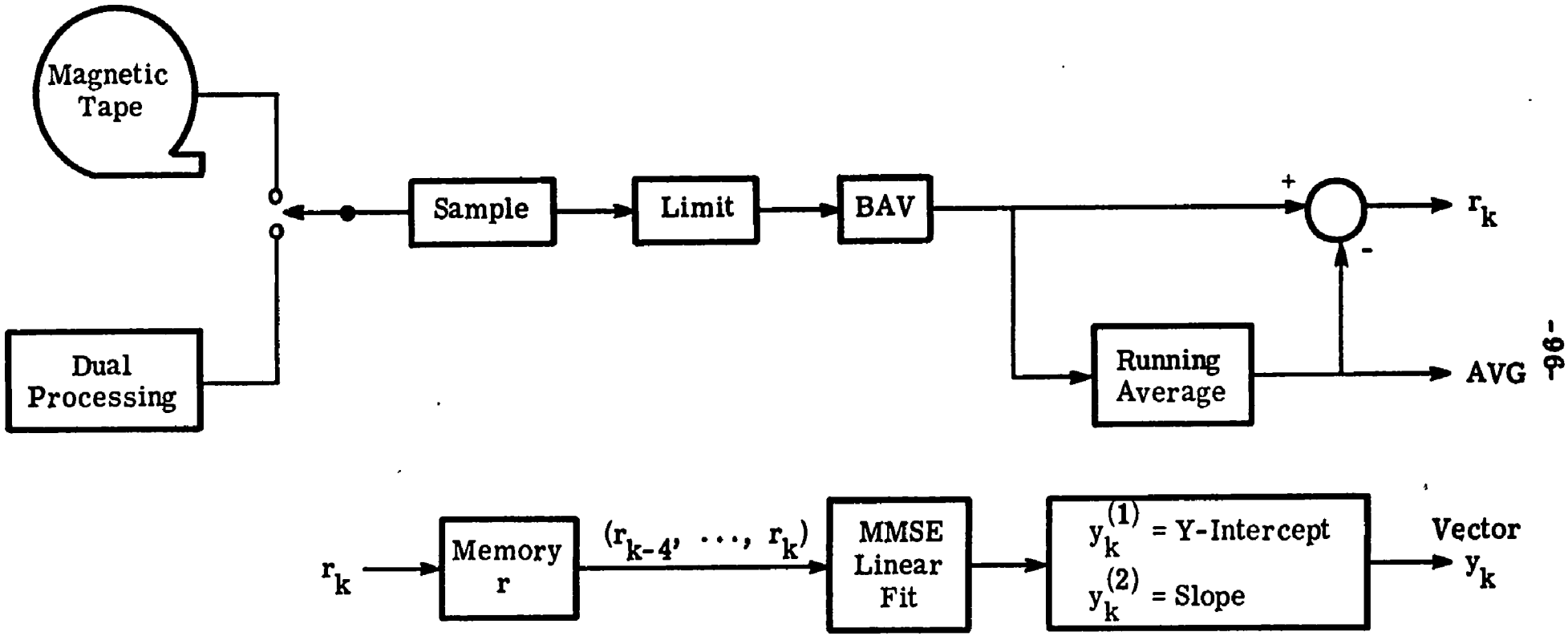


Fig. 6.1. Preprocessing of data

this chapter and related to an on-going acoustic signal processing program.

A computer flow diagram is presented to show the recursive computations in yet another form. A Monte-Carlo test of the parameter estimation equations is included to study the model under controlled and known statistics.

6.2 Computer Program Flow Diagram

The heart of this program is the set of equations (Eqs. 6.4 - 6.13) described above. However, certain ancillary features make the program very flexible in its use. These features include the calculation of initial conditions, a separate evaluation of performance, a minimum mean square error fit of a line to the observed data, the selection of several modes of parameter estimation and the selection of several output formats. The last two items are briefly discussed here.

Due to the experimental nature of this work, it is deemed desirable to estimate the parameters Φ, Q and R simultaneously and also separately in various combinations. The following modes of estimation are allowed.

1. Φ, Q, R, as in the equations
2. Φ, Q, R fixed at initial values
3. Φ, with Q, R fixed at initial values
4. Φ, Q, with R fixed at initial value

There may be physical reasons for fixing Φ, Q or R. Furthermore, during early phases of modeling, such procedures often help to get a ball-park estimate of the parameters.

Also desirable is a selection of output formats of such a program. Of obvious interest are the series of computed estimates and the filter output vector. The list of options allowed is given below.

1. Φ_k, Q_k, R_k, A_k, P_k, $\hat{x}_{k|k}$
2. Performance Indices $J0_k$, $J1_k$, $J2_k$, $J3_k$, $J4_k$
3. Forecasts of x
4. Statistics S_{00}, S_{12}, S_{02}, B_0, B_1
5. Program variables
6. Observed data

These options were found useful in particular when field data was used. Item 3 will be explained in the next chapter.

The flow diagram is given in Fig. 6.2. The choice of options is made by the "mode". The program was written in Fortran, for an 8K PDP-8 system and for the Level-G Fortran system in MTS (Michigan Terminal System). Figure 6.3 contains the flow chart for the initial condition input subroutine ICIN; it consists of reading the control and run parameters and the initial conditions for Φ, Q and R, from which the initial conditions of the statistics are computed. We found this particular flow of action very convenient. Figure 6.4

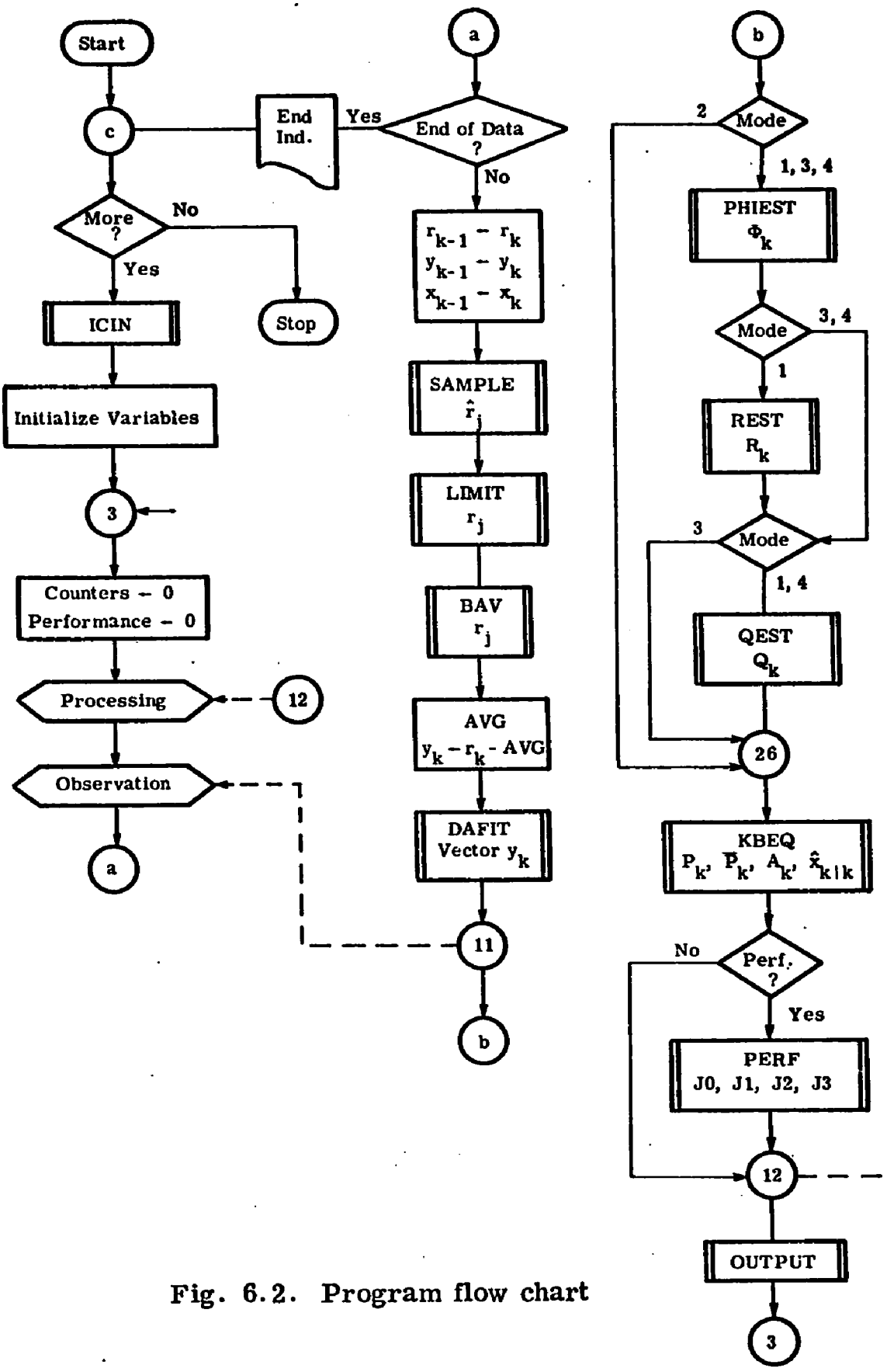


Fig. 6.2. Program flow chart

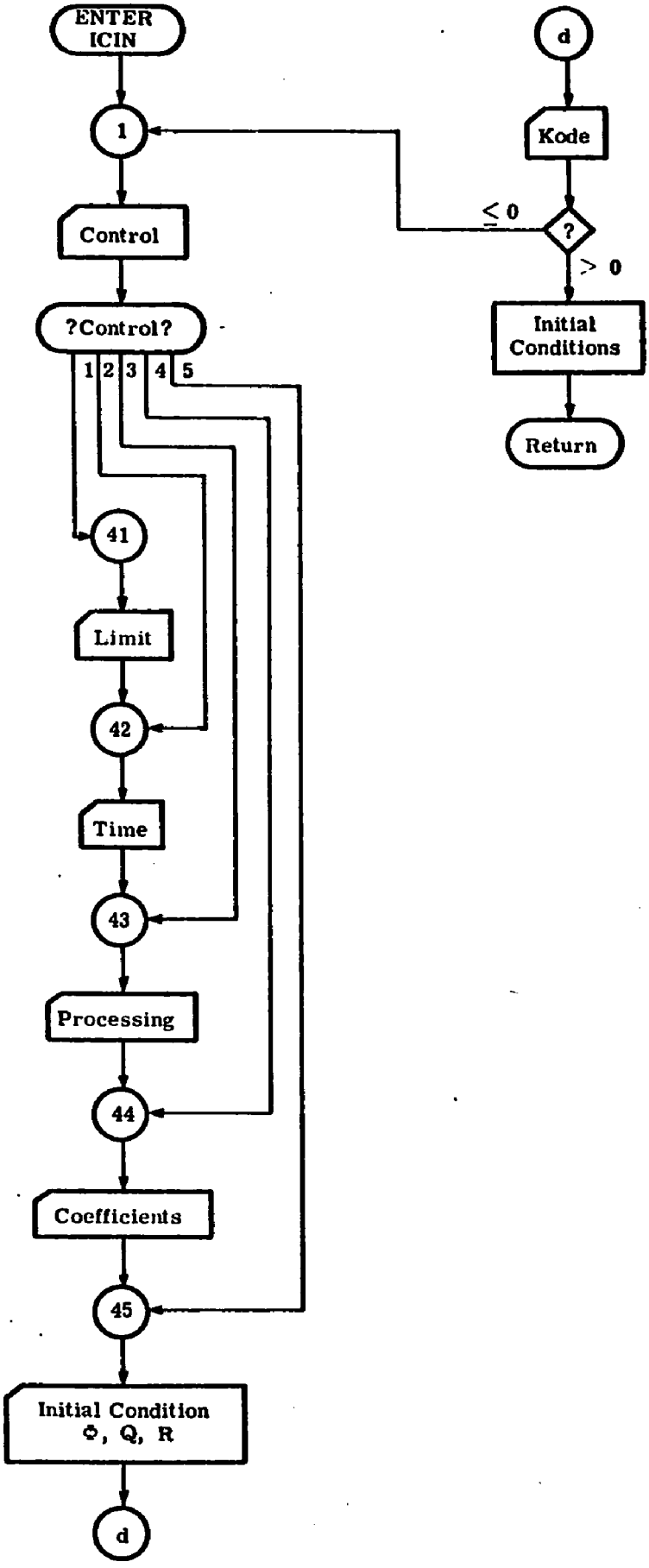


Fig. 6.3. Flow chart for ICIN

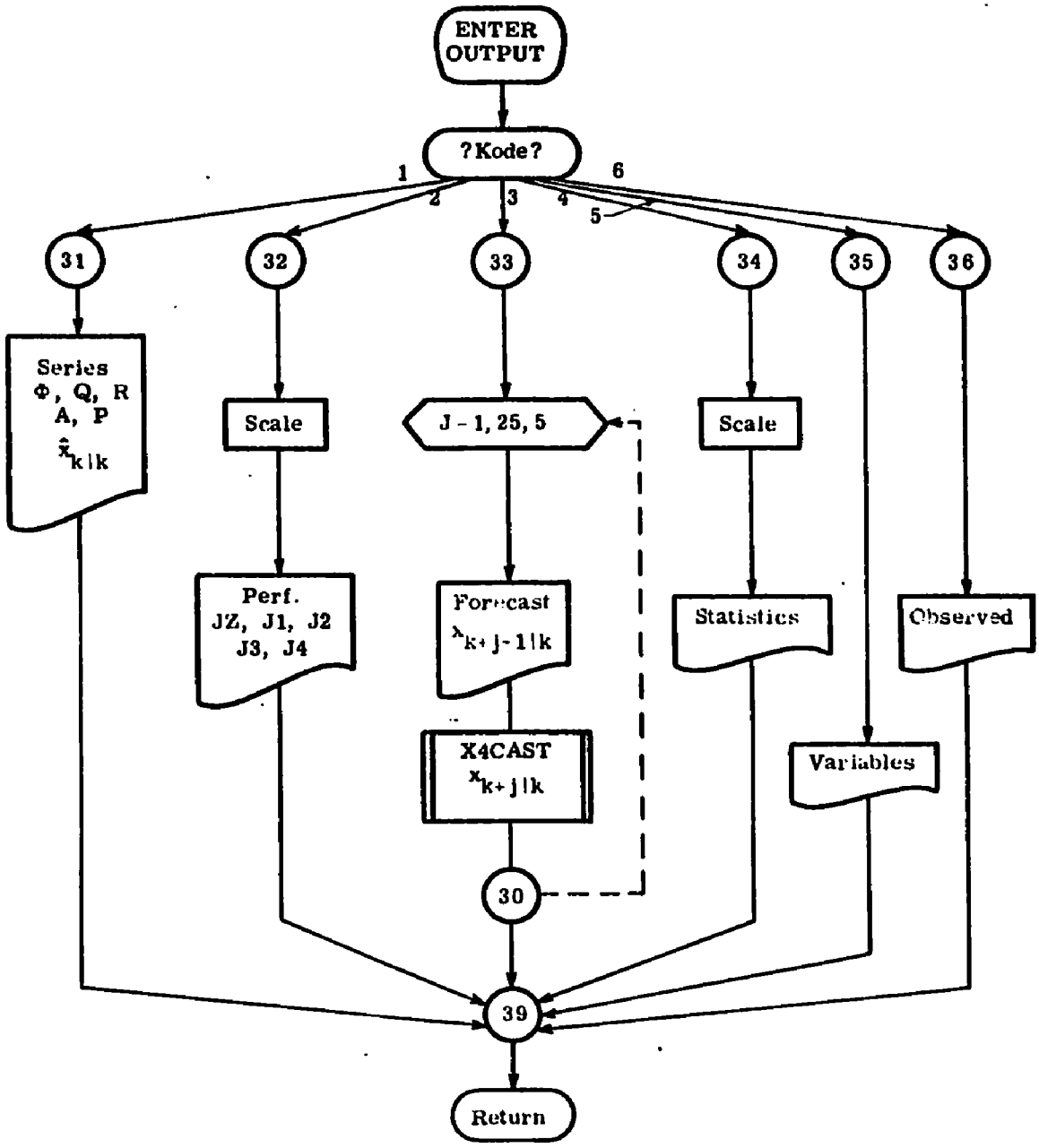


Fig. 6.4. Flow chart for OUTPUT

is the flow chart for the output subroutine OUTPUT; the selection of output quantities is called the "kode".

6.3 Monte-Carlo Tests

6.3.1 Gauss-Markov Sample Functions. A few sample functions of a discrete-time Gauss-Markov process (called a Gauss-Markov sequence) are illustrated in Fig. 6.5. All vectors are simply scalars (n = 1) and

$$x_k = \Phi x_{k-1} + u_k ,$$

where $\{u_k\}$ is a sequence of independent Gaussian variables with zero mean and variance Q. A standard random number generator routine was used.

The sample functions are typical for a Markov sequence with order $n=1$. Observe that no particular pattern, no consistent oscillations are apparent. The values of x are correlated with correlation coefficient Φ. The x process has zero mean and variance $Q/(1-\Phi^2)$.

6.3.2 Testing a Stationary Model. The parameter estimation schemes were tested using a time-invariant model, $n=1$. The values are $\Phi = .90$, $Q = 1.0$, $R = .40$; this is a moderately noisy observation system. The results of 38 runs are summarized in Table 6.1; they agree with previous experiments of AKKW (Ref. 1). Note that Φ is slightly lower than its "true" value. Q is nominally

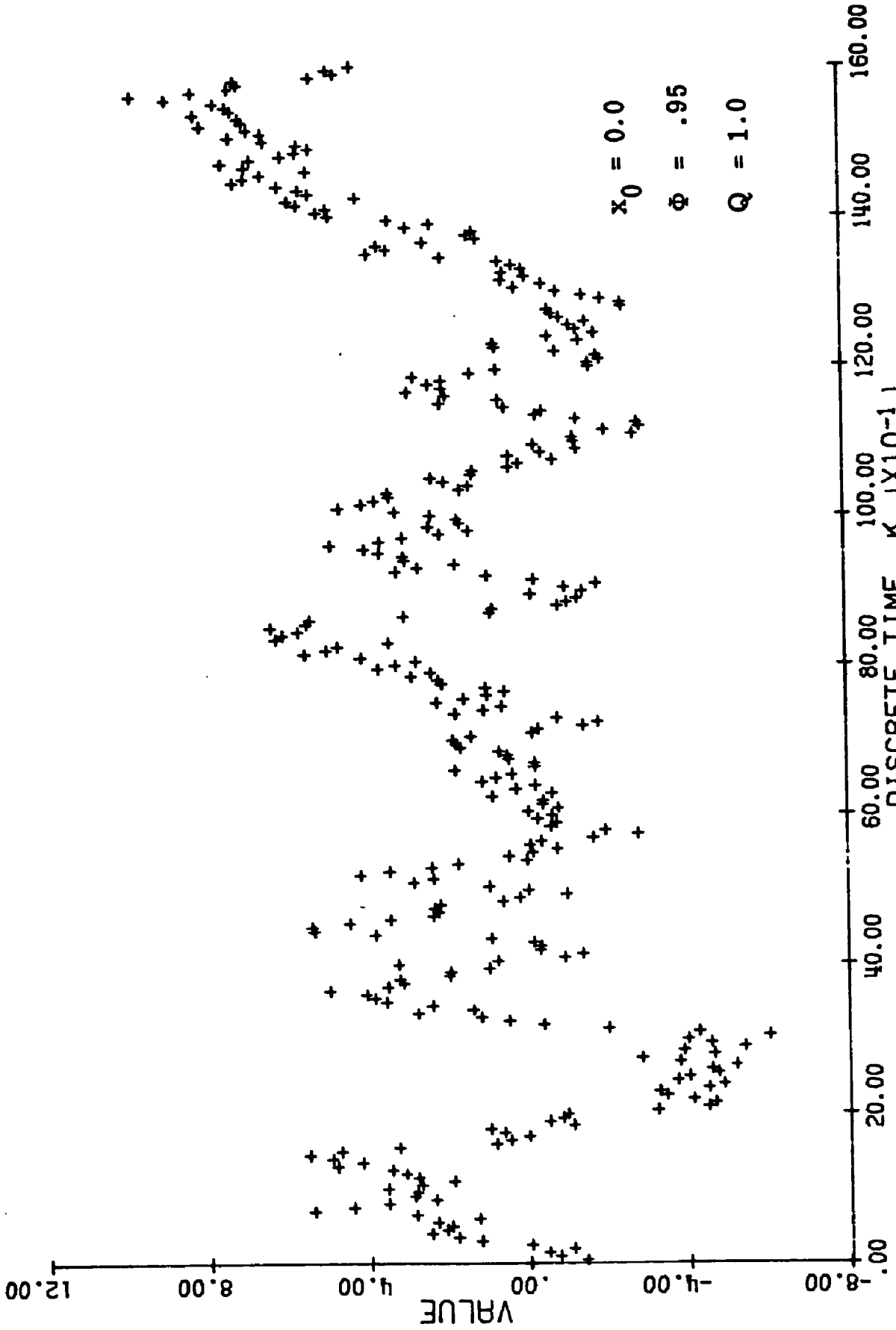


Fig. 6.5. Gauss-Markov sequence sample function

estimated correctly. The biggest problem is in estimating R; it is under-estimated in all cases.

k	AVG. $\hat{\Phi}$	VAR. $\hat{\Phi}$	AVG. $\hat{Q}$	VAR. $\hat{Q}$	AVG. $\hat{R}$	VAR. $\hat{R}$
500	.8848	.0014	1.0133	.0523	.1329	.0152
1000	.8837	.0015	1.0082	.0258	.1409	.0076
1500	.8960	.0008	.9920	.0357	.1440	.0089
2000	.8887	.0023	.9571	.0297	.1547	.0108

Table 6.1. $\Phi = .90$, $Q = 1.0$, $R = .40$

6.4 Source of Field Data

6.4.1 Underwater Acoustics. The following chapters are applications of this theory to a study of certain underwater acoustic processes. The data is part of a long-term and continuing study of underwater acoustics and oceanographic data. The experimental site and the laboratory facilities are described in a report by Steinberg and Birdsall (Ref. 17). Previous data acquisition is described in the reports by Under and Veenkant (Refs. 19 and 20) and Hatter (Ref. 7).

6.4.2 Dual Comb Data Processing. The term "Dual Comb" will be clarified presently. The previous chapter has presented a general treatise on a class of digital processors; there it was pointed out that these digital "filters" are readily implemented on a digital computer; a member of this class has been given the name MPR

processor.

A dual comb filter is set up as follows: one comb is made to have its peaks or teeth at a set of frequencies and then a second comb is placed exactly mid-way between these teeth; the analogy is quite figurative and the name is aptly applied.

It is common practice at this site to use a bandwidth with center frequency at 420 Hz. The sampling of data is done at four times this frequency, 1680 Hz. The data-acquisition processing of the dual comb filter is illustrated in Fig. 6.6.

