Reports of the Department of Geodetic Science

Report No. 244

PROBABILISTIC AND DETERMINISTIC ASPECTS OF LINEAR ESTIMATION IN GEODESY

by

Athanasios Dermanis

Prepared for the

National Aeronautics and Space Administration
Washington, D.C.

Grant No. NGR 36-008-204
OSURF Project No. 3820-A1

The Ohio State University
Research Foundation
Columbus, Ohio 43212

October, 1976

Reports of the Department of Geodetic Science

Report No. 244

PROBABLISTIC AND DETERMINISTIC ASPECTS OF LINEAR ESTIMATION IN GEODESY

by

Athanasios Dermanis

Prepared for the

National Aeronautics and Space Administration
Washington, D.C.

Grant No. NGR 36-008-204
OSURF Project No. 3820-A1

The Ohio State University
Research Foundation
Columbus, Ohio 43212

October, 1976

PREFACE AND ACKNOWLEDGEMENTS

This project is under the supervision of Professor Ivan I. Mueller, Department of Geodetic Science, The Ohio State University, and under the technical direction of Mr. James P. Murphy, Special Programs, Office of Applications, Code ES, NASA Headquarters, Washington, D.C. The contract is administered by the Office of University Affairs, NASA, Washington, D.C. 20546.

I wish to thank Dr. Ivan I. Mueller without whose continuous support and encouragement during the past four years this work would not have been possible.

I would like to thank my colleague and friend C. C. Tscherning of the Danish Geodetic Institute who initiated me into Hilbert space techniques during his stay at OSU.

I would also like to thank Prof. U.A. Uotila, of this Department, who, through his teaching and during many hours of enjoyable discussions, introduced me to today's problems and techniques of estimation in geodesy.

Special thanks must be extended to Prof. J.T. Scheick of the OSU Department of Mathematics who helped me formulate the necessary mathematical background for this study. His thoughtful criticism resulted in revision and clarification of significant parts of this work. Prof. T.P. Schwartzbauer of the same department also helped in the critical evaluation of the manuscript.

I wish also to thank Mrs. Irene Tesfai for her help in preparing this manuscript for publication.

This report has been submitted to the Graduate School of The Ohio State University in partial fulfillment of the requirements for the degree Doctor of Philosophy.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

The object of geodesy is usually defined to be "the determination of the shape and gravity field of the earth." Not being happy with the idea of devoting ourselves to a hopeless task, we take the liberty of revising the definition to read: "the estimation of the shape and gravity field of the earth." Faithful to this definition, this work is about estimation, the process of extracting information about parameters and/or functions from observational data. To extract any information about the physical world from observations, it is first necessary to construct a simplified and idealized image of the real world, i.e., a model.

The complexity of model building is very much dependent on the accuracy of the observational techniques. In the presence of large observational errors, the adoption of a simplified model is warranted since an elaborate model would not compensate for the loss of information associated with observational errors.

Indeed, simple geometric models of axiomatic validity have been of great use in geodetic work, although observables could not be directly identified with their simple counterparts in the model. This problem was bypassed by the appropriate a priori reduction of the observations. Although uncertainties were present in these reductions, their order of magnitude was insignificant compared to the level of observational noise. On the other hand, recent advances in observational techniques related to geodetic work (VLBI, laser ranging) make it imperative that more consideration should be given to modeling problems. Uncertainties in the effect of atmospheric refraction, polar motion and precession-nutation parameters,

etc. cannot be dispensed with in the context of "centimeter level geodesy." Even physical processes that have generally been previously altogether neglected (station motions) must now be taken into consideration.

The problem, in essence, is one of modeling functions of time or space, or at least their values at observation points (epochs). Modeling each function value as an independent unknown parameter may be possible in some cases (geometric methods in satellite geodesy), though it may generally result in overparameterization; and therefore the interdependence of function values must obviously be taken into consideration. When the nature of the function to be modeled is unknown, one may resort to representing the function in terms of a finite number of parameters using polynomials, trigonometric series, step functions, etc (polynomial fits in short arc satellite geodetic techniques, truncated spherical harmonic expansions of the gravity field of the earth). The need to include a limited number of terms and to a priori decide upon a specific form may result in a representation which fails to sufficiently approximate the unknown function.