The bandlimiting filter with transfer function $H_a(f)$ is of paramount importance. All subsequent processing by processors of the MPR class is broadband. It is this front-end filter which prevents aliasing of all these frequencies. The filter has a nominal bandwidth of 100 Hz, center frequency of 420 Hz and $|H_a(f)|$ has slopes of 18 db/octave.

The first MPR filter is complex and performs what in the past has been called demodulation. Its filtering action is clear: it has a peak at $f = \frac{1}{2\Delta_{IN}} - \frac{1}{4\Delta_{IN}} = 420$ Hz and a bandwidth of about 210 Hz. It produces a complex output $X + jY$ at intervals $\Delta_{OUT} = \frac{1}{210}$ sec. Its smoothing is very slight but it produces the orthogonal results X and Y.

X and Y are then subject to identical processing. The four MPR processors with $M = 2$ accomplish the interleaving of the peaks

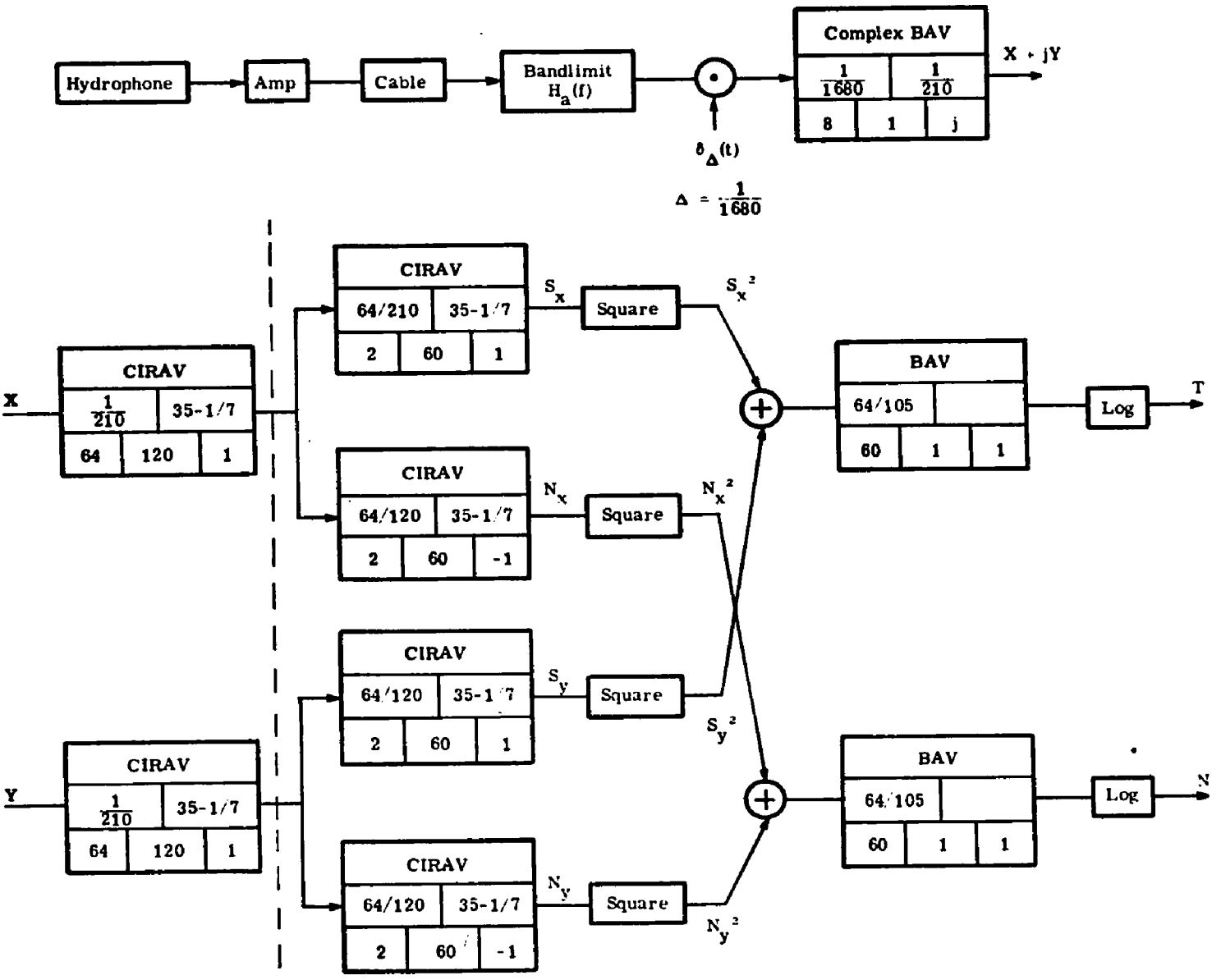


Fig. 6.6. Data-acquisition processing

of the comb filters. One has output S, the other N. The sum of the squares of the x and y components are averaged in the final BAV processors; the logarithm of both of these numbers is calculated and subsequently stored on magnetic tape.

The data is over an interval of 35-1/7 sec. The "post-processing" to the right of the dashed line in Fig. 6.6 is done in about 6 sec. and a CRT display is run for about 10 sec. Hence, the pair of observations arrive every 52 sec., approximately.

6.4.3 Processes T, N. The T process is defined* as the log of the average of the quadratic content of the S filters. The N process is defined as the log of the average of the quadratic content of the N filters.

$$T = \log \left(\Sigma \, (S_x^{\,2} + S_y^{\,2}) \right)$$

$$N = \log \left(\Sigma \, (N_x^{\,2} + N_y^{\,2}) \right)$$

To this point, no mention has been made concerning the nature of the process at the hydrophone. The processing is meaningful for any input. In the event that no signal is being transmitted, these two processes T and N should be about the same.

---

*In the experimental development of Dual Comb processing at the Cooley Electronics Laboratory, there have been several variations on this definition, differing only in how the carrier power is treated.

The primary purpose of the dual comb processing is to study the received signal power and noise power when a known signal of known period is being transmitted. The signal used in these experiments is a 420 Hz carrier amplitude modulated (switched ON-OFF) by a binary pseudo-random sequence of period 15 bits. The details of such signals are adequately discussed in previous articles. The essential feature is that the binary sequence is clocked at $\frac{1}{32} \cdot 1680$ Hz.

$$\Delta_s = 32\,\Delta, \qquad \Delta = \frac{1}{1680} \text{ as above .}$$

$$|H_s(f)|^2 = \left(\frac{\sin \pi f \Delta_s}{\pi f \Delta_s}\right)^2$$

This envelope of the signal power spectrum has 0 values at $f = \frac{m}{\Delta_s}$, $m = \pm 1, \pm 2, \ldots$. The spectral lines are located at $f = \frac{p}{15\,\Delta_s}$, $p = 0, \pm 1, \pm 2, \ldots$.

6.5 Experimental Procedure

Considerable experimentation has pointed to the need for an experimental procedure in establishing the model parameters. The following steps are useful.

1. Inspection of sample functions.

An inspection of a sufficient amount of data can give information on the model parameters Φ, Q, R and their initial conditions. For example, the T process exhibits sinusoidal behavior

which is attributable to a second order system with a stochastic process input. Periodicity of 50 to 150 sample intervals is observed.

Also, the sample functions indicate that, for the T process, Q is small, being less than .50 and more likely .050. Likewise R is small due to the long averaging which produces each data point. However, it is significant to keep in mind that the initial values of Q and R are not very critical since the estimation equations will reduce their effect as time progresses.

2. Checking initial conditions.

Having selected a nominal set of initial conditions for Φ, Q and R, it is often worthwhile to use the time-varying model to check these initial conditions. For example, if the results show a decreasing driving noise covariance Q but the data does not support such behavior, then the initial value of Q was too high. Iteration of this procedure easily determines reasonable values for initial conditions.

3. Inspection of filtered process.

The filtered process is the output $\hat{x}_{k|k}$ of the optimal filter. For a fixed model with the nominal values for Φ, Q, R, the filtered process should be a smoothed version of the observed data. If it is not very smooth but rather seems to follow the observed points, then the gain A is too high. If the filtered process is smooth but seems to lag the observed points, the gain A is too low. These

points are clarified by rewriting the filter equations (Eqs. 6.5 and 6.6) to get

$$\hat{x}_{k|k} = (I - A_k H_k)\, \Phi_k\, \hat{x}_{k-1|k-1} + A_k\, y_k$$

4. Performance.

This step will indicate the quality of the model by evaluating the performance indices J0, J1, ..., JN defined in Chapter 3. Ideally, $\tilde{x}_k = x_k - \hat{x}_k$ should be used in this evaluation, but of course x_k is not available; $y_k = x_k + v_k$ is observed. Therefore $\tilde{y}_k = y_k - \hat{y}_k = y_k - \hat{x}_k$ is used and consequently a lower bound for J0 is $\bar{P} + R < Q + R$. The estimation error variance can not be any less than the sum of the forcing process variance and the observation process variance.

This performance index might be evaluated for several sets of values for Φ, Q, R. An improved fixed model (time-invariant) will often be evident. However, little effort is warranted to accomplish this, because the time-varying model should exhibit considerable improvement.

5. Time-varying model.

Having established a set of initial conditions for Φ, Q, R, inspected the filtered process and evaluated the performance, the time-varying model can be applied to improve the performance. The parameters Φ, Q, R will be "tracked" by the estimation scheme

and better performance at all times k should be expected. The reasons for not getting a better performance at a particular time k will be discussed for each process under study.

CHAPTER 7

MODEL APPLIED TO THE T PROCESS

Synopsis

7.1 Introduction

7.2 T Process Sample Functions

7.3 One-State Model, n = 1

- Coefficients in the estimators
- Initial conditions
- Filtered process
- Performance results

7.4 Modified Two-State Model, n = 2

- Coefficients in the estimators
- Initial conditions
- Filtered process
- Performance results

7.5 Comments on the T Process

7.1 Introduction

The T-process, as defined in Section 6.4.3, is a broadband signal power measurement expressed in db. For the reasons given in Chapter 1, the Gauss-Markov conditions are reasonable to assume for this process. The questions to answer in this chapter are the following:

1) What minimum n-state model is adequate for the process?

2) Is the time-varying model better?

3) What are the performance indices?

7.2 T Process Sample Functions

To illustrate the T process, a few sample functions have been plotted. Figures 7.1 a), b) each represent about two hours of observations and Fig. 7.1 c) is about four hours of observations. These sample functions are exactly those produced by the data acquisition program called "Dual". Observe that the process is considerably over-sampled. By taking only 1 of every 5 data points, the sample function of Fig. 7.2 results. This is about 18 hours of data with samples spaced about 4.3 minutes apart.

7.3 One-State Model, n = 1

The first study is for the scalar case n=1 and all equations are scalar. This is a strong simplification, but is justifiable on three points. It is the elementary starting place to study the models

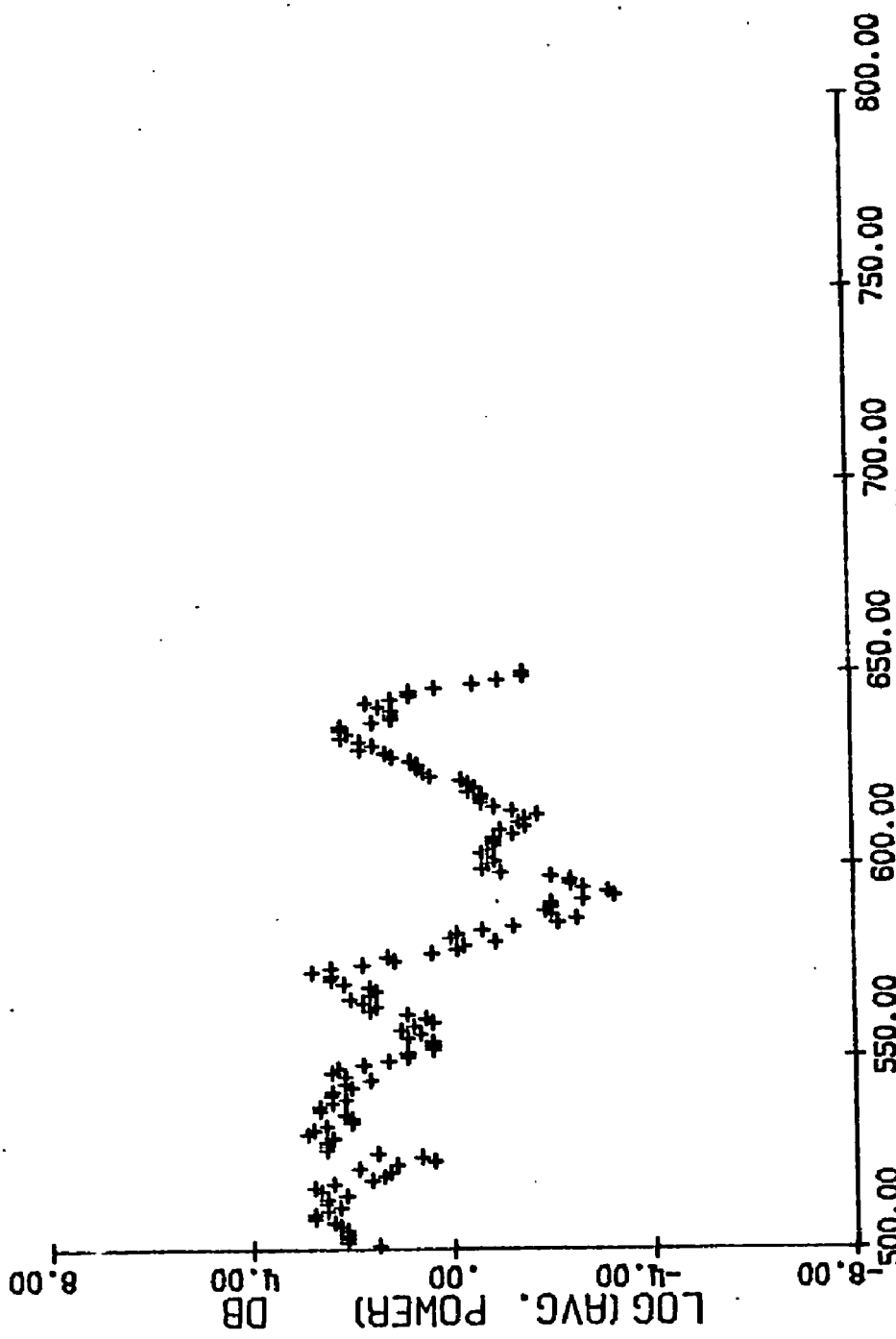


Fig. 7.1. a) T process sample function

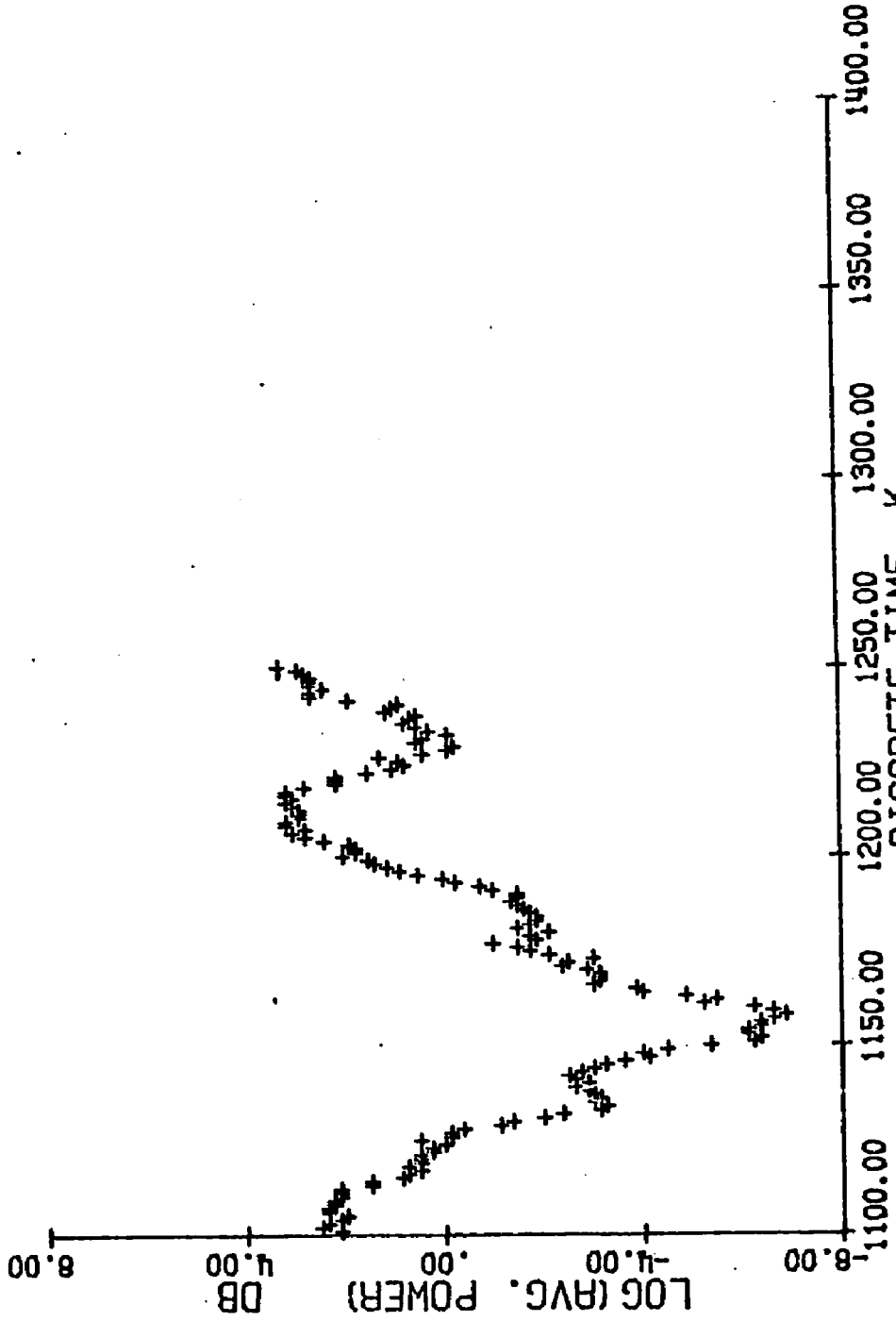


Fig. 7.1. b) T process sample function

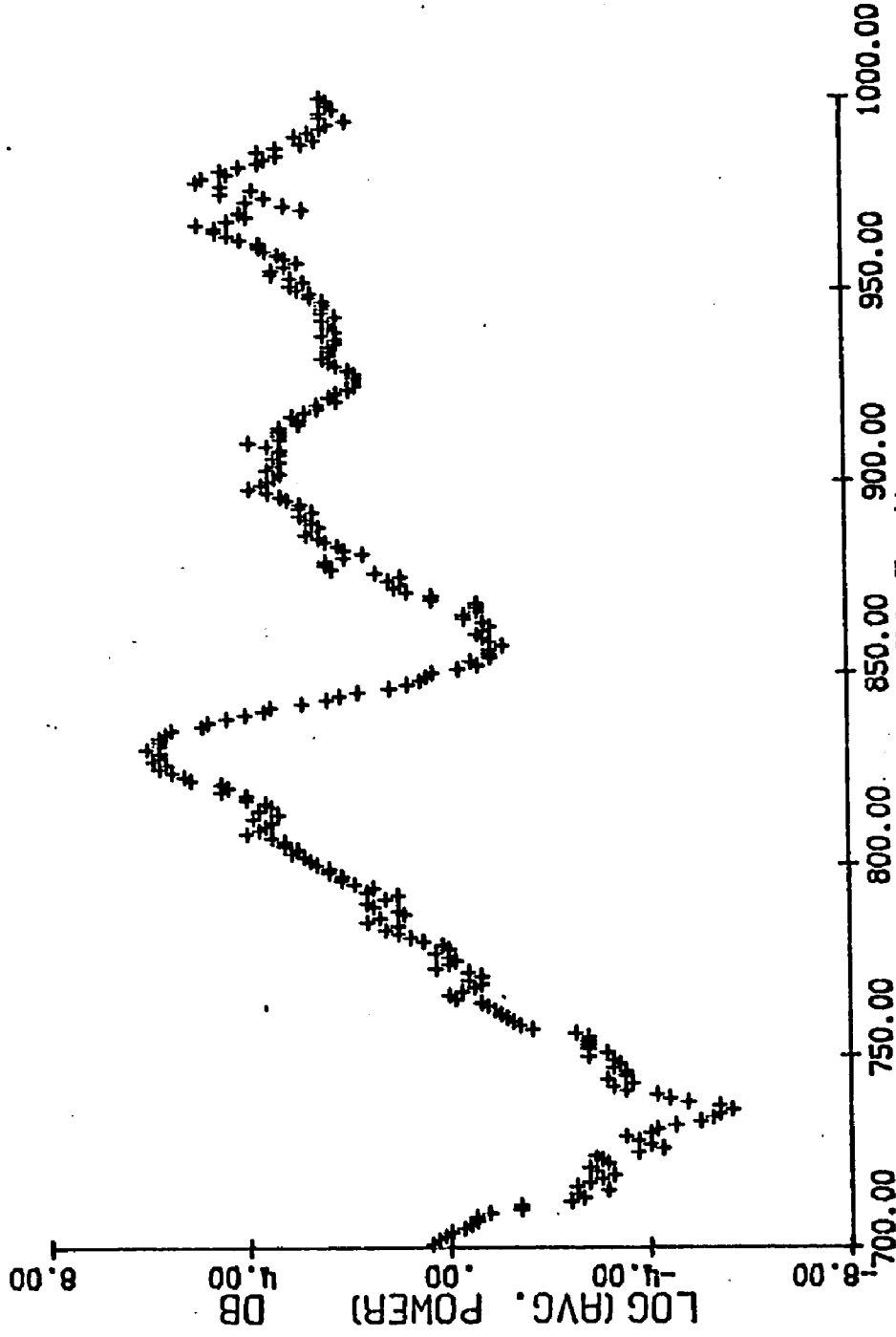


Fig. 7.1. c) T process sample function

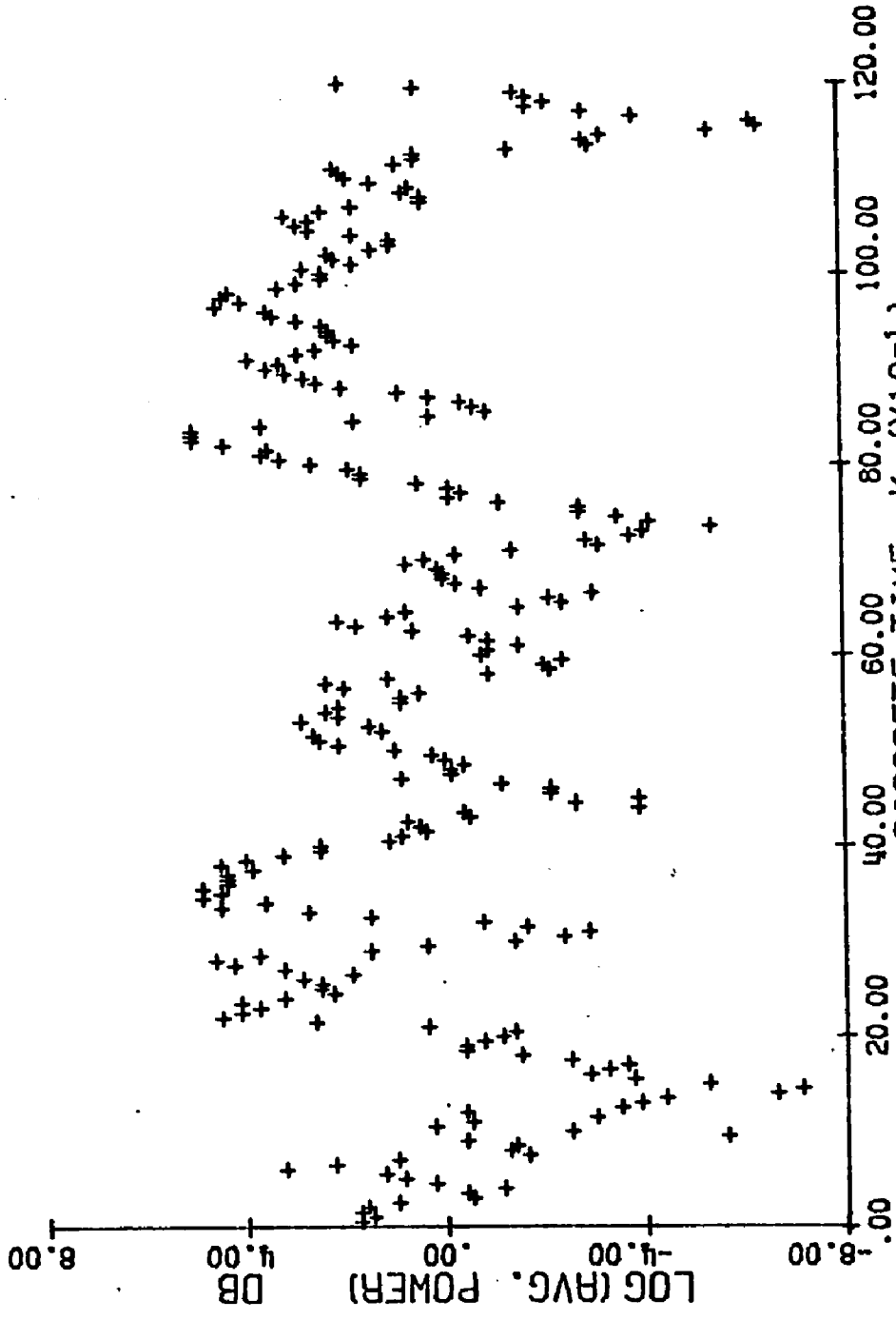


Fig. 7.2. T process sample function, one of every 5 data points

under field test. The observations y_k are often available only in scalar form; other state variables are often difficult or impossible to observe or very noisy. Thirdly, the scalar models can be expected to have acceptable performance for some processes, but if not, they will indicate which higher-order models to consider. This has been the case in this study, as the sequel will show.