An alternative approach of increasing application in several scientific disciplines nowadays is the use of stochastic models, i.e., the modeling of unknown functions as stochastic processes. Although the functions under consideration are not known, they are, in general, not completely unknown either. For example, in spite of the fact that the position of the earth-rotation-pole at some current epoch is not exactly known, some places on the earth can be considered much more likely candidates than others. This situation of relative uncertainty suggests the use of probability as a measure of this uncertainty. Thus, stochastic models are means of describing the "likely" behavior of the unknown function.

Relative uncertainty is hardly the usual framework for introducing probability concepts. On the contrary, one is usually made aware of these concepts through the relative frequency definition [Papoulis, 1965, p. 8]. According to this approach, probability theory describes the statistical or average behavior of random entities over an infinite ensemble of possible

realizations. This poses some disturbing questions about the relevance of probabilistic techniques to geodetic problems. Indeed, in geodesy we deal with unique functions rather than ensembles of likely-behaving functions whose average behavior would call for statistical-probabilistic descriptions.

Such difficulties are not particular to geodesy and can be circumvented with the introduction of the concept of ergodicity. A function is assumed to behave likely over different parts of its domain of definition so that ensemble averages can be replaced by "time" averages.

However, the limiting frequency interpretation of probability is not the only one. The meaning of probability has been a matter of much controversy over the last two hundred years [Fine, 1973]. Related arguments have acquired more practical importance as the use of probabilistic ideas has spread into so may scientific activities. We need only mention the distinctly different interpretation given by subjectivists who even come to the explicit aphorism that "probability does not exist" [Finetti, 1974, p. x]. According to this school of thought, the uncertainty that probability is meant to describe is our own "subjective" uncertainty about reality and not inherent to the physical phenomena themselves. This point of view is indeed very comforting to the geodesist eager to use probabilistic techniques; for even though it does not answer the question of relevance of probabilistic concepts to geodetic work, it nevertheless dispenses with the problem altogether. The classification of physical processes into random and deterministic ones becomes irrelevant; and since stochastic models are meant to describe our own indisputably limited knowledge about physical processes, their relevance is obvious.

In essence, the interpretation of probability, although of relevance to geodetic work, is more of a problem in philosophy of science. However, the justification for the use of probabilistic tools in applied science relies on such an interpretation; and one should feel obligated to be at least aware of the questions, if not of giving an answer.

Beyond such philosophical considerations, there are some problems of practical importance associated with estimation in geodesy that have to be addressed.

Within a "second-order theory" framework, existing estimation techniques call for the a priori specification of the first- and second-order statistics (means and covariances) of random parameters and functions involved.

It has been found that the gravity field of the earth cannot be modeled as a random field which is both ergodic and Gaussian [Lauritzen, 1973]. This would disqualify the use of statistical sampling techniques for obtaining an estimate of the relevant covariance function.

Even in the case of ergodicity, where stochastic models might be appropriate, we encounter situations where the functions to be modeled cannot be directly observed, thus prohibiting the use of sampling techniques for covariance estimation.

With respect to the first problem, the probabilistic justification of estimate optimality is dispensed with by exposing the deterministic aspects of the probabilistically derived estimation algorithm, and strictly deterministic criteria for estimate optimality are established.

In regard to the second problem, the possibility of parameterization of the required statistics of stochastic processes and the simultaneous recovery of the relevant parameters along with other unknown parameters is investigated. This leads to "adaptive estimation" schemes where the stochastic model is adapted to the observational evidence.

In order to secure the proper use of existing estimation techniques, it is considered necessary to bring into light the interrelations (similarities and dissimilarities) of such techniques through the use of a unified approach. In the hope of contributing to the understanding of estimation techniques, the separation between the deterministic aspects of estimation algorithms as related to linear best approximation theory and the probabilistic justification of estimate optimality criteria are especially emphasized.

2. FUNDAMENTAL CONCEPTS

2.1 Introductory Remarks

The purpose of this chapter is to present a short exposition of certain fundamental concepts, definitions and results which can be found in a number of textbooks of mathematics and probability theory and which are necessary for reference in the following chapters. Such an exposition will help to avoid a large number of references in the main text, and especially through the introduction of a unified notation will hopefully make the following chapters easier to read.

Without trying to conceal or entertain the definite need for more mathematical rigor, an honest attempt has been made to bring the discussion to a level consistent with the usual mathematical foundation of most geodetic work. Whenever a more rigorous treatment is though to be necessary, proper references to appropriate works are given.

2.2 Random Variables—Concept and Mathematical Model

The concept of a random variable is the cornerstone of this entire discussion. As already explained in the introduction, our concept of a (real) random variable is that of a real valued quantity, where uncertainty exists about its true value. This uncertainty reflects limitations of one's subjective knowledge about the quantity in question and is completely irrelevant to the "true nature" of the quantity. This means that an a priori classification of physical quantities and processes into deterministic and random ones is meaningless. Starting from the fundamental assumption that a "true value" exists, such a value is axiomatically deterministic. On the contrary, our image of reality, based upon our imperfect mental and observational capabilities, is always governed by uncertainty and is therefore random.

These images, or "mathematical models" to use a more familiar term, are always governed by uncertainty, even when this is purposely ignored and a deterministic mathematical model is used.

The mere assertion of the presence of uncertainty falls short from justifying the introduction of a mathematical theory of uncertainty. What is further needed is a measure of uncertainty, i.e., something that helps to distinguish between "much" and "little" uncertainty, including the limiting cases of absolute ignorance (total uncertainty) and complete knowledge (zero uncertainty).

This naturally leads to the use of measure theory. Measure is a function m(A), which maps sets A into nonnegative reals, and furthermore $m(\emptyset) = 0$, where $\emptyset$ is the empty set, i.e., a set with no elements.

In its nonmathematical context, measure is one of the first "mathematical" concepts that man conceived. It is directly related to notions such as long-short, large-small, i.e., to comparison of lengths, volumes, time intervals, etc. Unfortunately, in contemporary mathematical education measure theory is a rather advanced topic and therefore a somewhat lengthy discussion appears to be in order at this point.

Returning to the concept of a random variable, let us think of the uncertain set of values that the random variable may obtain. This will be a (not necessarily proper) subset S of the set of reals R. The concept of measure can now be used to answer questions of the form: "How likely is it that the value of the random variable belongs to a certain subset of S?"

This corresponds to assigning a nonnegative real number to the subset in question, or more generally to defining a measure on a collection of subsets of S. Following a universal convention, the range of this measure will be restricted to the finite interval [0, 1]. This leads to the concept of probability measure. The two critical questions about probability measure that have to be answered next are the following:

(a) Should this probability measure be defined for all possible subsets of

S (the so-called power set $\mathcal{P}(S)$ of S), or for some other collection of subsets?

(b) What are the properties that probability measure should have?

Starting with the second question, some obviously desirable properties are:

$$P_X(\emptyset) = 0$$

$$P_X(S) = 1$$

$$P_X(A \cup B) = P_X(A) + P_X(B), \text{ for } A, B \subset S \text{ and } A \cap B = \emptyset$$

where X denotes the random variable, and P_X the corresponding probability measure on S.