7.3.1 Coefficients in the Estimators. To use the estimation scheme C, the coefficients c_{k-j}, d_{k-j}, e_{k-j} must be chosen. To this point, a wide selection of statistics is possible by choice of these coefficients. However, the coefficients are now constrained to make the algorithm simple (for ease of programming, to conserve computer core space and to yield economy in run-time). The constraint is that they be powers of a fixed number less than 1; see Chapter 5 for $r < 1$.

$$c_{k-j} = r_{\phi}^{k-j} \qquad r_{\phi} < 1 \tag{7.1}$$

$$d_{k-j} = r_0^{k-j} \qquad r_0 < 1 \tag{7.2}$$

$$e_{k-j} = r_1^{k-j} \qquad r_1 < 1 \tag{7.3}$$

r_{ϕ} is associated with the Φ estimate, r_0 with B_0 and r_1 with B_1. Then for each k, k-j measures time into the past and

$$c_0 = 1, \quad c_1 = r_\phi, \quad c_2 = r_\phi^2, \quad \ldots \tag{7.4}$$

$$d_0 = 1, \quad d_1 = r_0, \quad d_2 = r_0^2, \quad \ldots \tag{7.5}$$

$$e_0 = 1, \quad e_1 = r_1, \quad e_2 = r_1^2, \quad \ldots \tag{7.6}$$

and the computations of the statistics are greatly simplified.

$$S_{12,k} = r_\phi S_{12,k-1} + y_{k-1} y_{k-2} \tag{7.7}$$

$$S_{02,k} = r_\phi S_{02,k-1} + y_k y_{k-2} \tag{7.8}$$

$$B_{0,k}/b_0 = r_0 B_{0,k-1}/b_0 + (y_k - \hat{\Phi}_k y_{k-1})(y_k - \hat{\Phi}_k y_{k-1}) \tag{7.9}$$

$$B_{1,k}/b_1 = r_1 B_{1,k-1}/b_1 + (y_k - \hat{\Phi}_k y_{k-1})(y_{k-1} - \hat{\Phi}_{k-1} y_{k-2}) \tag{7.10}$$

The gain terms b_0 and b_1 are determined for large k.

$$b_0^{-1} = \lim_{k\to\infty} \sum_{j=3}^{k} d_{k-j} = \lim_{k\to\infty} \sum_{i=0}^{k} r_0^i = \frac{1}{1-r_0} \tag{7.11}$$

$$b_1^{-1} = \frac{1}{1-r_1} \tag{7.12}$$

Nominal values of r_ϕ, r_0 and r_1 can be selected by inspection of the sample functions. These coefficients will be refined during testing and will remain fixed when found.

A qualitative discussion may help to introduce these calculations. If the statistic were nearly constant, selecting r just below

1, at 1 - .001 for example, would yield a filter coefficient of .10 for data at time k - 2300. For r = 1 - .01, a coefficient of .10 is applied to data at k - 230. In either case, the statistic would be insensitive to observation noise and they would not follow local changes. On the other hand, if the statistic has significant time-variation, then r should not be close to 1. This would permit tracking of the variations but the estimates would be more in error (noisy). There must be a compromise between these two extremes.

The following discussion yields a nominal value for r_ϕ. The sample function in Fig. 7.2 is observed to have considerable "up-down" motion. For this model, any regularity in this behavior should be ignored, for the one-state model (n=1) can make use of only the observation, and not its derivative. Assuming that Φ_k is almost constant (steady), the value of r_ϕ should be very close to 1.0, so take a nominal value of

$$r_\phi = 1.0 - .001 = .999 \qquad (7.13)$$

If Φ_k should be allowed to be time-varying, then a smaller value of r_ϕ is needed. The shortest "ramp" in the data is of the order of 10 observation intervals. Take

$$r_\phi = 1.0 - \frac{1}{2 \times 10} = 1.0 - .05 = .95 \; . \qquad (7.14)$$

The calculations below give a nominal value for r_0 used in

calculating B_0. This statistic is based on difference terms of the form

$$y_k - \Phi_k y_{k-1} = u_k + v_k - \Phi_k v_{k-1} \quad . \tag{7.15}$$

Due to the dependence on v_k and v_{k-1}, these terms are not independent. However, since v is postulated as having smaller variance than u, approximate independence exists. Furthermore, this condition is used only in this calculation of r_0 and an upper bound will result.

A digression to recall the chi-squared distribution, denoted by χ^2, and its properties is necessary.

Theorem

Let $x_1, x_2, \ldots, x_n$ be a random sample from a normal distribution with mean 0 and variance σ^2, often denoted as $N(0, \sigma^2)$. Then

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i^2 \tag{7.16}$$

has a chi-squared distribution with n degrees of freedom.

The estimator $A_{0,k}$ for A_0 is of this form and although $B_{0,k}$ is not exactly χ^2, its behavior will be grossly similar to χ^2. Based on chi-squared properties, a value of n will be chosen from which r_0 will be evident.

Given a random sample of size n from a normal population with mean = 0, the maximum-likelihood estimator for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 = \frac{\sigma^2}{n} \chi^2 \tag{7.17}$$

The density for the estimator is therefore

$$f(\hat{\sigma}^2) = \frac{1}{[n|2-1]!} \left(\frac{n}{2\sigma^2}\right)^{n|2} (\hat{\sigma}^2)^{(n|2)-1} e^{-n\hat{\sigma}^2/2\sigma^2} \tag{7.18}$$

$$\text{for } \hat{\sigma}^2 > 0 .$$

A confidence interval with level .80 means that the true value of the parameter being estimated will be in this interval with probability .80. We arbitrarily seek n to make the confidence interval shorter than $.4\,\hat{\sigma}^2$. Using the Handbook (Ref. 6) and the condition that n is large, the results are tabulated in Table 7.1.

On the basis of the length of the confidence interval, n = 90 will suffice.

To get the corresponding r_0, gains are equated.

$$\frac{1}{n} = b_0 = 1 - r_0 \tag{7.19}$$

$$r_0 = 1 - \frac{1}{n} = .989 \tag{7.20}$$

This is the nominal value for r_0 but it should not be larger than this.

n	$\chi^2_{.90}$	$\chi^2_{.10}$	$(n/\chi^2_{.10},\ n/\chi^2_{.90})\ \hat{\sigma}^2$ 80% Confidence Interval	Length
40	29.0505	51.8050	(.772, 1.375) $\hat{\sigma}^2$	.603 $\hat{\sigma}^2$
50	37.6886	63.1671	(.792, 1.325) $\hat{\sigma}^2$	.533 $\hat{\sigma}^2$
60	46.4589	74.3970	(.807, 1.290) $\hat{\sigma}^2$	.483 $\hat{\sigma}^2$
70	55.3290	85.5271	(.818, 1.265) $\hat{\sigma}^2$	.447 $\hat{\sigma}^2$
80	64.2778	96.5782	(.829, 1.242) $\hat{\sigma}^2$	.413 $\hat{\sigma}^2$
90	73.2912	107.565	(.836, 1.228) $\hat{\sigma}^2$	.392 $\hat{\sigma}^2$
100	82.3581	118.498	(.844, 1.215) $\hat{\sigma}^2$	.371 $\hat{\sigma}^2$

Table 7.1. Confidence intervals

By reducing r_0, better tracking of the parameter may result, but a larger confidence interval results for the same level of .80.

Such a simple procedure is not possible for r_1, because B_1 is a sum of product pairs, not a sum of squares. Hence $B_{1,k}$ has increments of about the same magnitude as has $B_{0,k}$, but clearly they may be positive or negative. So $|B_{1,k}|$ is smaller than $B_{0,k}$ and has larger variance. A larger sample size n (or r_1 closer to 1) is required. A development of the density for B_1 and a calculation of the confidence interval have not been fruitful. A nominal value is needed, so a sample size 5 times that for r_0 is used.

$$r_1 = 1 - \frac{1}{5(90)} = 1 - .0022 \qquad (7.21)$$

7.3.2 Initial Conditions. In Chapter 3, initial conditions for the Kalman-Bucy equations have been given in terms of Φ, Q, R. P_2 was determined from a steady state condition and $x_0 = 0$ was assumed. In Chapter 4, initial conditions on the estimators for Φ, Q, R have been expressed in terms of Φ, Q, R. Only initial conditions for Φ, Q, R are needed and this is the scalar case.

Viewed as a first-order Markov process, the sample functions of the T-process in Fig. 7.1 are not unlike those of the computer-generated sample functions in Fig. 6.5. No marked trends are apparent and the up-down motion might be attributed to the random nature of the received signal. By comparison with the sample functions of the Gauss-Markov processes in Section 6.3, the following nominal values are assigned to the initial conditions. Initial time is at $k = 0, 1, 2$, according to the notation of Chapter 4.

$$\begin{aligned} \Phi_0 &= \Phi_1 = \Phi_2 = .95 \\ Q_0 &= Q_1 = Q_2 = .50 \\ R_0 &= R_1 = R_2 = .05 \end{aligned} \qquad (7.22)$$

where the last choice of $R_2 = .10Q_2$ is quite arbitrary. These values may be changed as explained in Section 6.5 on experimental procedure.

7.3.3 Filtered Process. The filtered process is the output $\hat{x}_{k|k}$ produced by the Kalman filter. Figure 7.3 shows the result for the model being studied. Observe that $\hat{x}_{k|k}$ follows the observed

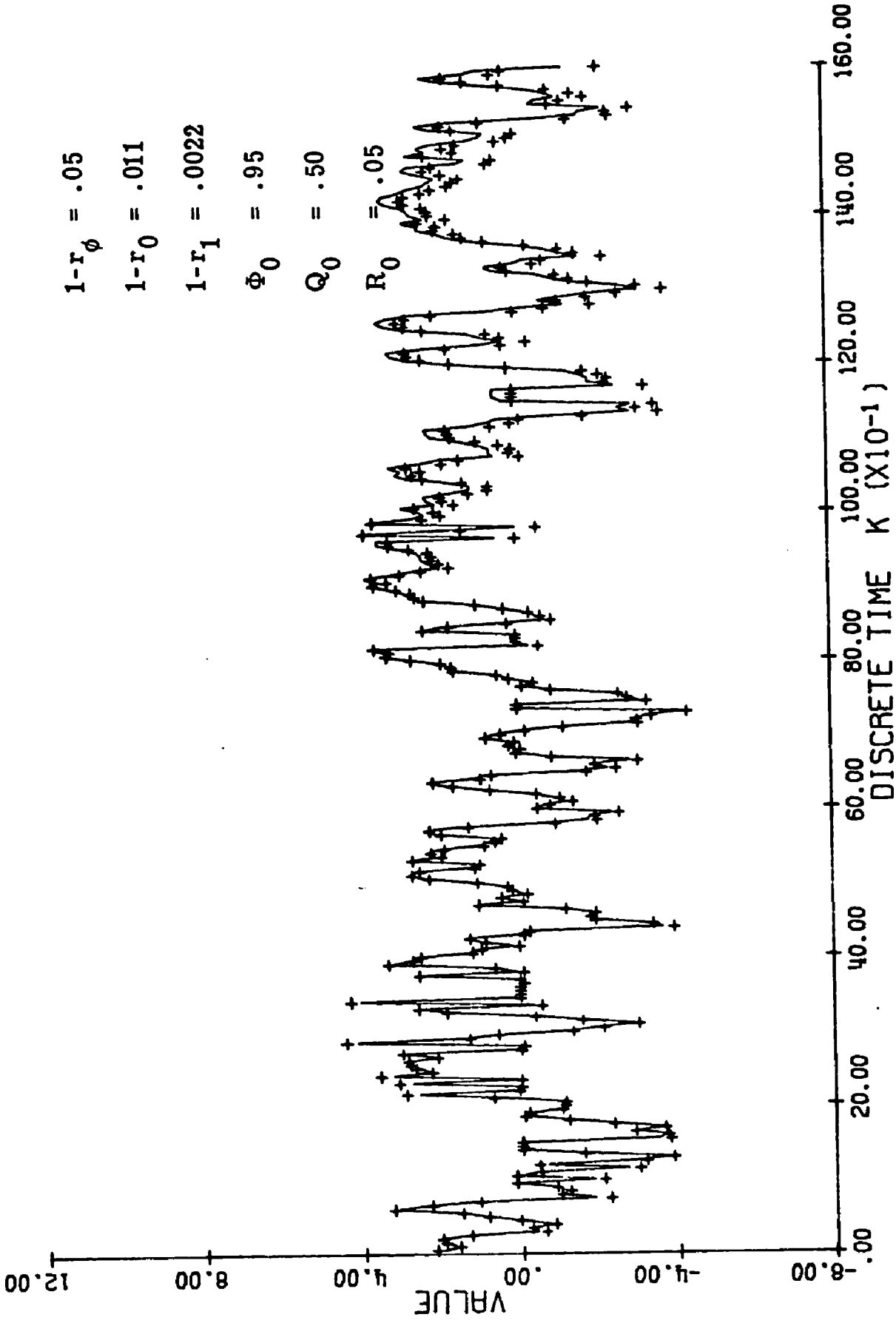


Fig. 7.3. T process: estimate $\hat{x}_{k|k}$ for n = 1

process fairly closely but that there is some smoothing. This is due to the high values of A_k as computed by the gain equation.

For comparison to other models below, the forecast $\hat{x}_{k+j|k}$ for $j = 1, \ldots, 20$ and for several k are given in Fig. 7.4. For $n = 1$, and for a stable system with $\Phi < 1$, the prediction is an exponential, decaying toward the mean of the process.

The corresponding series of $\hat{\Phi}_k$, $\hat{Q}_k$ and $\hat{R}_k$ are given in Fig. 7.5. These are for initial conditions of .97, .120 and .006 respectively. A_k is very close to 1.0, and so P_k is approximately $Q_k + R_k$, which is just slightly larger than Q_k, and is not plotted.

7.3.4 Performance Results. This one-state model has poor performance, as measured by several indicators. The filtered process seems to continually lag the observed process, for the simple model cannot follow rapid changes. The prediction error covariance P_k is small, being very close to its lower bound $Q_k + R_k$, but the performance index $J0_k$ is much larger than P_k. P_k and $J0_k$ are plotted in Fig. 7.6.

Consequently, the one-state model of the T process may be rejected. This model is not adequate as a smoothing filter, nor as a prediction filter (forecasting). The resultant sequences $\{\Phi_k\}$, $\{Q_k\}$, $\{R_k\}$ are therefore not very reliable. These conclusions are in part based on the significant improvements made by extending the model to a two-state model.