Another not so obvious property, but nevertheless necessary in connection with the concept of limit and convergence of a sequence of random variables, is the following:

$$P_X\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P_X(A_i)$$

where

$$\bigcup_{i=1}^{\infty} A_i \subset S \text{ and } A_i \cap A_j = \emptyset \quad \text{for } i \neq j.$$

We have now arrived at the concept of a countably additive measure. This final property makes the definition of such a measure impossible for $\mathcal{P}(S)$ [Royden, 1968, p. 53]. Another collection of subsets of S must be found as the domain of P_X. Such an appropriate collection is a σ-algebra of subsets of S.

An algebra (or Boolean algebra) of sets is a collection $\mathcal{A}$ of subsets of S, such that

(i) $A \cup B$ is in $\mathcal{A}$ whenever A and B are,

(ii) $A^c = \{t: t \in S, t \notin A\}$ is in $\mathcal{A}$ whenever A is.

An algebra $\mathcal{A}$ of sets is called a σ-algebra or a Borel field if every union of a countable collection of sets in $\mathcal{A}$ is again in $\mathcal{A}$ [Royden, 1968, pp. 16-17].

If for simplicity S is identified with R, we are then primarily concerned with sets of the form $\{x;\ a \leq x \leq b\}$, $\{x;\ a < x < b\}$, $\{x;\ a \leq x\}$, $\{x;\ x < b\}$, etc. Such sets can be constructed from complements and countable intersections and unions of closed intervals of R. The smallest σ-algebra that contains the closed intervals of R is called the Borel sets $\mathcal{B}$ of R. We have now constructed a triplet $\{R, \mathcal{B}, P_X\}$, called a (probability) measure space, and consisting of the set of possible values R that the random variable X may obtain, and a collection $\mathcal{B}$ of subsets of R, on which a probability measure P_X is defined. However, the mathematical context of the random variable itself has not been defined yet. The idea of a set of values R "obtained" by the random variable leads to the identification of R with the range of a function which is consequently identified with the random variable. But a function further needs a domain, and such a domain is constructed through the introduction of an abstract space Ω of elements ω, together with a σ-algebra $\mathcal{A}$ of subsets of Ω and a measure P defined on $\mathcal{A}$, i.e., a probability space $\{\Omega, \mathcal{A}, P\}$.

Now the random variable X can be defined as a function $X: \Omega \to R$, and furthermore a measureable function, i.e., a function satisfying

$$X^{-1}(B) = \{t: X(t) \in B\} \in \mathcal{A} \qquad \text{for every } B \in \mathcal{B}$$

This means that the inverse image under X of every Borel subset of R belongs to the σ-algebra $\mathcal{A}$ of subsets of Ω measureable under P. Up to now we have a mapping

$$X: \{\Omega, \mathcal{A}, P\} \to \{R, \mathcal{B}\}$$

but the measure P induces a measure P_X on $\mathcal{B}$ (the measure we have been originally interested in) through the relation:

$$P_X(B) = P(X^{-1}(B)) \qquad \text{for every } B \in \mathcal{B}$$

The probability measure P_X is called the distribution of the random variable X.

2.3 Integration, Expectation and Moments of a Random Variable

A concept related to measure is that of integration. As a matter of fact, it was questions in integration theory that historically gave rise to the study of measure theory. An exposition of integration theory is beyond the scope of this work, and we shall therefore take for granted knowledge of the definition of integrals of the form

$$\int_A \varphi(\omega)\ dP(\omega)\ , \quad A \in \mathcal{A} \quad \text{and} \quad \int_\Omega \varphi(\omega)\ dP(\omega)$$

where $\varphi(\omega)$ is a measurable function $\varphi: \Omega \to R$.