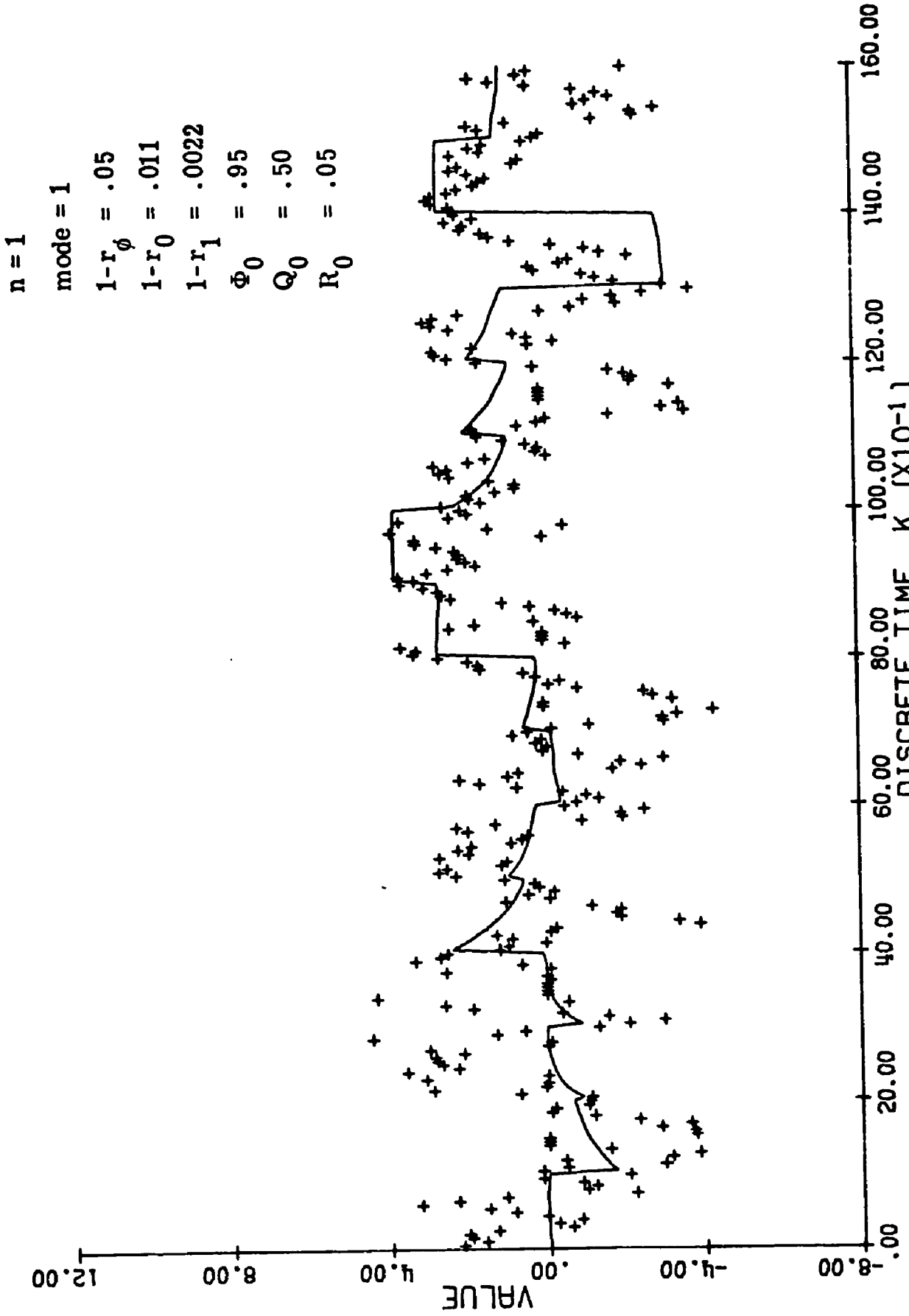


Fig. 7.4. T process: forecasts $\hat{x}_{k+j|k}$ for n = 1

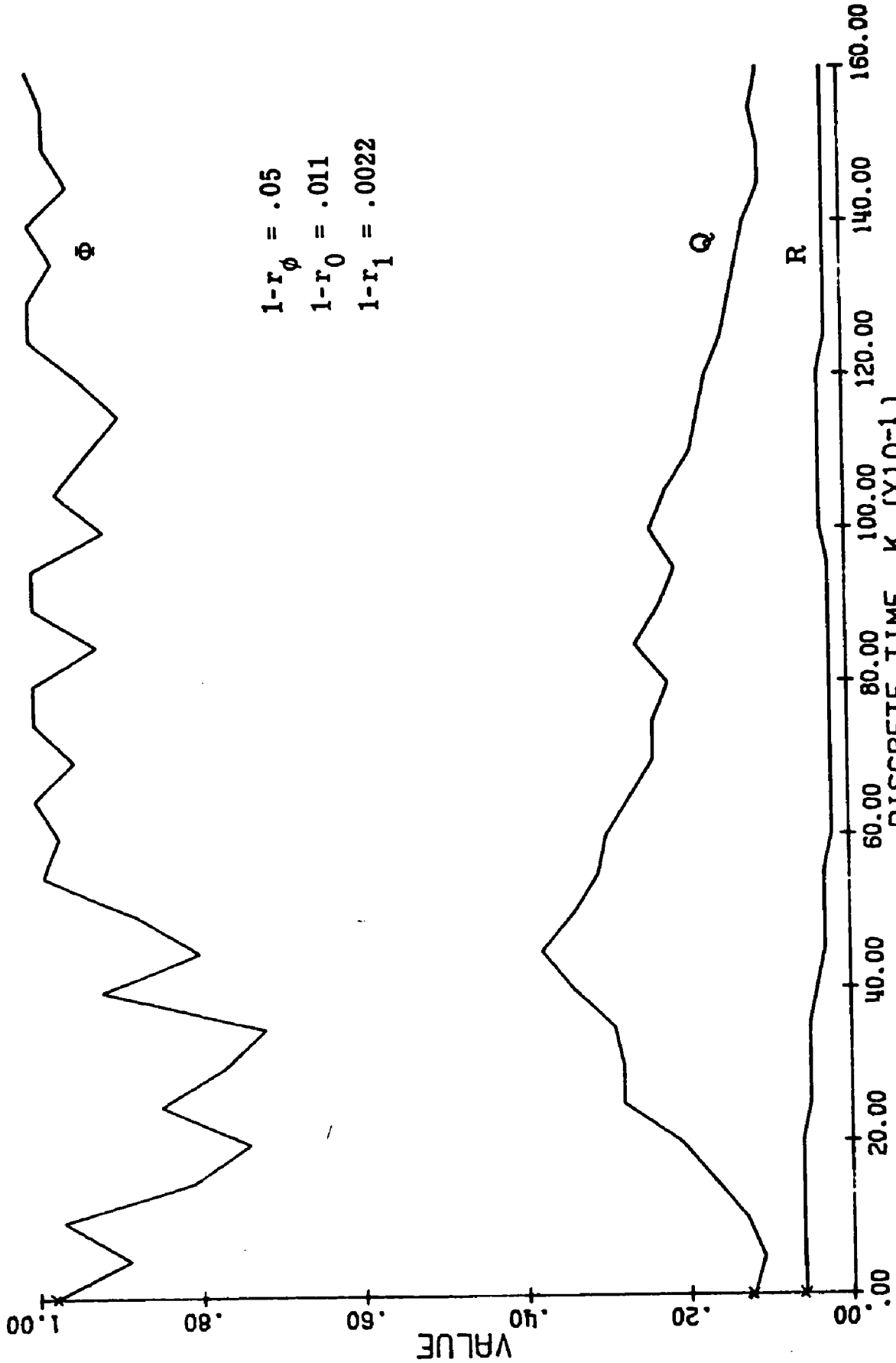


Fig. 7.5. Estimates $\hat{\Phi}_k$, $\hat{Q}_k$, $\hat{R}_k$ for n = 1

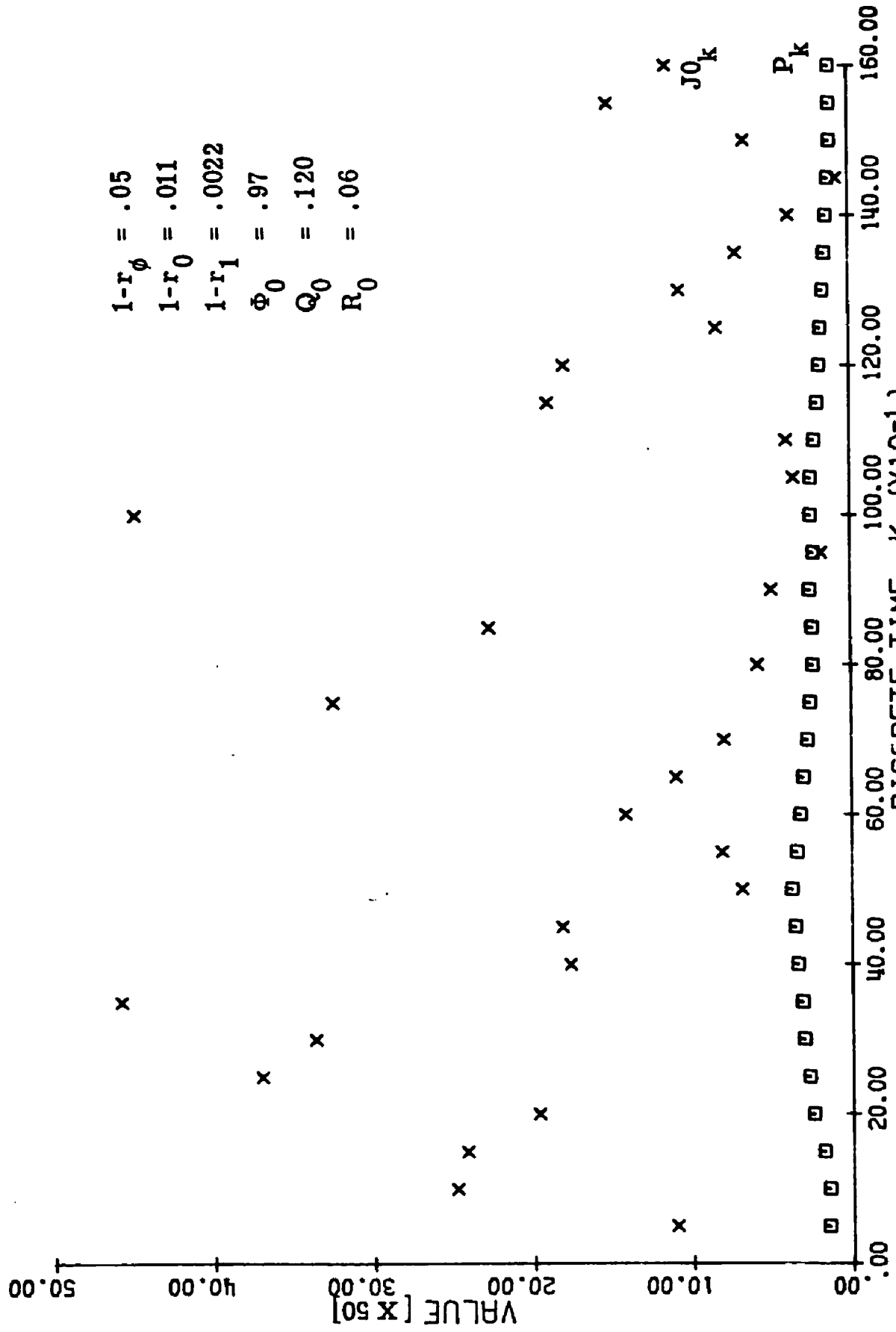


Fig. 7.6. P_k, JO_k for n = 1

7.4 Modified Two-State Model, n = 2

The model to be considered in this section is a two-state model from the optimal estimate vantage point, but is a one-state model from the parameter estimation viewpoint; hence the term modified two-state model. To clarify this modification, a table of variables and dimensions is given.

	Kalman Filter	Parameter Estimation
y_k	2 x 1	scalar $y_k^{(1)}$
$\hat{x}_{k\|k}$	2 x 1	
Φ_k	2 x 2	scalar ϕ_k
A_k	2 x 2	
H_k	2 x 2	
P_k	scalar	
Q_k	scalar	scalar
R_k	scalar	scalar
A_{11}	scalar	

Table 7.2. Dimensions in modified model

The estimation equation for $\hat{x}_{k|k}$ is 2-dimensional but other calculations are scalar. The first component of y_k is $y_k^{(1)}$ and is used ot obtain the scalars ϕ_k, Q_k, R_k.

The matrices Φ_k, H_k, A_k are obtained by the following

extension, which uses the approximation discussed in Section 2.3.3.

$$\Phi_k = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} = \begin{bmatrix} \phi_k & 1 - (1-\phi_k)/3 \\ -2(1-\phi_k) & \phi_k \end{bmatrix} \tag{7.23}$$

$$H_k = I(2 \times 2) \tag{7.24}$$

$$A_k = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \tag{7.25}$$

where $A_{11} = A_{22}$ is the gain from the scalar computation. These matrices are required in the filter equations (Eqs. 6.5 and 6.6).

7.4.1 Minimum Mean-Square Fit to Obtain Vector y. This section contains a general discussion of a technique which can be used to get multi-dimensional observations of a process and a particular scheme to get a 2-dimensional observation.

In general, the observation matrix H_k is $m \times n$, where m is the dimension of the observation y and n is the dimension of the process x. Usually $m \leq n$ and it is desirable to have $m = n$. But if the observations are restricted to $m = 1$, the processing can still be achieved; there are two possibilities.

i) Accept H_k as $1 \times n$. The Kalman equations are formulated for the general case, given Φ_k, Q_k, and R_k.

ii) Obtain the full observation vector y, where $m = n$ and H_k is $n \times n$.

As already pointed out in Chapter 4, the estimation schemes require an n-dimensional observation. Furthermore, to account for the oscillating tendencies of the process, higher-order models and observations are required. For these reasons, item i) is rejected and an algorithm is presented to obtain the two-dimensional y.

The algorithm is based on a minimum mean-square fit of a line to the scalar observations. It will be successful if the original observations can be taken at much higher rate than the minimum sampling rate. The result will be a higher-order observation taken at a slower rate. The details for the case $n = 2$ follow.

A minimum mean square (MMS) fit of a straight line to the data points is sought. Let the line be

$$\hat{y}_k = sk + \hat{y}_0 \tag{7.26}$$

$$\text{Error} = \sum_{k<0}^{0} w_k (y_k - sk - \hat{y}_0)^2 \tag{7.27}$$

where w_k is a weighting coefficient and y_k is the data point. The partial derivatives of this mean square error with respect to s and $\hat{y}_0$ are set to 0 to obtain two equations in s and $\hat{y}_0$.

$$\begin{bmatrix} \sum w_k k^2 & \sum w_k k \\ \sum w_k k & \sum w_k \end{bmatrix} \begin{bmatrix} s \\ \hat{y}_0 \end{bmatrix} = \begin{bmatrix} \sum w_k k\, y_k \\ \sum w_k\, y_k \end{bmatrix} \tag{7.28}$$

Select the number of data points as 5, the weights $w_k = 10+k$ for $-4 \leq k \leq 0$ (chosen rather arbitrarily); this corresponds to having observed the data point at $k = 0$ and the last 5 being fitted by a straight line. The required solution for s and $\hat{y}_0$ is readily accomplished.

$$s = .226\, y_0 + .087\, y_{-1} - .026\, y_{-2} - .113\, y_{-3} - .174\, y_{-4} \tag{7.29}$$

$$\hat{y}_0 = .645\, y_0 + .378\, y_{-1} + .155\, y_{-2} - .023\, y_{-3} - .155\, y_{-4} \tag{7.30}$$

As a check, observe that if the y_k sequence is a constant, then $s = 0$ and $\hat{y}_0 =$ constant. If the y_k sequence is a straight line, then indeed $s =$ slope of that line and $\hat{y}_0$ is the ordinate intercept.

To summarize, this processing of the field data can be diagrammed as shown. Henceforth, the observation vector is the output of such an algorithm.

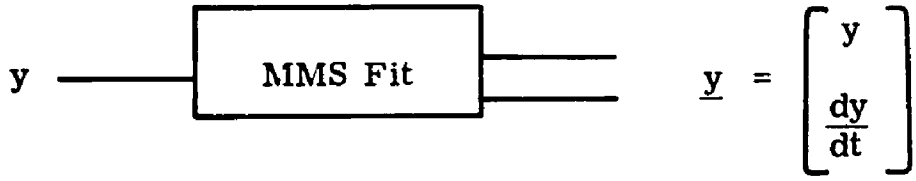


Fig. 7.7. Line fit for vector y_k

As a parenthetical note, numerical differentiation followed by averaging has also been investigated to achieve the higher-order observation vector, but the results were not satisfactory. The methods of numerical differentiation assume that the data is almost noise-free and that the process is a polynomial. However, the approach cannot be disregarded categorically and should be studied at greater length.

7.4.2 Coefficients in the Estimators. The coefficients r_ϕ, r_0, r_1 are needed for the estimators. The nominal values of r_0 and r_1 are as determined for $n = 1$.

$$r_0 = .989 \tag{7.31}$$

$$r_1 = .9978 \tag{7.32}$$

A nominal value for r_ϕ when $n = 2$ will be obtained. From the sample function in Fig. 7.2, a marked sinusoidal character is observed; a minimum period occurs at $k = 1050$.

sampling interval $= \Delta_s = 52$ sec.

minimum period $= 40\,\Delta_s$

maximum frequency $= f_m = \dfrac{1}{40\,\Delta_s} = .00048$ Hz

reciprocal time constant $= \dfrac{1}{\tau} = 2\lambda_m = 4\pi f_m = .00603$

$$r_{\phi} = 1 - \frac{1}{\tau} = .99397 \qquad (7.33)$$

Thus we have chosen the digital equivalent of a filter which has a single pole at $-\frac{1}{\tau}$, so that the gain of the double frequency term will be 3 db down. Returning again to the sinusoidal behavior, the following table lists some of the observed periods and shows ϕ.

Period = a	$\lambda = \frac{2\pi}{a}$	$\phi = \cos 5\omega$
50	.1255	.80920
75	.0838	.91349
150	.0418	.97824

Table 7.3. Observed periods

ϕ is discussed in Section 2.3.3 and $S = 5$ is the number of data samples used to get an observation vector y_k. These are trial values for ϕ in the modified second-order model.

7.4.3 Filtered Process. The filtered process $\hat{x}_{k|k}$ for this model is illustrated in Fig. 7.8. It is readily apparent that $\hat{x}_{k|k}$ is a smoothed version of the observed process. Figures 7.9, 7.10 and 7.11 are detailed illustrations of this filtering action of the Kalman filter.

7.4.4 Performance Results. The performance of this modified two-state model is substantially better than that of the one-state