Definitions of these integrals and related discussions on integration and measure can be found in several texts [Pitt, 1963; Royden, 1968; Kolmogorov and Fomin, 1960; Burril, 1972]. For the random variable $X: \{\Omega, \mathcal{A}, P\} \to \{R, \mathcal{B}, P_X\}$, we have

$$\int_\Omega X(\omega)\ dP(\omega) = \int_R x\ dP_X(x) \tag{2.1}$$

the latter being a Lebesgue integral. To transform this integral into a Riemann-Stieltjes one, we need a description of the measure P_X. Such a description becomes possible through the introduction of the distribution function $F_X(\lambda)$ of the random variable X, defined as

$$F_X(\lambda) = P\{\omega; X(\omega) \leq \lambda\} = P_X(-\infty, \lambda]$$

When $F_X(\lambda)$ is properly defined, then P_X is defined for all sets of the form $(-\infty, \lambda]$, and consequently for every set $B \in \mathcal{B}$. This follows from the fact that B being a Borel set can be written as a union and/or intersection of as many as countable sets or complements of sets of the form $(-\infty, \lambda]$, and from P_X being countably additive.

Now the expectation of a random variable X is defined as

$$E\{X\} = \int_{\Omega} X(\omega)\, dP(\omega) = \int_{R} x\, dP_X(x) = \int_{-\infty}^{+\infty} \lambda\, dF_X(\lambda) \tag{2.2}$$

the latter being a Riemann-Stieltjes integral. The definition can be extended to any function $\varphi(X)$, such that the composite mapping $\varphi \circ X: \Omega \to R$ is a measurable function

$$E\{\varphi(X)\} = \int_{\Omega} \varphi(X(\omega))\, dP(\omega) = \int_{-\infty}^{+\infty} \varphi(\lambda)\, dF_X(\lambda) \tag{2.3}$$

In the particular case that $f_X(\lambda) = \frac{\partial}{\partial\lambda} F_X(\lambda)$ exists and is continuous, expectation can be defined by means of a simple Riemann integral

$$E\{\varphi(X)\} = \int_{-\infty}^{+\infty} \varphi(\lambda)\, f_X(\lambda)\, d\lambda \tag{2.4}$$

$f_X(\lambda)$ is called the (probability) density function of the random variable.

Among the admissible functions $\varphi(X)$, the following are of most interest:

(a) The identity function $\varphi(X) = X$, giving rise to the mean μ_X of the random variable X

$$\mu_X = E\{X\}$$

(b) Functions of the form $\varphi(X) = (X - \mu_X)^n$, $n = 1, 2, \ldots$, giving rise to the nth central moment of X

$$M_X^{(n)} = E\{(X - \mu_X)^n\}$$

For $n = 1$ we have trivially $M_X^{(1)} = 0$, while for $n = 2$ we obtain the variance σ_X^2 of X:

$$\sigma_X^2 = E\{(X - \mu_X)^2\} = E\{X^2\} - \mu_X^2$$

Not any arbitrary function $F_X(\lambda)$, (or $f_X(\lambda)$), corresponds to some random variable X. $F_X(\lambda)$ must have the properties

$$F_X(-\infty) = 0,\ F_X(+\infty) = 1,\ F_X(a) \leq F_X(b) \text{ for } a < b$$

(i.e., $F_X(\lambda)$ is nondecreasing), and if $a_1 < a_2 < \ldots < a_n < \ldots$ is a sequence of reals with $\lim_{n\to\infty} a_n = a$, then

$$\lim_{n\to\infty} F_X(a_n) = F_X(a)$$

i.e., F_X is continuous in the left.

Up to now only a single random variable has been considered. For a number of random variables $X_1, X_2, \ldots, X_n$, we can construct a vector $\overline{X} = [X_1, X_2, \ldots, X_n]^T$ and the mapping $X: \Omega \to R^n$ is called a random vector.

We have taken here all the random variables X_i to be defined on the same underlying probability space $\{\Omega, \mathcal{A}, P\}$.