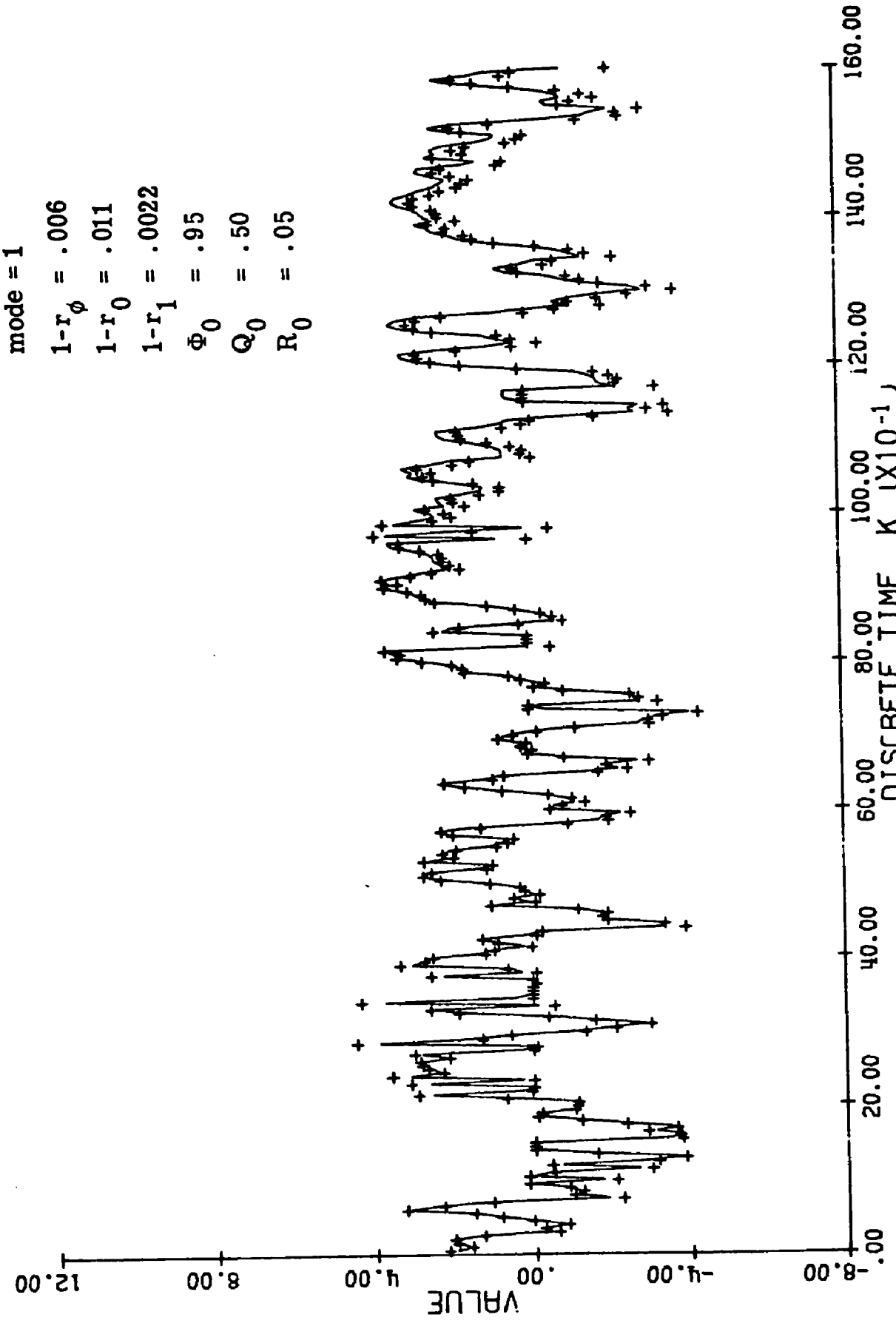


Fig. 7.8. T process: estimate $\hat{x}_{k|k}$ for n = 2

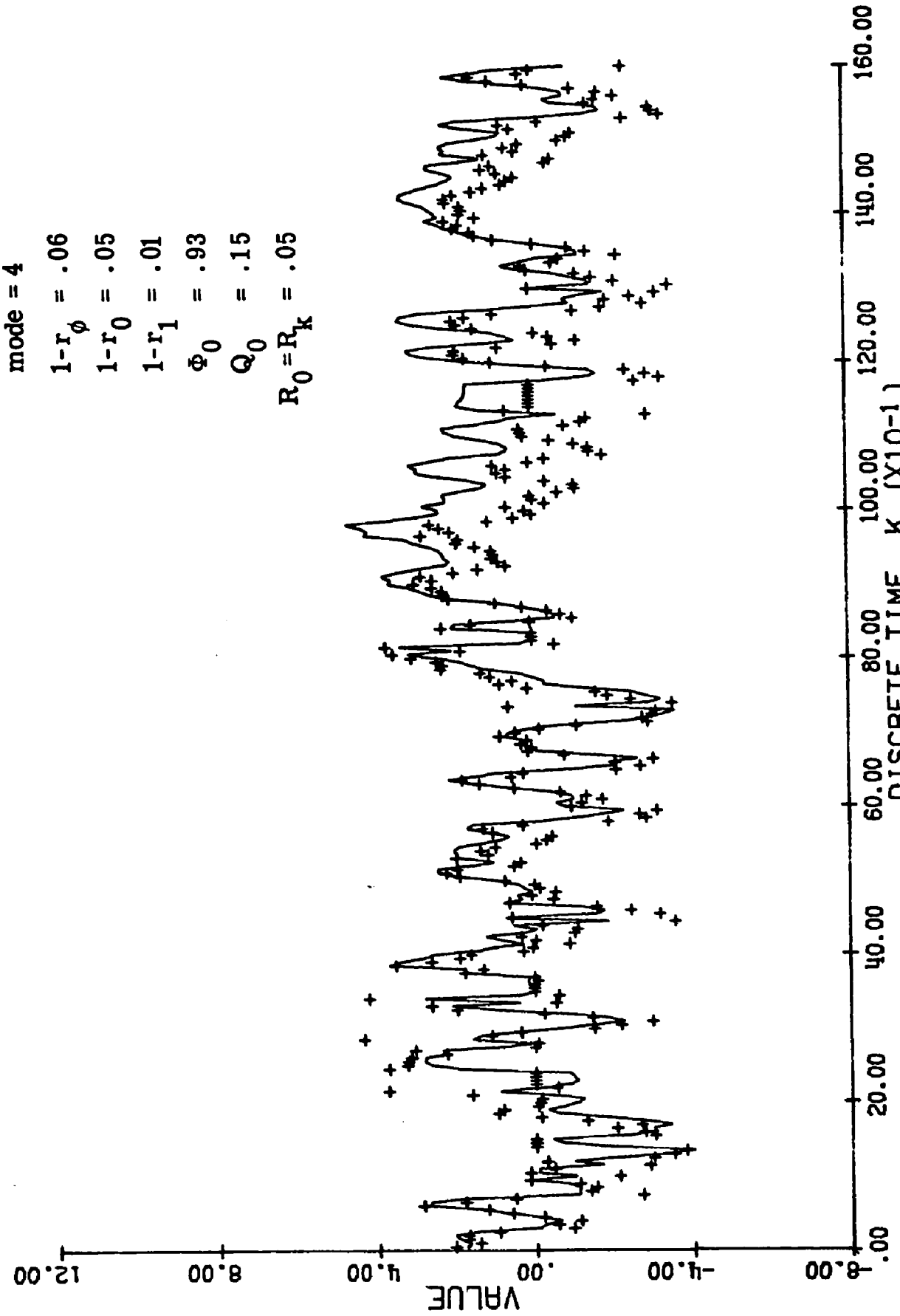


Fig. 7.9. T process: estimate $\hat{x}_{k|k}$ for n = 2

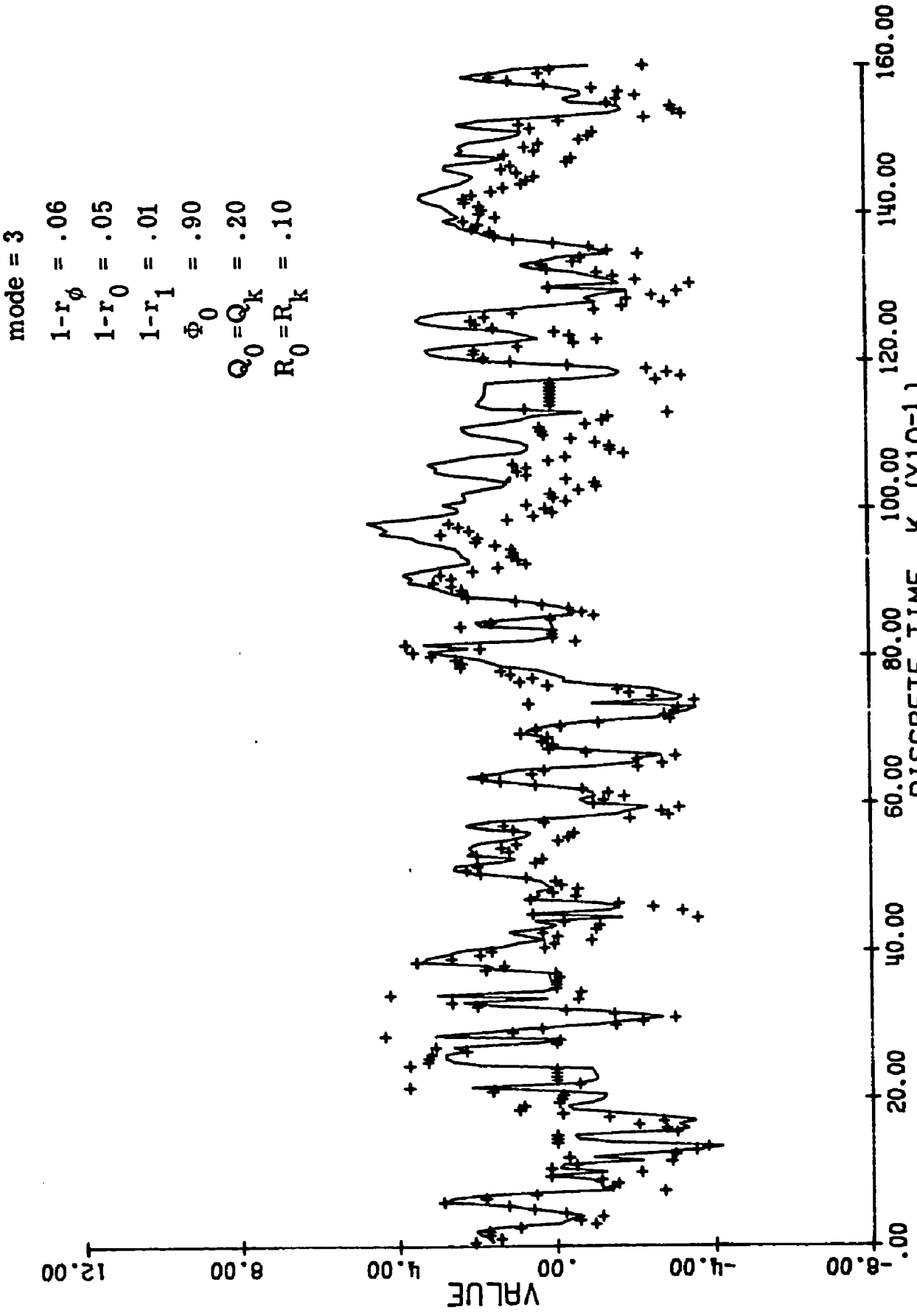


Fig. 7.10. T process: estimate $\hat{x}_{k|k}$ for n = 2

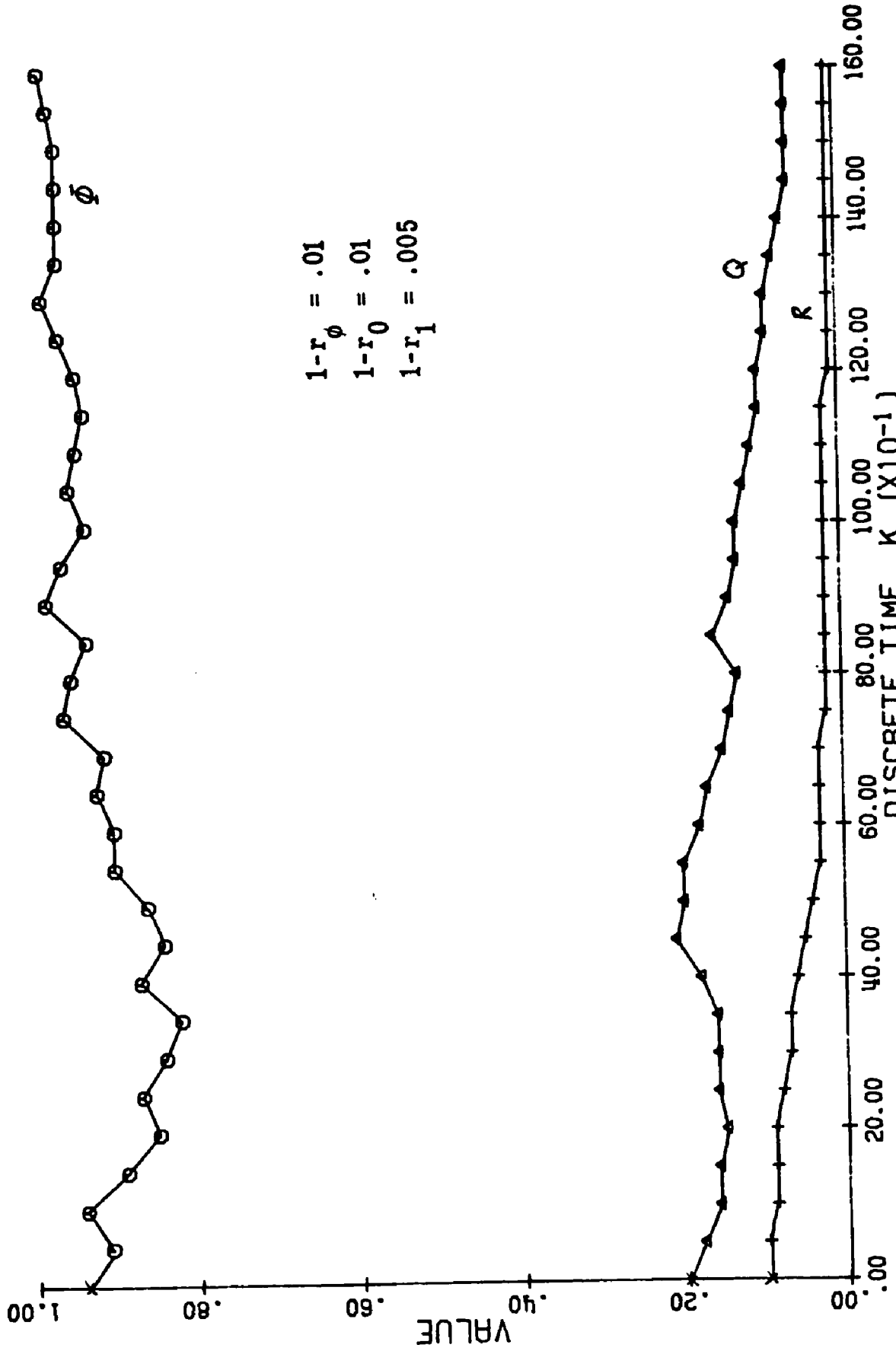


Fig. 7.11. Estimates $\hat{\Phi}_k$, $\hat{Q}_k$, $\hat{R}_k$ for n = 2

model. The illustrations of $\hat{x}_{k|k}$ have already made this clear in a qualitative way. In this section, the performance will be put to a quantitative test by inspection of the error covariance sequence $\{P_k\}$ and the performance index sequence $\{J0_k\}$.

The error covariance P_k is shown in Fig. 7.12 along with the performance index $J0_k$. P_k is the solution of the variance equation and depends on Φ_k, Q_k, and R_k but $J0_k$, defined in Eq. 3.52, is another measure of the mean square error between the observed and filtered process. The two statistics are not the same, but they both indicate mean quadratic properties, so it is not surprising to find them nearly the same.

A significant result which the model should produce is a reasonable forecast of the future values $\hat{x}_{k+j|k}$ of the process based on the present state (as estimated by $\hat{x}_{k|k}$) and the knowledge of the system parameters (as estimated by $\hat{\Phi}_k$, $\hat{Q}_k$, and $\hat{R}_k$). That such forecasting, or prediction, is possible and sensible is a fact which will be demonstrated now.

Given that $\hat{x}_{k|k}$ is the state of the process at the time k, then by the very definition of a state variable, subsequent states can be obtained from the system if the input is known. In the problem being studied, the input is a Gaussian white noise sequence with zero mean. A direct consequence is that the best forecast, in the minimum mean square sense, is the response of the system with initial

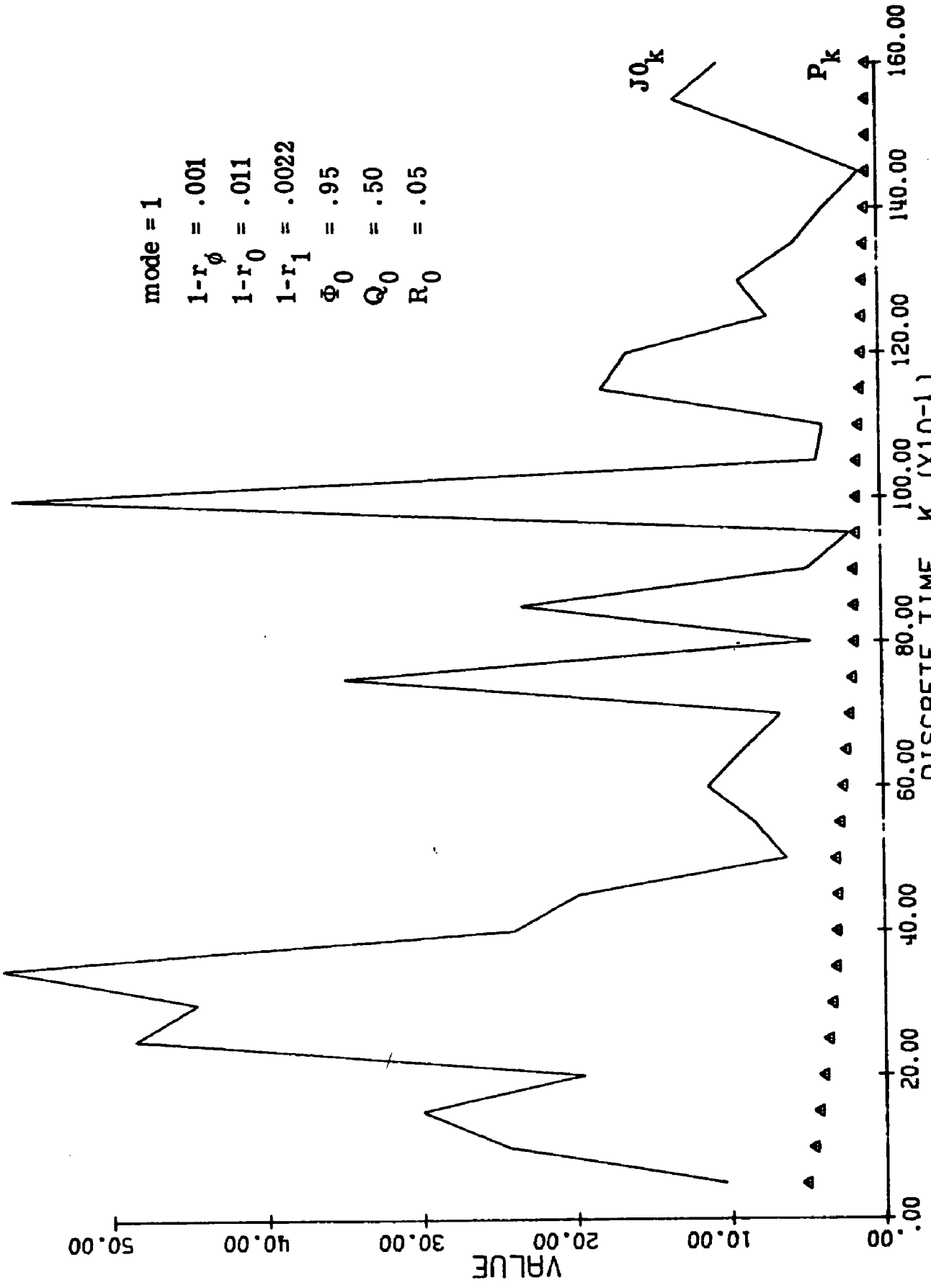


Fig. 7.12. Error performance P_k and performance index JO_k

state as $\hat{x}_{k|k}$ and no further input. In control terminology, the free response of the system from any time k with state $\hat{x}_{k|k}$ is the best prediction of future values of the stochastic process x.

There are two general sources of error in these forecasts. First, the state of the process is not known exactly, for $\hat{x}_{k|k}$ is an estimate, being the conditional mean using all past observations. Second, as hypothesized in this work, the underlying system Φ_k and the noise covariances Q_k and R_k are not known exactly. Such imprecise knowledge of Φ_k and its variation in time make forecasting far into the future very difficult.

However, for short-range forecasting such as 20 observation intervals, sensible results occur. Figure 7.13 is a composite of several such forecasts; the filter has been operating as a Kalman-Bucy estimator of x_k up to the left-most point of each curve; then a 20-point forecast is drawn. Observe that specific values are not predicted very accurately, but general trends are often correctly indicated. Unfortunately, not all such trends are predicted, for the model is only a second-order, undamped system.

7.5 Comments on the T Process

The T process is average acoustic signal power, measured in a fixed frequency band by a comb filter, and expressed in db., as defined in Section 6.4. A periodic signal with known frequency spectrum is presumably being transmitted. At the hydrophone, the

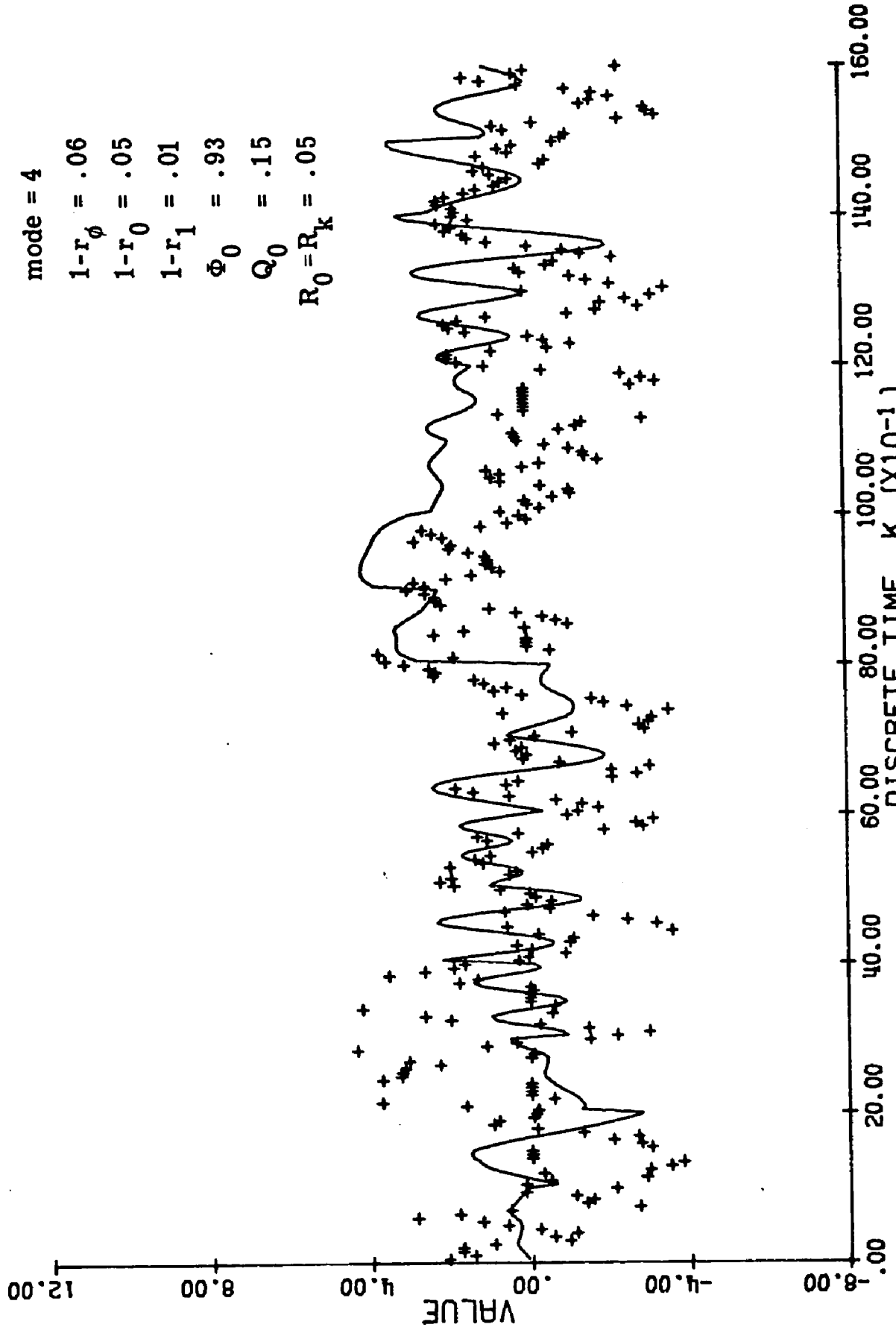


Fig. 7.13. T process: forecasts $\hat{x}_{k+j|k}$ for n = 2

received signal has impressed upon it the effects of the ocean, including propagation variations, surface effects and noise.

In modeling this T process, conditions assumed were that the process is Gauss-Markov and that a linear, stochastic model of order n is adequate. The results for $n = 1$ were not appealing, but the results for $n = 2$ and the model modified for the estimation equations, showed some interesting features of the process. The amount of data was insufficient for a definite conclusion, but indications are that the process is second-order stochastic. The time-varying model has been shown to be feasible; however, the resultant sequences $\{\hat{\Phi}_k\}$, $\{\hat{Q}_k\}$ and $\{\hat{R}_k\}$ of the model parameters have not been studied for physical significance. The capability of the model to forecast the future values of the process is noteworthy.

CHAPTER 8

MODEL APPLIED TO THE N PROCESS

Synopsis

8.1 Introduction

8.2 N Process Sample Functions

Shipping noise

Biological noise

8.3 One-state model, n = 1

Coefficients in the estimators

Initial conditions

Filtered process

Performance results

8.4 Comments on the N Process

8.1 Introduction

The N process, as described in Section 6.4.2, is a broad-band noise power measurement expressed in db. It is the output of the N portion of the Dual Comb Processing in Fig. 6.6. Unlike for the T process, no ready-made demonstration is known to the author to show that the N process is Gauss-Markov. The argument used for the T process is not applicable, because the process represented by $|N_x + j N_y|$ is not Gauss-Markov. However, the N process is assumed Gaussian on the grounds that it is a time-average power measurement of the process $|N_x + j N_y|$, which is itself the effect of a multitude of macroscopic causes. Concerning the Markov condition, a plausibility argument is that the N process samples represent a physical phenomenon and that these samples are close together in time. The physical phenomenon is called ambient noise, and has some finite bandwidth. The Markov-n condition is satisfied for some n. The model for $n = 1$ is used below and evaluated.

8.2 N Process Sample Functions

The N process is illustrated by the sample functions in Fig. 8.1; they are over the same time intervals as in Fig. 7.1. Cbserve that the process is somewhat over-sampled and that it has a much larger variance in comparison to the T process. Very interesting aspects of this process are the extremely large excursions from the mean value. These high noise power "pulses" are attributed to

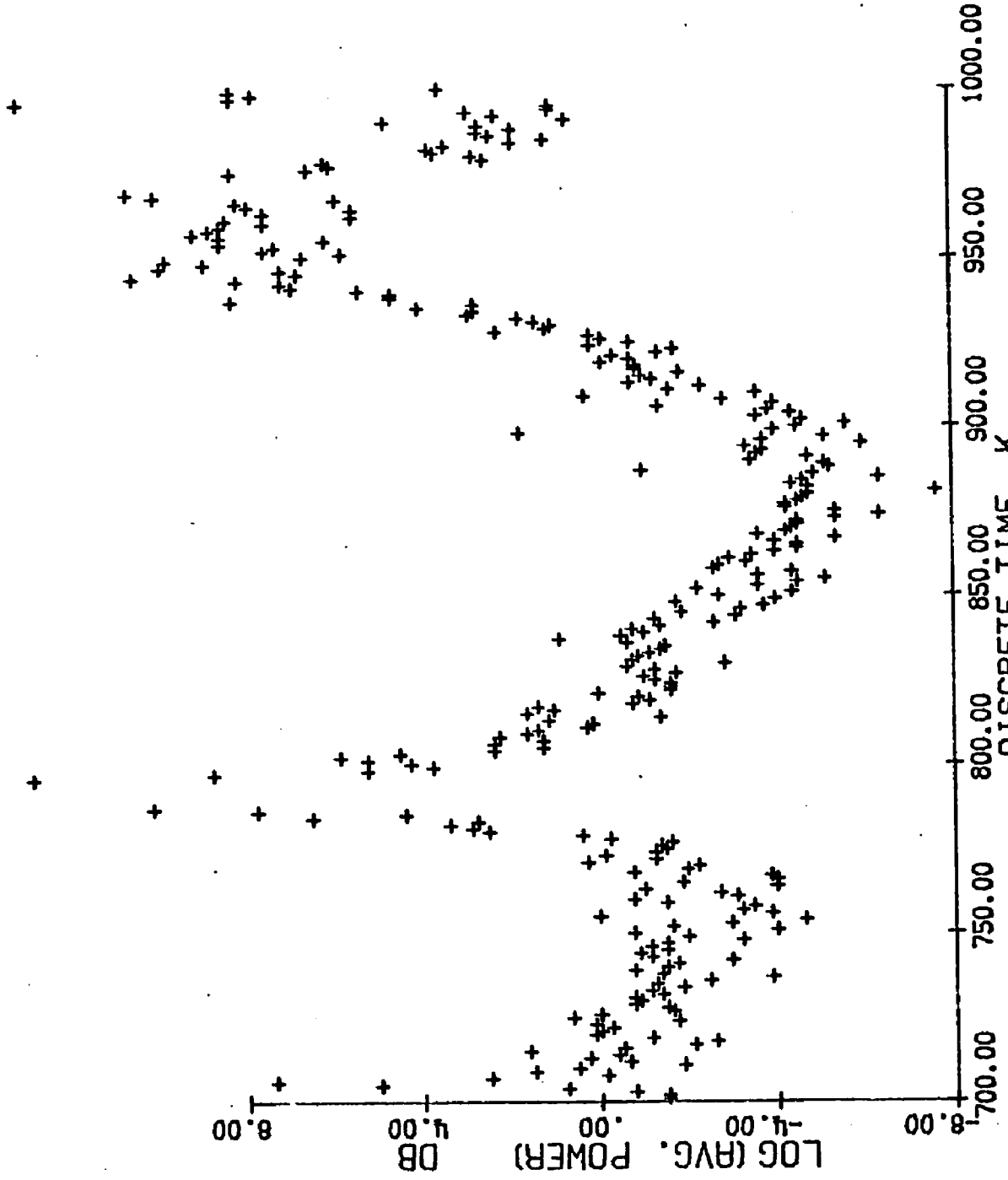


Fig. 8.1. N process sample function

shipping noise and biological noise; their duration is typically 15 to 20 minutes.

Shipping noise is a very strong contributing factor to the ambient noise in the Straits of Florida. When a ship is close to the (bottom-mounted) receiving hydrophone or when propagation conditions are such as to allow distant shipping to be heard, the noise level rises sharply, only to fall as the condition fades. The Dual Comb Processing has an analog prefilter which has a bandwidth of about 100 Hz centered at 420 Hz, and this is in the high, attenuated portion of the shipping noise spectra (Ref. 21). Low frequencies are much more predominant, so by listening to the total band, shipping can be identified. Positive, visual identifications were not made.

Biological noise is also a strong contributor. The swimming action of schools of fish, the feeding noises of masses of aquatic animals and the sounds emitted by some species are examples of noise sources. Usually, such sources just contribute to the ambient noise in the ocean, but occasionally a source is near the sensitive hydrophone or propagation conditions enhance a far-field source, so that a large increase in noise level results.

There are several other factors involved. Weather conditions affect the noise level directly and propagation conditions change due to surface effects. Strong correlation between wave heights and noise level have been reported under various conditions by many

researchers, as documented by Urick (Ref. 21). There are also micro-tremors in the ocean floor and major and minor earthquakes. This list is not intended as a catalog of noise sources, but rather as an indication of the ambient and transient sources contributing to ocean noise, which the N process measures in db.

8.3 One-State Model, n = 1

Only the simple one-state model, for which all equations are scalar, will be studied. Some of the comments made in Section 7.3 about this model for the T process apply for the N process as well.

8.3.1 Coefficients in the Estimators. The model uses the estimators of Model C, in Fig. 4.1, so that parameters for the nonlinear processor and the coefficients c_{k-j}, d_{k-j}, e_{k-j} must be chosen. The limiting values in the nonlinear processor are chosen as ± 4.0 db about the running average of the filtered data; inspection of the sample functions shows that this processor will eliminate most of the large sample values attributed to shipping or biological noise. It is not a purpose of this model to include such effects and their elimination is essential for the proper functioning of the estimators. The sample functions at the output of the nonlinear processor are illustrated in Fig. 8.2.

The coefficients are selected as in Section 7.3.1 as powers of $r < 1$. The nominal choices of r_ρ, r_0, r_1 are made intuitively

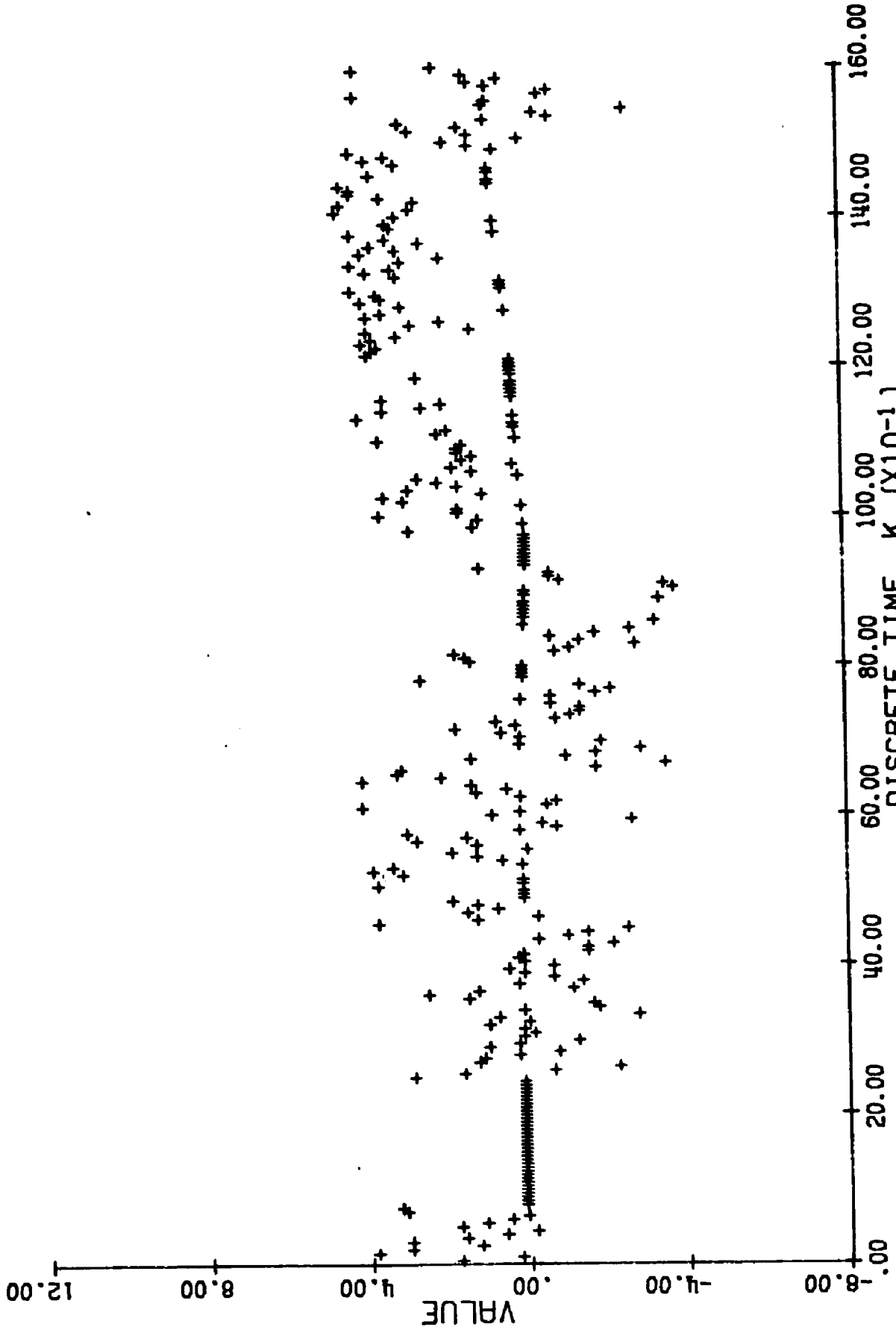


Fig. 8.2. Output of the nonlinear processor

and by inspection of the sample functions. For this study, Φ_k is assumed steady and a high value of $r_\phi = .99$ is taken. In Section 7.3.1, r_0 was chosen to put the noise statistic $B_{0,k}$ in a confidence interval with level .80; $r_0 = .989$ was the result. Here, to allow for some time-variation in $B_{0,k}$ but at the expense of a poorer estimate, $r_0 = .95$ is the selected value. As previously, the ad-hoc choice for r_1 is to have a sample size of 5 times that for r_0: $r_1 = .99$ is the result.

8.3.2 Initial Conditions. As discussed in Section 7.3.2, initial values for only Φ, Q, R are needed. The state transition matrix Φ is just a scalar and has the meaning of a correlation coefficient. An initial value of

$$\Phi_0 = \Phi_1 = \Phi_2 = .80$$

is assumed, because the N process is not as correlated as the T process. The observations y_k of the N process in Fig. 8.2 have an approximate variance of 4 db^2. If the observation noise variance is given an initial value of

$$R_0 = R_1 = R_2 = 1.0 \ ,$$

this allows a variance of 3 db for the random variable x_0: $\Lambda_0 = 3.0$. Then from Eq. 3.45,

$$Q_0 = Q_1 = Q_2 = (1 - \Phi_0^2)\,\Lambda_0 = 1.08 \ .$$

8.3.3 Filtered Process. The filtered process $\hat{x}_{k|k}$ is the output of the Kalman filter, and is illustrated in Fig. 8.3 for these coefficients and initial conditions; the parameters Φ, Q and R were held fixed. Consequently, the filter gain A is constant and has a value $A = .593$ computed by the equations. It is apparent that this filter does much smoothing. Figure 8.4 shows the results if the same initial conditions are applied, but in this instance the parameters Φ_k, Q_k and R_k are estimated and time-varying. There is not so much smoothing because A_k has an average value of about .85. The plots of $\hat{\Phi}_k$, $\hat{Q}_k$ and $\hat{R}_k$ are given in Fig. 8.5. Once again, the estimate for $\hat{R}_k$ seems to be low; selecting a constant value $R_k = 1.0$ and repeating, produces $\hat{\Phi}_k$ and $\hat{Q}_k$ in Fig. 8.6.

8.3.4 Performance Results. The prediction filter produces the forecasts $\hat{x}_{k+j|k}$ for the process x given observations up to time k. The noise process N is being modeled with a model of order one $(n = 1)$, so that the forecasts are exponentials decaying to the running average, as for the T process, $n = 1$ model. The results are illustrated in Fig. 8.7. Due to frequent shipping and biological noises, the performance indices P_k and JN_k apparently convey little information about model performance.