For a real valued function $\varphi(\overline{X})$, $(\varphi: R^n \to R)$, the corresponding composite function $\varphi(\overline{X}(\omega))$, $(\varphi \circ \overline{X}: \Omega \to R)$, is a random variable (provided $\varphi \circ \overline{X}$ is measurable), and its expectation is simply defined as:

$$E\{\varphi(\overline{X})\} = \int_{\Omega} \varphi(\overline{X}(\omega))\, dP(\omega) \qquad (2.5)$$

If we define the distribution function

$$F_{\overline{X}}(\overline{\lambda}) = F_{\overline{X}}(\lambda_1, \lambda_2, .., \lambda_n) = P\{\omega;\ X_1(\omega) \leq \lambda_1, .., X_n(\omega) \leq \lambda_n\} \qquad (2.6)$$

and if the density function

$$f_{\overline{X}}(\overline{\lambda}) = f_{\overline{X}}(\lambda_1, \lambda_2, .., \lambda_n) = \frac{\partial^n F_{\overline{X}}(\lambda_1, \lambda_2, \ldots, \lambda_n)}{\partial \lambda_1\, \partial \lambda_2 \ldots \partial \lambda_n} \qquad (2.7)$$

exists, then

$$E\{\varphi(\overline{X})\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} \varphi(\overline{\lambda})\, f_{\overline{X}}(\overline{\lambda})\, d\lambda_1\, d\lambda_2 \ldots d\lambda_n \tag{2.8}$$

2.4 Stochastic Processes

The concept of a random variable is the probabilistic counterpart of the deterministic concept of a real valued constant. A similar counterpart to the deterministic concept of a function is provided by the concept of a stochastic process (or random process, or random function).

We now introduce the following notational convention: For a given function $y(t, s)$ ($y: T \times S \to Y$) of two variables $t \in T$ and $s \in S$, we denote by y^t and y^s the mappings $y^t: S \to Y$ and $y^s: T \to Y$ resulting by fixing t or s respectively in y to a constant value.

A stochastic process $\xi(t, \omega)$, $t \in T$, $\omega \in \Omega$ is a mapping $\xi: T \times \Omega \to R$, where T is an index set and $\{\Omega, \mathcal{A}, P\}$ a probability space. For every fixed $t \in T$, the corresponding mapping $\xi^t: \Omega \to R$ is a random variable, while for fixed ω the mapping $\xi^\omega: T \to R$ is an ordinary function with domain T and range in R. For the various fixed ω values the corresponding deterministic functions $\xi^\omega(t)$ are called the sample functions of the stochastic process.

If Φ denotes the set of functions $\varphi: T \to R$, then a stochastic process $\xi(t, \omega)$ can be alternatively viewed as a mapping $\xi: \Omega \to \Phi$, i.e., as a "function valued" random variable. In our discussion we shall take T to be the set of reals R or some interval in R. If T is the set of positive integers N^+, the term random (or stochastic) sequence will be used. If $T \in R^n$, then the corresponding mapping $\xi(t, \omega)$ will be called a random field.

In a description of a stochastic process, the probability measure P corresponding to the probability space $\{\Omega, \mathcal{A}, P\}$ is usually not given. Instead the joint distribution function of the random variables $\xi^{t_1}(\omega)$, $\xi^{t_2}(\omega)$, $\ldots$, $\xi^{t_n}(\omega)$ is given for any finite set of values $t_1, t_2, \ldots, t_n$ in T:

$$F_{\xi, t_1, t_2, .., t_n}(\lambda_1, \lambda_2, \ldots, \lambda_n) =$$

$$= P\left\{\omega;\ \xi^{t_1}(\omega) \leq \lambda_1,\ \xi^{t_2}(\omega) \leq \lambda_2,\ \ldots,\ \xi^{t_n}(\omega) \leq \lambda_n\right\} \tag{2.9}$$

Such a given family of distribution functions corresponds to some stochastic process, provided the two following properties are satisfied [Kolmogorov, 1950]:

(a) If $p_1, p_2, \ldots, p_n$ is a permutation of the indices 1, 2, ..., n, then

$$F_{\xi, t_{p_1}, t_{p_2}, \ldots, t_{p_n}}(\lambda_{p_1}, \lambda_{p_2}, \ldots, \lambda_{p_n}) =$$

$$= F_{\xi, t_1, t_2, \ldots, t_n}(\lambda_1, \lambda_2, \ldots, \lambda_n) \tag{2.10}$$