8.4 Comments on the N Process

Having applied the model to the N process and seeing the

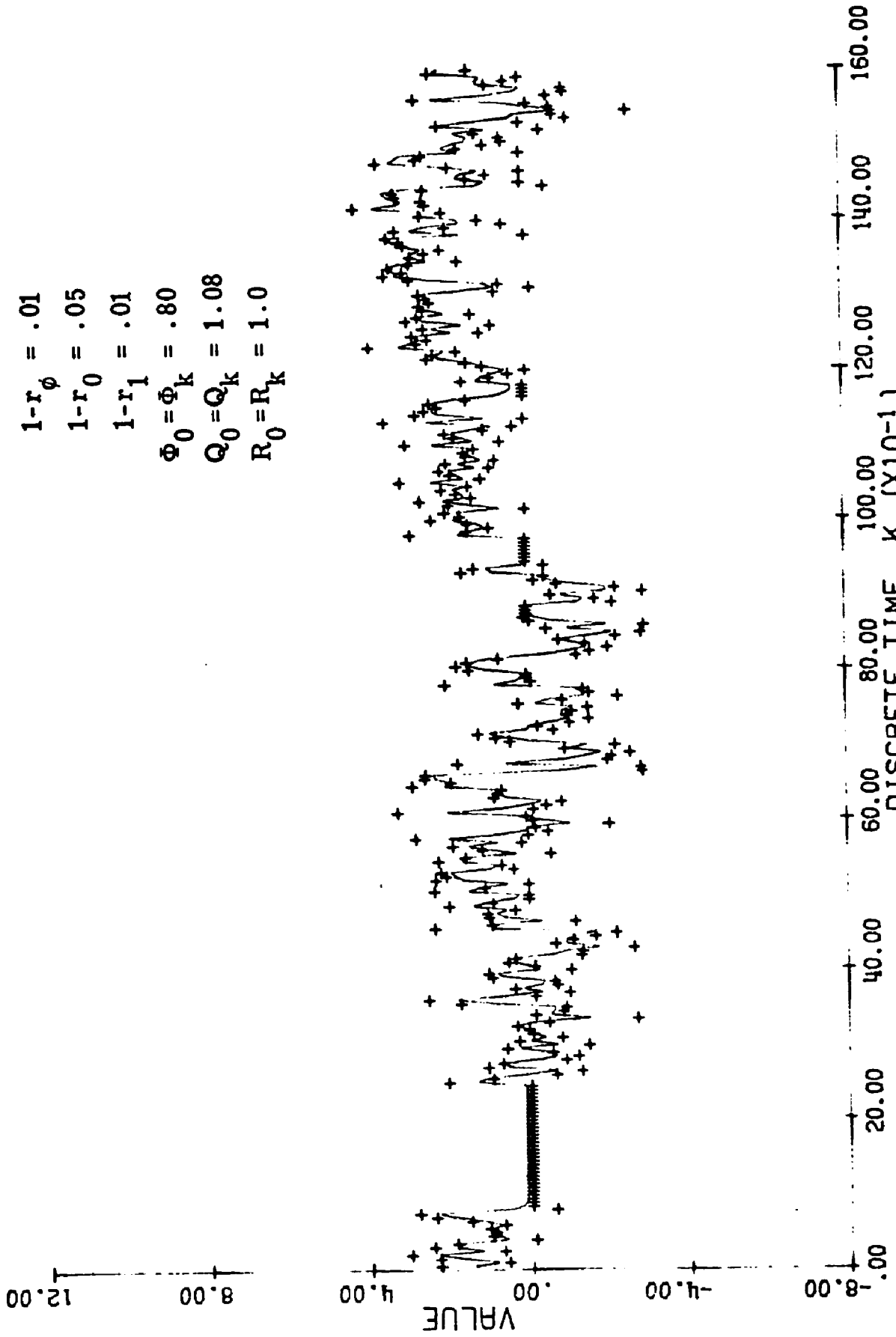


Fig. 8.3. N process: estimate $\hat{x}_{k|k}$ for n = 1, time-invariant model

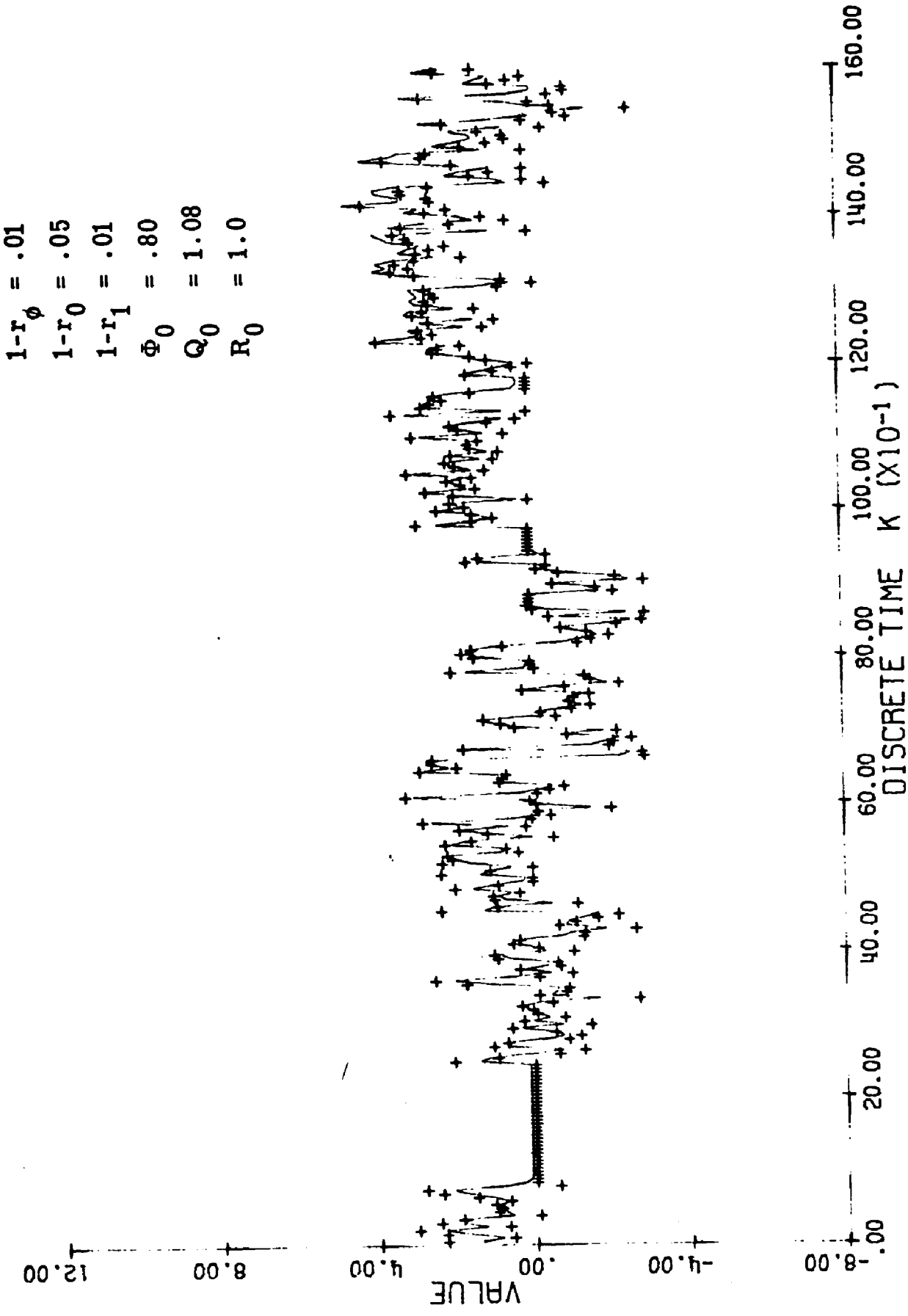


Fig. 8.4. N process: estimate $\hat{x}_{k|k}$ for n = 1, time-varying model

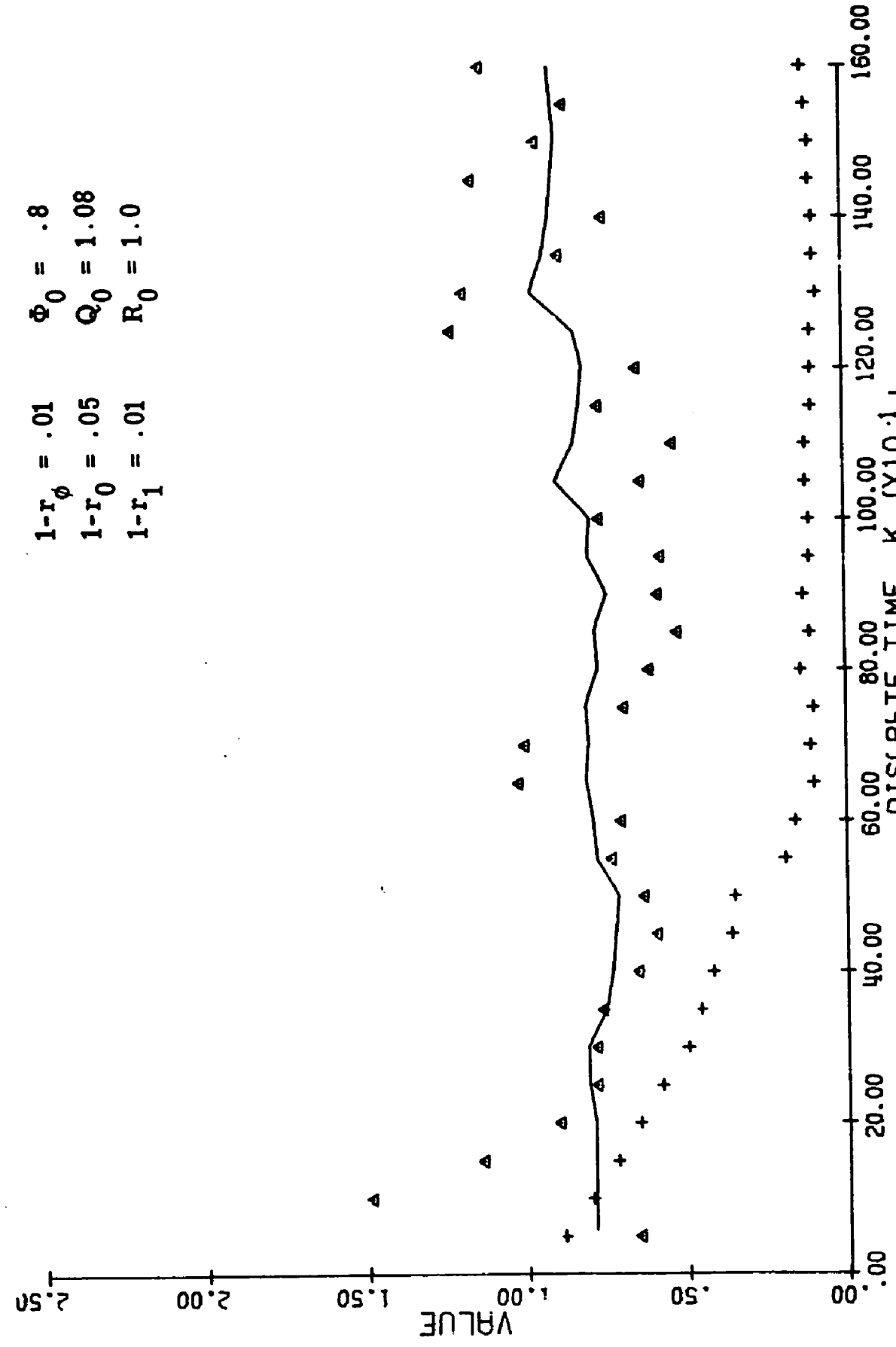


Fig. 8.5. Model parameters $\hat{\Phi}$, $\hat{Q}_k$, $\hat{R}_k$

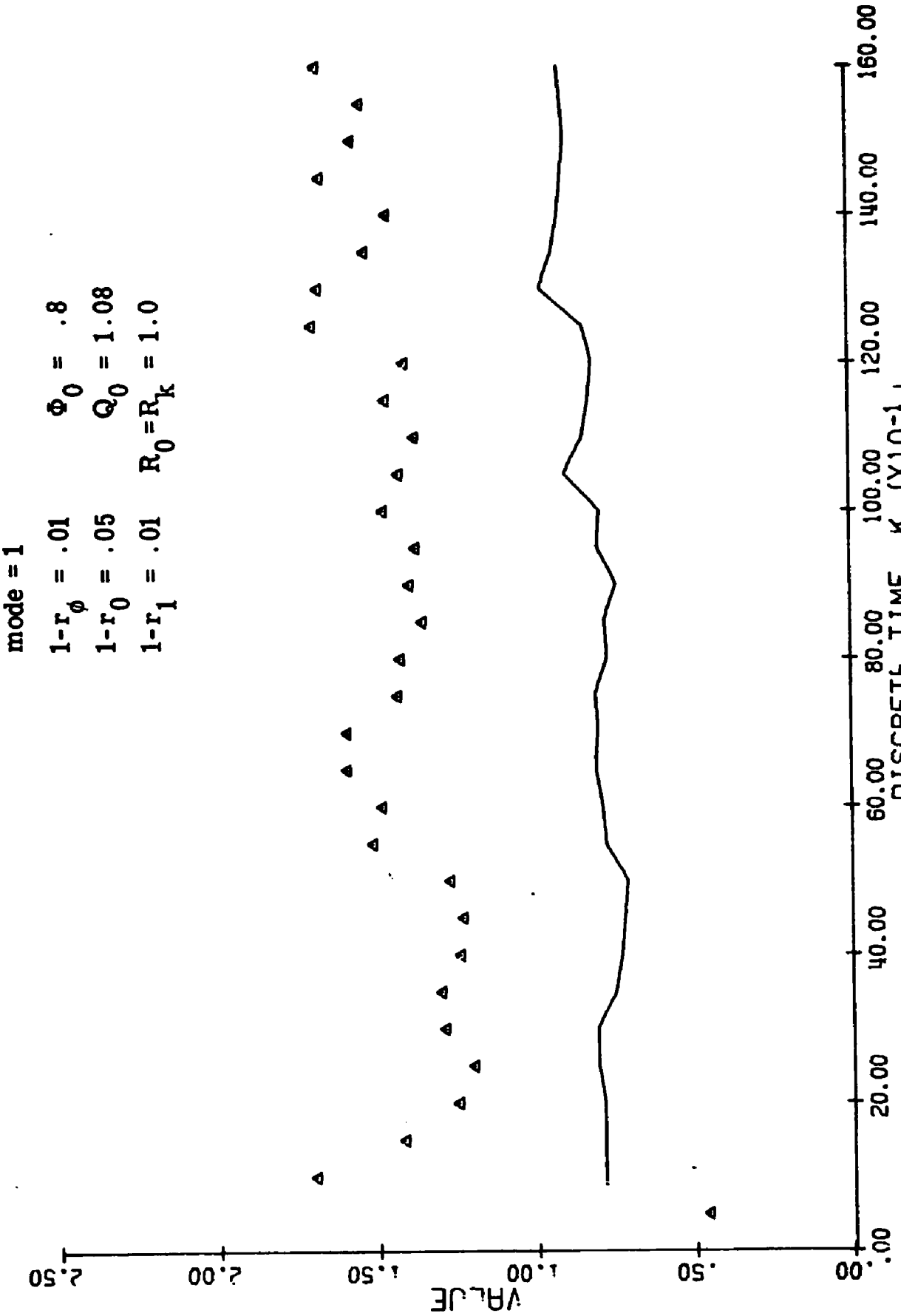


Fig. 8.6. Model parameters $\hat{\Phi}_k$, $\hat{Q}_k$ with $R_k = 1.0$

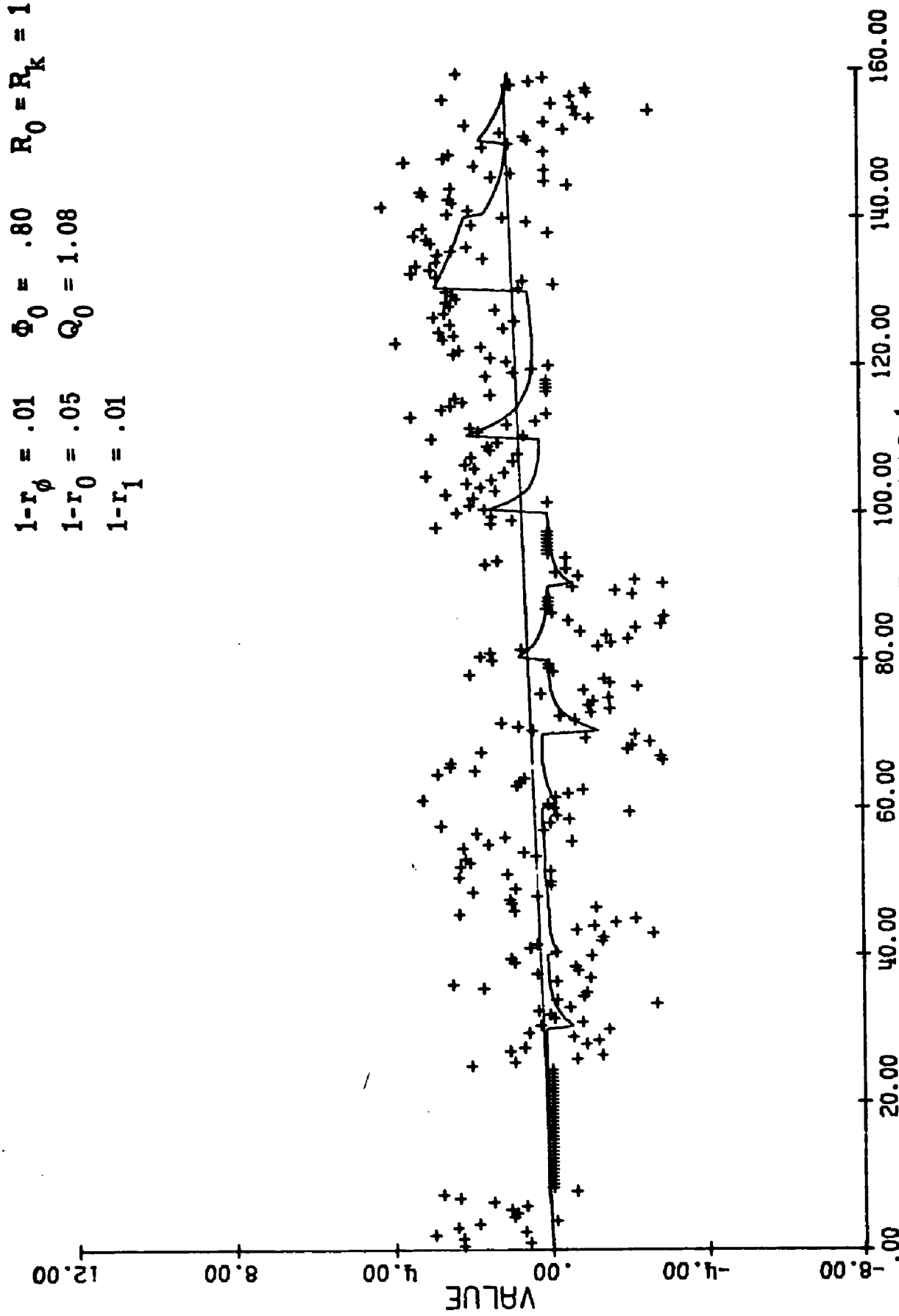


Fig. 8.7. N process: forecast $\hat{x}_{k+j|k}$ for n = 1

results, it is apparent that the model is valid. There is obviously correlation in the observations, and furthermore, the estimator for Φ shows that it is time-varying. The shipping and biological noise "pulses" detract from the effectiveness of the modeling of the actual ambient noise. The model does not, and could not have been intended to, model such "pulses". The absolute values of $\hat{Q}_k$, $\hat{R}_k$ are therefore in doubt, but their relative values and their time-variations can be regarded as correct. The smoothing of the data by the filter and the short-range forecasting of the process are acceptable.

CHAPTER 9

CONCLUSION

Synopsis

9.1 Summary of Results

Theory and methods

Data processing

Achieved objectives

Contributions

9.2 Future Work

9.1 Summary of Results

The major results and contributions of this thesis are summarized. The major purpose has been to study nonstationary stochastic processes. A general linear model has been given and under the Gauss-Markov condition, viable solutions to the filtering and prediction problems are well known. Since such a linear model has parameters which are usually unknown, a set of estimators was devised to obtain these parameters from the data sequence.

The data processing parts of the thesis include a general discussion of a class of linear filters and processors and a descriptive flow-diagram of the computer processing. The model was tested briefly with randomly generated data and applied to two real acoustics processes. The highlights of each of these items are reiterated below.

9.1.1 Theory and Methods. In this study of nonstationary stochastic processes, strong emphasis has been placed on the nonstationary properties of the physical processes which have need of being investigated. Such time-variations of statistics are often exhibited by natural phenomena and it is only the exceptional process which is adequately described as stationary. The need for a stochastic study has been amply demonstrated by the two acoustics processes: acoustic signal power, the T process and acoustic noise power, the N process. Both are affected to varying degrees by winds, tides and

weather, by natural and man-made noises. Fortunately, this multitude of factors does not need to be described one by one, but rather their gross result is of interest to researchers in detection theory and communication theory.

This state of affairs has led many to model-building and this work has been no exception. For compelling reasons listed in Chapter 1, models for nonstationary stochastic processes were developed and studied. For mathematical tractability, the models are linear and the processes are constrained to be Gauss-Markov. Due to the stochastic nature of the model and the time-varying parameters, the state space approach is particularly suited to these problems and the n'th order model is readily treated. The model equations are given in Eqs. 6.1, 6.2, 6.3.

The filtering and prediction problems, stated in general form in Chapter 1, have viable solutions due to the three key conditions: linear, Gauss-Markov. The driving process u_k and the observation noise v_k are assumed Gaussian white noise sequences. By linearity, the output process x_k and observation process y_k are also Gaussian, so that the conditional probability density of the process x_k is determined by the conditional mean and conditional covariance. Coupled with the Markov condition, the solutions take the form known as the Kalman or Kalman-Bucy equations, as given in Eqs. 6.4 to 6.10.

Two significant observations have been made concerning these solutions. The process estimate $\hat{x}_{k|k}$ is a weighted sum of the previous estimate $\hat{x}_{k-1|k-1}$ and the present observation y_k; the optimal weighting is determined by the system matrices Φ_k and H_k and the optimal gain A_k. This has the practical implication that the filter is quite simple and does not have a growing memory. Secondly, the prediction error covariance P_k and hence the optimal gain A_k are not dependent on the observations; in some applications, advance computations are possible.

An important part of modeling is the determination of the system parameter matrices, such as the state transition matrix Φ, the noise covariances Q and R for this model. For the processes studied, little information exists to determine the parameters from physical laws. Parameter estimation was adopted and a set of consistent estimators was discussed. Simple modifications to these estimators were made in Chapter 4 to permit "tracking" of the time-varying parameters. The parameter estimation equations are given in Eqs. 6.11, 6.12 and 6.13.

9.1.2 Data Processing. Actual data was used in the experimental part of this thesis. Data acquisition, the orderly collection of considerable amounts of observations of the real world, is a very important prerequisite to a study such as this. The previous work of others (Refs. 7, 19 and 20) and the contemporary development of

the Dual Processing by T. G. Birdsall were the foundations of the acquisition of data used in Chapters 7 and 8. The Dual Processing of Fig. 6.6 uses a class of processors which were given the name MPR processors and treated at length in Chapter 5.

Two important features in the data processing should be emphasized. It was necessary to use a nonlinear processor to eliminate certain observations, such as those due to shipping in the N process study. The other feature is the use of curve fitting to obtain the vector observations y_k, as presented in Section 7.4.1. Full-dimensional observations are needed by the parameter estimators, yet the Dual Processing provides only scalar observations; the MMSE Linear Fit shown in Fig. 6.1 bridges this gap.

The model was applied to an acoustic signal process called the T process. A periodic pseudo-random signal at constant power level was transmitted from Fowey Rocks, near Miami. The received signal at the hydrophone, 7 miles east in the Straits of Florida, was cabled back to the laboratory for processing by Dual. T is the log [time-average quadratic of a processor output]. The model assumes the T process is Gauss-Markov and the parameter estimation and filtering equations were used in the computations.

The T process observations exhibit cyclic behavior, so that a second-order model is appropriate. The filtered output $\hat{x}_{k|k}$ of the Kalman filter of this model is a smoothed version of the observation

vector y_k, as was to be expected. This estimate $\hat{x}_{k|k}$ is the conditional mean of the process x_k, given all observations y_k in the past, and is optimum for any convex criteria. Furthermore, the Kalman prediction filter produces $\hat{x}_{k+j|k}$, which is a forecast of the expected value of the process x_{k+j} given all observations y_k in the past. This forecasting ability of the model is a natural result of the state-space formulation. For example, as shown in Section 7.4.4, short-range forecasting over 20 observation intervals is apparently reasonable, and is a technique with potential applications.

The estimates of the model parameters, the state transition matrix Φ and the noise covariance matrices Q and R, are apparently adequate. No absolute criterion of judgement is available, because the true values are unknown to the experimenter. However, Monte-Carlo tests indicated acceptability of $\hat{\Phi}_k$ and $\hat{Q}_k$ although $\hat{R}_k$ was troublesome and needs more work.

Chapter 8 applied the scalar model to the N process, which is the noise power process produced by the Dual processing. The process contains shipping and biological noise "pulses" far greater than the ambient ocean noise. This fact and insufficient data for the process really prevent sweeping conclusions. The estimators and Kalman filter for the scalar case were operable and produced sensible results.

9.1.3 Achieved Objectives. The objectives stated in Chapter 1

have been achieved, albeit for only a special class of processes, the Gauss-Markov. The density function is a complete statistical description of a random variable or process. Since the conditional probability densities of x_k are Gaussian, knowledge of the conditional mean and conditional covariances completely solves the filtering and prediction problems. The conditional means are denoted by $\hat{x}_{k|k}$ and $\hat{x}_{k+j|k}$ and the conditional covariances can be determined from the prediction error covariance P_k. The solutions have the intuitive appeal of being cast as filters and predictors.

An objective in this study of models for nonstationary stochastic processes is to provide information on the behavior of the model parameters. These results have been limited to sequences of the parameters $\hat{\Phi}_k$, $\hat{Q}_k$, $\hat{R}_k$ as estimated from the data. No further studies of this data have been done.

9.1.4 Contributions. Though much of the work reported here is that of others, as documented, some of the small contributions to the practical aspects of stochastic modeling are briefly described.

1. In Chapter 3, a set of steady-state initial conditions were found for the Kalman-Bucy equations. This was found useful in working with real data.

2. In Chapter 4, estimators were given which are called tracking estimators, in the hope that a proper choice of coefficients by the engineer will permit time variations of the parameters to be

followed. Also, a reasonable set of initial conditions were found for the statistics used by the estimators.

3. Several forms of digital processors were discussed. Chapter 5 is a new and unique treatment of the class of processors which have been labelled MPR processors. The phrase "multi-positional recirculators" is descriptive of these processors. The nonlinear processor to eliminate shipping noise and the curve fit to extract the vector observation are examples of ad hoc designs which are very important.

4. To the author's knowledge, these applications of stochastic models are the first made in the field of ocean acoustics. Indeed, it should be pointed out that the Dual Processing data represents the first large quantity of data taken uniformly in time and made available in digital form for post-processing. The data reported on the T process and N process is but an initial step and part of more such data in a continuing study.

9.2 Future Work

One of the interesting characteristics of research is that it usually points to the need for a wider theory, the possibility of another application or opens up new avenues to existing problems. Exceptional and the exception is the report which completely documents a field of research and leaves nothing to be done.

Future work on this and related research problems is abundant.

Some of these areas have been mentioned in the body of the text and some are offered for the reader's consideration. If some of these suggestions seem vague, I caution that they are not yet well-defined and that is often part of the problem.

1. Concerning the oscillations in the T process in particular, a propagation model is suggested which may account for the fluctuations. Consider distant noise sources and propagation effects which cause constructive and destructive interference at the hydrophone. It is not clear that such a propagation model, incorporated in the stochastic model, could account for such changes.

2. More sophistication in observation techniques is certain to have good results for some processes. For example, if it were possible to observe the y_k vector rather than just a scalar sequence, better forecasts can be expected. Also, wind speed and wave height have been shown to correlate with noise level (Ref. 22); by increasing the observation vector dimension and the model order, better estimators of the noise levels should result.

3. This obviously points to the need to study other processes (such as wind speed data), both by themselves and jointly. The theory exists but the application is lacking.

4. Significant improvements are expected in modeling by finding better estimators. The shortcomings of the estimator $\hat{R}_k$ have been discussed. Perhaps a higher order "filter" in the statistics

is required.

5. Asking for more dimensions and better estimators increases the computational needs. For a study of computational aspects, in particular for large dimensional models, the researcher should consult the thesis by Pentecost (Ref. 14).

6. For sensitivity studies, concerning the effect of estimates on the Kalman filter, see the work reported by Bucy and Joseph (Ref. 4).

7. Having obtained better models and sufficient data on the process, extensive studies on long-range forecasting and its reliability seem necessary before this concept can be correctly applied.

As for other areas of application, there come to mind weather system studies and forecasting, ionospheric propagation studies and the field of economic systems. Considerable data is currently being gathered on these topics and some modeling has been done. For other applications, see the text by Bucy and Joseph for examples.

APPENDIX A

A TECHNICAL GLOSSARY

This appendix collects the definitions of all the technical terms which may be unfamiliar or new to the reader. All underlined terms in the main body of the text are defined here. The starting point for these definitions is a knowledge of terms in probability theory. The probability space is denoted by $(\Omega, \mathcal{F}, P)$, where Ω is an abstract set, called the sample space, $\mathcal{F}$ is a σ-field of subsets of Ω and P is a probability measure on $\mathcal{F}$. Let $\mathcal{D}$ be a subset of $\mathcal{F}$ and let $\mathcal{B}$ be the σ-field on the real line and $\mathcal{B}(s)$ is the field generated by the space s.

Almost surely

A condition is said to hold almost surely if the set on which it does not hold has probability zero.

Chi-squared distribution

The chi-squared (chi-square) distribution is a particular form of a gamma distribution with $\alpha = \frac{n}{2} - 1$ and $\beta = 2$,

$$f(x;n) = \frac{1}{\left[\frac{n}{2} - 1\right]!} \frac{1}{2^{n/2}} x^{\frac{n}{2} - 1} e^{-x/2}$$

where n is called the number of degrees of freedom.

Conditional expectation

The conditional expectation $E\{x|\mathcal{D}\}$ is any random variable measurable $(\Omega, \mathcal{D})$ such that

$$\int_D E\{x|\mathcal{D}\}\, dP = \int_D x dP, \qquad \text{all } D \in \mathcal{D}$$

Any two versions differ on a set of probability zero. If $\mathcal{D} = \mathcal{F}(y_1, y_2, \ldots)$, denote $E\{x|\mathcal{D}\} = E\{x|y_1, y_2, \ldots\}$.

Let y be a random variable on $(R,\mathcal{B})$, then for $\lambda \in R$, the conditional expectation

$$\phi(\lambda) = E\{x|y=\lambda\}$$

is any random variable on $(R,\mathcal{B})$, where $\hat{P}(B) = P(y \in B)$ satisfying

$$\int_B E\{x|y=\lambda\}\, d\hat{P} = \int_{y \in B} x\, dP, \qquad \text{all } B \in \mathcal{B}.$$

Conditional probability

Let $\mathcal{D}$ be a sub-σ-field of $\mathcal{F}$. The conditional probability of $A \in \mathcal{F}$ given $\mathcal{D}$ is a random variable $P(A|\mathcal{D})$ on $(\Omega,\mathcal{D})$ satisfying

$$\int_D P(A|\mathcal{D})\, dP = P(A \cap D), \qquad \text{all } D \in \mathcal{D}.$$

Consistent estimator

An estimator is said to be consistent if the sequence of estimates converges in probability. $e_n \to e$ in probability.

Convergence in mean-square

The sequence x_n tends to x in the mean square sense if

$$E\{|x_n - x|^2\} \to 0 \qquad \text{for} \quad n \to \infty .$$

This is also called $\ell.i.m.$ convergence and written

$$\ell.i.m.\ |x_n - x|^2 \to 0 .$$

Convergence in probability

The sequence x_n tends to x in probability if

$$\text{for all } \lambda, \quad P\{|x_n - x| > \lambda\} \to 0 \quad \text{for} \quad n \to \infty .$$

This is also called stochastic convergence or convergence in measure.

Convergence with probability 1

The sequence x_n is said to converge to x with probability 1 if

$$P\{x_n \to x\} = 1 \text{ for } n \to \infty .$$

This is also called convergence almost everywhere (a.e.).

Convolution

The convolution of functions $f_1(x)$ and $f_2(x)$ is denoted by

$$f(x) = f_1(x) * f_2(x)$$

and is defined by

$$f(x) = \int_{-\infty}^{\infty} f_1(y)\, f_2(x-y)\, dy = \int_{-\infty}^{\infty} f_1(x-y)\, f_2(y)\, dy \quad .$$

Delay line

A delay line is a physical device which accepts input (voltage waveforms or numbers) to deliver as output, at a fixed time later, the value of the input.

Ergodicity

Let T be measure-preserving on $(\Omega, \mathcal{F}, P)$. T is called ergodic if for every $A \in \mathcal{I}$, $P(A) = 0$ or 1.

Gamma distribution

A random variable x is distributed as the gamma distribution if its density is

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\alpha!\ \beta^{\alpha+1}}\ x^{\alpha}\ e^{-x/\beta} & 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

where $\beta > 0$ and $\alpha > -1$.

Gaussian process

A process x_k (with random variables $x_0, x_1, x_2, \ldots$) is called Gaussian if every finite subset of variables has a joint Gaussian distribution.

Generalized inverse

A matrix $A^{\#}$ is called the generalized inverse of a rectangular matrix A if it satisfies

i) $AA^{\#}A = A$

ii) $A^{\#}AA^{\#} = A^{\#}$

iii) $(AA^{\#})' = AA^{\#}$

iv) $(A^{\#}A)' = A^{\#}A$

Input-output stable

A system is said to be input-output stable if bounded inputs cause only bounded outputs.

Joint Gaussian distribution

The array $Y = (y_1, \ldots, y_n)$, $E\{y_i\} = 0$ has a joint Gaussian (or joint normal) distribution with zero means if there are m independent random variables $X = (x_1, \ldots, x_m)$, each with the Gaussian distribution with zero mean and unit variance (called $N(0, 1)$), and if an $m \times n$ matrix A exists such that

$$Y = XA \ .$$

Limiter

A limiter is a transformation of a process such that the (output) range is bounded above and below.

$$y = \begin{cases} x_a & \text{for } x \geq x_a \\ x & x_b \leq x \leq x_a \\ x_b & x \leq x_b \end{cases}$$

It is often illustrated by the diagram shown and is sometimes called a clipper.

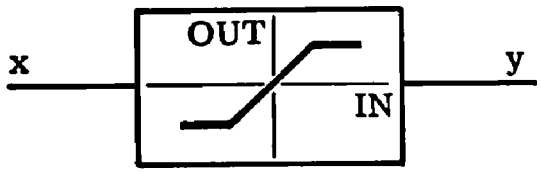


Fig. A.1. A limiter

Linear function

A function $f(\cdot)$ is linear if

$$f(ax + by) = a\,f(x) + b\,f(y)$$

for all a, b, x and y in its domain.

Markov process

A process $x(\omega, t)$, $\omega \in \Omega$, $t \geq 0$ is called Markov with state space $S \in \mathcal{B}$, if $x(t) \in S$, $t \geq 0$, and for any $B \in \mathcal{B}_1(S)$, t and

$\tau \geq 0$,

$$P(x(t+\tau) \in B \mid x(s),\ s \leq t) = P(x(t+\tau) \in B \mid x(t)) \qquad \text{a.s.}$$

Matrix inverse

A square matrix Λ has an inverse, denoted by Λ^{-1}, if its left and right inverses exist, so that

$$\Lambda^{-1} \Lambda = I = \Lambda \Lambda^{-1} .$$

Matrix Riccati equation

The matrix Riccati equation for continuous time is the matrix differential equation

$$\frac{d P(t)}{dt} = F(t) P(t) + P(t) F'(t) + Q(t)$$

$$P(t_0) = \Lambda_0 .$$

Measure-preserving

A measurable transformation T on $\Omega \to \Omega$ is called measure-preserving if $P(T^{-1} A) = P(A)$, all $A \in \mathcal{F}$. A set $A \in \mathcal{F}$ is invariant if $T^{-1} A = A$ for such T; the class of invariant sets is a σ-field $\mathcal{I}$.

Norm of a matrix

Let R^p be the space of $p \times p$ matrices with real entires. The Euclidian norm of $G \in R^p$ is denoted and defined by

$$\| G \| = [\mathrm{tr}(GG')]^{\frac{1}{2}}$$

Orthogonal random variable

The random variables $x_1, x_2, \ldots x_n$ are orthogonal if

$$E\{x_i x_j\} = 0 \qquad \forall i \neq j .$$

Periodic analog of sinc x

The function with parameter M

$$\mathrm{sinc}(x ; M) = \frac{\sin \pi x}{M \sin\left(\frac{\pi x}{M}\right)}$$

is called the periodic analog of sinc x.

Positive, nonnegative definite

A square matrix Λ is called positive definite if $x' \Lambda x > 0$ for all column vectors x. It is called nonnegative definite if $x' \Lambda x \geq 0$ (where the distinction is the allowed 0 value).

Pseudo-inverse

A matrix $\underline{A}^+$ is called a pseudo-inverse of a rectangular (not necessarily square) matrix $\underline{A}$ if it satisfies the relation

$$AA^+A = A .$$

Random sample

Let the variables $x_1, x_2, \ldots, x_n$ be independent identically distributed with density $f(x)$. Then $x_1, \ldots, x_n$ is a random sample of size n from the population with density $f(x)$.

Random variable

A random variable $r(\omega)$ is a measurable function defined on the space $(\Omega, \mathcal{F})$.

Sinc x

The sinc function is defined by

$$\text{sinc } x = \frac{\sin \pi x}{\pi x}$$

Spectral radius of A

Given the square real matrix A, the spectral radius of A is denoted by $\rho(A)$ and is defined as

$$\rho(A) = \lim_{n \to \infty} \|A^n\|^{1/n}$$

State equation

The state equation is the vector-matrix first-order differential equation

$$\frac{d\,x(t)}{dt} = F(t)\,x(t) + w(t)$$

where x is called the state vector and F is the coefficient matrix.

State of a system

The state of a system is the minimal amount of information at t_0 about the effects of past inputs necessary to describe completely the output for $t \geq t_0$. This does not include the description of the input, but when the input is included, the result is an extended state vector.

State transition matrix

The state transition matrix is usually denoted by $\Phi(t, t_0)$ and is formally defined as the solution to the matrix-differential equation

$$\frac{d\,\Phi(t, t_0)}{dt} = F(t)\,\Phi(t, t_0)$$

$$\Phi(t_0, t_0) = I \ .$$

Strict-sense stationary process

A strict-sense stationary process has all its statistics not affected by a shift in the time origin.

Stochastic process

A stochastic process is a family of random variables $r(\omega, t)$ indexed by a set $\mathcal{A}$, $t \in \mathcal{A}$. (t is often time.)

Strongly consistent estimator

An estimator is said to be strongly consistent if the sequence

of estimates converges with probability 1 .

$$e_n \rightarrow e \qquad \text{almost everywhere (a.e.)}$$

Symmetric matrix

A square matrix Λ is called symmetric if $\lambda_{ij} = \lambda_{ji}$.

Tapped delay line

A sequence of delay lines connected in series is a tapped delay line.

Tracking estimator

The name tracking estimator has been given in this study to the statistic

$$\hat{\Phi}_k = \left(\sum_{j=3}^{k} c_{k-j}\, y_j\, y'_{j-2} \right) \left(\sum_{j=3}^{k} c_{k-j}\, y_{j-1}\, y'_{j-2} \right)^{+}$$

because the coefficients c_{k-j} are chosen such that slow variations in the parameter Φ_k can be followed.

Uncorrelated random variable

The random variables $x_1, x_2, \ldots x_n$ are uncorrelated if

$$E\{x_i x_j\} = E\{x_i\}\, E\{x_j\} \qquad \text{for all} \quad i \neq j \ .$$

Wide-sense stationary process

A process with constant expected value and an autocorrelation which depends only on the time difference is called wide-sense stationary.

APPENDIX B

EXPECTATIONS OF PRODUCTS WITH Φ_k, Φ_{k-1} AS FACTORS

B.1 Required Expectations

In Chapter 4, the formulation of the estimators for Φ, R and Q requires the expectations of terms involving Φ_k and Φ_{k-1}, as shown in Table B.1.

$E\{\Phi_k v_{k-1} v_{k-1}'\}$	$\Phi_k R_{k-1}$	see Eq. B.20	(B.1)
$E\{\Phi_k v_{k-1} v_{k-2}'\}$	0	$-\Phi_{k-1} R_{k-1} R_{k-2} S^{+}_{12,k-1}$	(B.2)
$E\{\Phi_k v_{k-1} u_{k-1}'\}$	0		(B.3)
$E\{\Phi_k v_{k-1} u_{k-2}'\}$	0		(B.4)
$E\{\Phi_k v_{k-1} x_{k-2}'\}$	0		(B.5)
$E\{\Phi_{k-1} v_{k-2} u_k'\}$	0		(B.6)
$E\{\Phi_{k-1} v_{k-2} v_k'\}$	0		(B.7)
$E\{\Phi_k v_{k-1} v_{k-2}' \Phi_k'\}$	0		(B.8)

Table B.1. Required expectations

A first approximation is given in the second column, which is based on the assumption that Φ_k is independent of the indicated

products. This should be a good approximation because $\hat{\Phi}_k \approx \Phi$ and Φ is the system transition matrix, which is assumed independent of x, u and v in the linear model. Also, $E\{\hat{\Phi}_k\} = \Phi_k$ has been used. These are the conditions used in formulating Model B.

However, the estimate $\hat{\Phi}_k$ is certainly dependent on v_j and u_j for $j \leq k$. To display this slight dependence, the following recursion relations for $\hat{\Phi}_k$ are required and then the above expectations are found and recorded in column three.

B.2 Analysis of the Estimator for Φ

Recall that the estimator for Φ_k is

$$\hat{\Phi}_k = \left(\sum_{j=3}^{k} y_j y_{j-2}' \right) \left(\sum_{j=3}^{k} y_{j-1} y_{j-2}' \right)^{+} \qquad k \geq 3 \qquad (4.12)$$

For notational simplicity, let

$$\Delta_j \triangleq y_j y_{j-2}' \quad \text{and} \quad S_{02,k} \triangleq \sum_{j=3}^{k} \Delta_j \qquad (B.9)$$

$$\delta_j \triangleq y_{j-1} y_{j-2}' \quad \text{and} \quad S_{12,k} \triangleq \sum_{j=3}^{k} \delta_j \qquad (P.10)$$

These are $n \times n$ matrices and A^+ is the pseudo-inverse of A. Then

$$\hat{\Phi}_k = S_{02,k} S_{12,k}^{+}$$

$$\Phi_k S_{12,k} = S_{02,k}$$

$$\Phi_k(S_{12,k-1} + \delta_k) = S_{02,k-1} + \Delta_k \tag{B.11}$$

$$\Phi_k(I + \delta_k S_{12,k-1}^{+}) = \Phi_{k-1} + \Delta_k S_{12,k-1}^{+} \tag{B.12}$$

$$\Phi_k = (\Phi_{k-1} + \Delta_k S_{12,k-1}^{+})(I + \delta_k S_{12,k-1}^{+})^{+} \tag{B.13}$$

$$= (\Phi_{k-1} + \Delta_k S_{12,k-1}^{+})(I - \delta_k S_{12,k-1}^{+}) \tag{B.14}$$

$$= \Phi_{k-1} + (\Delta_k - \Phi_{k-1}\delta_k) S_{12,k-1}^{+} \tag{B.15}$$

In Eq. B.14 the approximation $(I+b)^{+} \doteq I - b$ is valid for "small" b. This is the matrix analog of $\frac{1}{1+b} \doteq 1 - b$. Equation B.15 results by retaining only first-order terms in Δ and δ. By letting $k \rightarrow k-1$ in Eq. B.15

$$\Phi_{k-1} = \Phi_{k-2} + (\Delta_{k-1} - \Phi_{k-2}\delta_{k-1}) S_{12,k-2}^{+} \tag{B.16}$$

So
$$\Phi_k = \Phi_{k-2} + (\Delta_{k-1} - \Phi_{k-2}\delta_{k-1}) S_{12,k-2}^{+} \tag{B.17}$$

$$+ (\Delta_k - \Phi_{k-1}\delta_k) S_{12,k-1}^{+}$$

B.3 Computations

A sample computation of the result in (B.1) will be shown. Φ_{k-2}, Δ_k, δ_{k-1} and $S_{12,k-1}^{+}$, $S_{12,k-2}^{+}$ are all independent of v_{k-1}. Hence,

$$E\{\Phi_k v_{k-1} v_{k-1}'\} = \Phi_{k-2} R_{k-1}$$

$$+ (\Phi_{k-1} \Phi_{k-2} \Sigma_{x,k-3} - \Phi_{k-2} \Phi_{k-2} \Sigma_{x,k-3}) S_{12,k-2}^{+} R_{k-1}$$

$$+ (\Phi_k \Phi_{k-1} \Sigma_{x,k-2} - \Phi_{k-1} \Phi_{k-1} \Sigma_{x,k-2}) S_{12,k-1}^{+} R_{k-1} \qquad \text{(B.18)}$$

These results are obtained by assuming the differences in Eq. B.17 are independent of v_{k-1}. Any error is reduced by the matrix S_{12}^{+}, which is small. Also,

$$E\{\hat{\Phi}_k\} = \Phi_k \quad \text{and} \quad E\{S_{12,k-2}^{+}\} = S_{12,k-2}^{+}$$

have been used. Now note that

$$S_{12,k} = \sum_{j=3}^{k} \delta_j \doteq E(y_{k-1} y_{k-2}') = \Phi_{k-1} \Sigma_{x,k-2}$$

Therefore,

$$S_{12,k}^{+} = (\Phi_{k-1} \Sigma_{x,k-2})^{-1} = \Sigma_{x,k-2}^{-1} \Phi_{k-1}^{-1} \qquad \text{(B.19)}$$

Finally,

$$E\{\Phi_k v_{k-1} v_{k-1}'\} = [\Phi_{k-2} + (\Phi_{k-1} - \Phi_{k-2}) \Phi_{k-2} \Sigma_{x,k-3} \Sigma_{x,k-4}^{-1} \Phi_{k-3}^{-1}$$

$$+ (\Phi_k - \Phi_{k-1}) \Phi_{k-1} \Sigma_{x,k-2} \Sigma_{x,k-3}^{-1} \Phi_{k-2}^{-1}] R_{k-1} \qquad \text{(B.20)}$$

A brief discussion of this result seems in order. The value in column one (B.1) is $\Phi_k R_{k-1}$. Φ_k is replaced by the square

bracket in (B.20). If Φ is stationary, the difference terms are 0 and $\Phi_k = \Phi_{k-2}$. If Φ is increasing ($\Phi_{k-2} \leq \Phi_{k-1} \leq \Phi_k$) or decreasing ($\Phi_{k-2} \geq \Phi_{k-1} \geq \Phi_k$), the new result is between Φ_{k-2} and Φ_k.

The result in (B.20) depends on Σ_x, which is undesirable from the modeling point of view. For Σ_x time-invariant,

$$\Sigma_{x,j} = \Sigma_{x,j-1} \qquad \text{for all } j \tag{B.21}$$

so that

$$E\{\Phi_k v_{k-1} v_{k-1}'\} = [\Phi_{k-2} + (\Phi_{k-1} - \Phi_{k-2}) \Phi_{k-2} \Phi_{k-3}^{-1} + (\Phi_k - \Phi_{k-1}) \Phi_{k-1} \Phi_{k-2}^{-1}] R_{k-1}$$

and Σ_x is not involved. When Σ_x is time-varying, this dependence on Σ_x remains, but only past values are required. Furthermore, inspection of (B.20) shows that large changes in Σ_x are needed before the effect begins to show.

Other values in the table (B.1 to B.8) are obtained in the same manner.

APPENDIX C

EXPECTATIONS OF QUARTIC TERMS FOR THE VECTOR CASE

The need arises for expectations of quartic terms. For the scalar case, $n = 1$,

$$\begin{aligned} E\{x_1 x_2 x_3 x_4\} &= E\{x_1 x_2\} E\{x_3 x_4\} \\ &+ E\{x_1 x_3\} E\{x_2 x_4\} \\ &+ E\{x_1 x_4\} E\{x_2 x_3\} \end{aligned} \qquad (C.1)$$

if x_i are jointly Gaussian and have mean 0. The extension to the vector case is given here.

$$\begin{aligned} E\{x_1 x_2' x_3 x_4\} &= E\{x_1 x_2'\} E\{x_3 x_4'\} \\ &+ E\{x_1 x_3'\} E\{x_2 x_4'\} \\ &+ E\{x_1 x_4'\} E\{x_2 x_3'\} \end{aligned}$$

if the x_i are n-vectors (column), have mean $\underline{0}$ and the x_i are jointly Gaussian.

Proof:

The proof will use the following general result.

$$\frac{\partial^r \mathcal{C}(0, 0, \ldots 0)}{\partial w_1^{k_1} \partial w_2^{k_2} \ldots \partial w_m^{k_m}} = j^r E\left\{x_1^{k_1} x_2^{k_2} \ldots x_m^{k_m}\right\}^*$$

where

$$r = k_1 + k_2 + \ldots k_m$$

$\mathcal{C}(w_1, w_2, \ldots w_m)$ = characteristic function.

For the jointly Gaussian x_1, x_2, x_3, x_4 vectors with mean 0, the characteristic function is

$$\mathcal{C}(w_1, w_2, w_3, w_4) = \exp\left[-\frac{1}{2}(w_1', w_2', w_3', w_4') R \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}\right]$$

and $R = [R_{ij}]$, $R_{ij} = E\{x_i x_j'\}$.

R_{ij} is $n \times n$, R is $4n \times 4n$ in dimension.

$$\left.\frac{\partial \mathcal{C}(w_1, w_2, w_3, w_4)}{\partial w_4'}\right|_{w_4=0} = -(R_{41}w_1 + R_{42}w_2 + R_{43}w_3)\,\mathcal{C}(w_1, w_2, w_3, 0)$$

$$\left.\frac{\partial^2 \mathcal{C}(w_1, w_2, w_3, 0)}{\partial w_3\, \partial w_4'}\right|_{w_3=0} = +\{(R_{41}w_1 + R_{42}w_2)(R_{31}w_1 + R_{32}w_2)' - R_{43}\}\,\mathcal{C}(w_1, w_2, 0, 0)$$

---

*Papoulis, page 244.

$$\left.\frac{\partial^3 \mathcal{C}(w_1, w_2, 0, 0)}{\partial w_2' \, \hat{c} w_3 \, \hat{c} w_4'}\right|_{w_2=0} = [-\{R_{41} w_1 w_1' R_{13} - R_{43}\} (R_{21} w_1) + R_{42} R_{31} w_1 + R_{32} R_{41} w_1] \, \mathcal{C}(w_1, 0, 0, 0)$$

$$\left.\frac{\partial^4 \mathcal{C}(w_1, 0, 0, 0)}{\partial w_1 \hat{c} w_2' \hat{c} w_3 \hat{c} w_4'}\right|_{w_1=0} = (+R_{43} R_{21} + R_{42} R_{31} + R_{32} R_{41})' = R_{12} R_{34} + R_{13} R_{24} + R_{14} R_{23}$$

which is the result claimed.

APPENDIX D

EXPRESSIONS FOR THE OBSERVATION VECTOR

In the development of the model equations and estimators for the model parameters Φ, Q and R, expressions for y_k in terms of previous observations are needed. For ready reference, these results are listed below. They are merely repeated applications of the model equations.

$$x_{k+1} = \Phi_{k+1} x_k + u_{k+1} \tag{4.1}$$

$$y_k = x_k + v_k \tag{4.2}$$

$$y_{k-4} = x_{k-4} + v_{k-4}$$

$$y_{k-3} = x_{k-3} + v_{k-3}$$

$$= \Phi_{k-3} x_{k-4} + u_{k-3} + v_{k-3}$$

$$y_{k-2} = x_{k-2} + v_{k-2}$$

$$= \Phi_{k-2} x_{k-3} + u_{k-2} + v_{k-2}$$

$$= \Phi_{k-2} \Phi_{k-3} x_{k-4} + \Phi_{k-2} u_{k-3} + u_{k-2} + v_{k-2}$$

$$y_{k-1} = x_{k-1} + v_{k-1}$$

$$= \Phi_{k-1} x_{k-2} + u_{k-1} + v_{k-1}$$

$$= \Phi_{k-1} \Phi_{k-2} x_{k-3} + \Phi_{k-1} u_{k-2} + u_{k-1} + v_{k-1}$$

$$= \Phi_{k-1} \Phi_{k-2} \Phi_{k-3} x_{k-4} + \Phi_{k-1} \Phi_{k-2} u_{k-3} + \Phi_{k-1} u_{k-2} + u_{k-1} + v_{k-1}$$

$$y_k = x_k + v_k$$

$$= \Phi_k x_{k-1} + u_k + v_k$$

$$= \Phi_k \Phi_{k-1} x_{k-2} + \Phi_k u_{k-1} + u_k + v_k$$

$$= \Phi_k \Phi_{k-1} \Phi_{k-2} x_{k-3} + \Phi_k \Phi_{k-1} u_{k-2} + \Phi_k u_{k-1} + u_k + v_k$$

$$= \Phi_k \Phi_{k-1} \Phi_{k-2} \Phi_{k-3} x_{k-4} + \Phi_k \Phi_{k-1} \Phi_{k-2} u_{k-3} + \Phi_k \Phi_{k-1} u_{k-2}$$

$$+ \Phi_k u_{k-1} + u_k + v_k$$

APPENDIX E

THE STEADY-STATE COVARIANCE Λ

The need exists for a symmetric positive definite solution Λ of the matrix equation

$$\Lambda = \Phi \Lambda \Phi' + Q \ .$$

Φ is a state-transition matrix and Q is a covariance matrix. For the scalar case, the solution is evidently

$$\Lambda = \frac{Q}{1 - \Phi^2}$$

which is positive if $\Phi^2 < 1$. For the matrix case, a similar condition is required but the solution is not so readily expressed.

Proposition: Let Φ have spectral radius[1] $\rho(\Phi) < 1$ and let Q be symmetric positive definite. Then the equation

$$\Lambda = \Phi \Lambda \Phi' + Q$$

has the solution

$$\Lambda = \sum_{i=0}^{\infty} \Phi^i Q(\Phi')^i$$

---

[1] See Appendix A for the definition.

where $\Phi^0 = I$; $\Phi^i = \Phi \text{ --- } \Phi$ is the i-fold product. The proof is by substitution of finite sums

$$\Lambda^{(k)} = \sum_{i=0}^{k} \Phi^i \, Q(\Phi')^i$$

and then letting $k \to \infty$. Existence and uniqueness have been shown.

Often, the actual computation of Λ can be accomplished with a small number of iterations k, because of the condition on Φ. Furthermore, convergence of the solution is accelerated by actually "solving" the equation recursively:

$$\Lambda^{(k)} = \Phi \, \Lambda^{(k-1)} \, \Phi' + Q \qquad k > 1$$

$$\Lambda^{(0)} = Q \; .$$

A good initial approximation greatly reduces the required iterations.

REFERENCES

1. W. N. Anderson, Jr., G. B. Kleindorfer, P. R. Kleindorfer and M. B. Woodroofe, Consistent Estimates of the Parameters of a Linear System, Department of Statistics, Carnegie-Mellon University, Pittsburgh, Pennsylvania, August 1968.

2. T. G. Birdsall, Adaptive Detection Receivers and Reproducing Densities, Cooley Electronics Laboratory Technical Report No. 194, The University of Michigan, Ann Arbor, Michigan, July 1968.

3. L. Breiman, Probability, Addison-Wesley Publishing Company, Menlo Park, California, January, 1968.

4. R. S. Bucy and P. D. Joseph, Filtering For Stochastic Processes With Applications to Guidance, Interscience Publishers, New York, May 1968.

5. J. L. Doob, Stochastic Processes, John Wiley and Sons, New York, 1963.

6. M. Abramowitz and L. A. Stegun (Eds.), Handbook of Mathematical Functions, National Bureau of Standards, AMS 55, Washington, D. C., 1964.

7. N. Hatter, Triband: The Three-Band Filter for the Continuing MIMI Experiment, Cooley Electronics Laboratory Technical Report No. 201, The University of Michigan, Ann Arbor, Michigan, February 1970.

8. D. Jaarsma, The Theory of Signal Detectability: Bayesian Philosophy, Classical Statistics, and the Composite Hypothesis, Cooley Electronics Laboratory Technical Report No. 200, The University of Michigan, Ann Arbor, Michigan, February 1970.

9. R. E. Kalman, "New Methods in Wiener Filtering Theory," Proc. 1st Symp. on Engrg. Applications of Random Function Theory and Probability, J. L. Bogdanoff and F. Kozin, Eds., New York: Wiley, 1963.

10. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Eng., A.S.M.E., 82 (1960), pp. 35-45.

REFERENCES (Cont.)

11. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," J. Basic Eng., A.S.M.E., 83 (1961), pp. 95-108.

12. P. P. Nuspl, "Probability Density Function of Time-Average Power and Log (Power)," to appear as a Letter-to-the-Editor, J. Acoustical Society of America, July 1970.

13. A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill, New York, 1965.

14. E. E. Pentecost, "Synthesis of Computationally Efficient Sequential Linear Estimators," Thesis, University of California at Los Angeles, 1965.

15. M. I. Schwartz, "Distribution of the Time-Average Power of a Gaussian Process," IEEE Trans. on Information Theory, IT-16, No. 1, (1970), pp. 17-26.

16. R. L. Spooner, Theory of Signal Detectability: Extension to the Double Composite Hypothesis Situation, Cooley Electronics Laboratory Technical Report No. 192, The University of Michigan, Ann Arbor, Michigan, April 1968.

17. J. C. Steinberg and T. G. Birdsall, "Underwater Sound Propagation in the Straits of Florida," J. Acoustical Society of America, Vol. 39, No. 2, February 1966, pp. 301-315.

18. I. Tolstoy and C. S. Clay, Ocean Acoustics: Theory and Experiment in Underwater Sound, McGraw-Hill, New York, 1966..

19. R. Unger and R. Veenkant, Underwater Sound Propagation in the Straits of Florida. The MIMI Experiment of 3 and 4 February 1965, Cooley Electronics Laboratory Technical Report No. 183, The University of Michigan, Ann Arbor, Michigan, May 1967.

20. R. Unger and R. Veenkant, Underwater Sound Propagation in the Straits of Florida. The MIMI Continuous and Sampled Receptions of 11, 12, and 13 August 1966, Cooley Electronics Laboratory Technical Report No. 186, The University of Michigan, Ann Arbor, Michigan, June 1967.

REFERENCES (Cont.)

21. R. J. Urick, Principles of Underwater Sound for Engineers, McGraw-Hill, New York, 1967.

22. H. L. Van Trees, Detection, Estimation and Modulation Theory, Part I, John Wiley and Sons, Inc., New York, 1968.

23. L. A. Zadeh and C. A. Desoer, Linear System Theory: The State Space Approach, McGraw-Hill, New York, 1963.