(b) If we let the variables $\lambda_{j+1}, \lambda_{j+2}, \ldots, \lambda_n$ approach infinity $(1 \leq j < n)$, then

$$F_{\xi, t_1, t_2, .., t_j, t_{j+1}, \ldots, t_n}(\lambda_1, \lambda_2, \ldots, \lambda_j, \infty, \infty, \ldots, \infty) =$$

$$= F_{\xi, t_1, t_2, \ldots, t_j}(\lambda_1, \lambda_2, \ldots, \lambda_j) \tag{2.11}$$

The above finite dimensional distribution functions of the process determine a good deal of the structure of the probability measure P but by no means all of it. For a discussion on this interesting problem we refer to [Lauritzen, 1973], where further references to the literature are also given.

With the help of the distribution functions or the corresponding finite dimensional density function (provided it exists),

$$f_{\xi, t_1, t_2, \ldots, t_n}(\lambda_1, \lambda_2, \ldots, \lambda_n) =$$

$$\frac{\partial^n F_{\xi, t_1, t_2, \ldots, t_n}(\lambda_1, \lambda_2, \ldots, \lambda_n)}{\partial\lambda_1\ \partial\lambda_2\ \ldots\ \partial\lambda_n} \tag{2.12}$$

we can now define a few more useful concepts.

The mean value function $\mu_\xi(t)$ of a stochastic process $\xi(t, \omega)$ is defined as

$$\mu_\xi(t) = E\{\xi(t,\omega)\} = \int_R \lambda \, dF_{\xi,t}(\lambda) = \int_{-\infty}^{+\infty} \lambda \, f_{\xi,t}(\lambda)\, d\lambda \tag{2.13}$$

Strictly speaking, for a fixed t value, the (constant valued) mean μ_ξ^t of the random variable $\xi^t(\omega)$ is defined in the usual way, and then the function $\mu_\xi(t)$ is constructed by letting t vary over the whole set T.

The (auto) covariance function $C_\xi(t, s)$ is a mapping $C: T \times T \to R$, defined as

$$C_\xi(t,s) = E\left\{ [\, \xi^t(\omega) - \mu_\xi(t) \,] \; [\, \xi^s(\omega) - \mu_\xi(s) \,] \right\} \tag{2.14}$$

The (auto)correlation function is a similar mapping defined as

$$R_\xi(t,s) = E\left\{ \xi^t(\omega) \; \xi^s(\omega) \right\} \tag{2.15}$$

A stochastic process with finite (auto)covariance is called a second-order stochastic process. Obviously,

$$C_\xi(t,s) = R_\xi(t,s) - \mu_\xi(t) \; \mu_\xi(s) \tag{2.16}$$

Given any stochastic process $\xi(t, \omega)$, a new stochastic process $\tilde{\xi}(t, \omega) = \xi(t, \omega) - \mu_\xi(t)$ can be constructed with the properties

$$\mu_{\tilde{\xi}}(t) = 0 \qquad \text{and} \qquad C_{\tilde{\xi}}(t,s) = R_{\tilde{\xi}}(t,s)$$

Given two stochastic processes $x(t, \omega)$ and $y(t, \omega)$, their cross-correlation and cross-covariance functions are simply defined as

$$R_{xy}(t,s) = E\left\{x^t(\omega) \; y^s(\omega)\right\} \tag{2.17}$$

and

$$C_{xy}(t,s) = E\left\{ [\, x^t(\omega) - \mu_x(t) \,] \; [\, y^s(\omega) - \mu_y(s) \,] \right\} \tag{2.18}$$

The autocovariance function has the following two properties:

(a) Symmetry: $C_\xi(t, s) = C_\xi(s, t)$

(b) The covariance function is positive in the sense that

$$\sum_i \sum_j a_i \; a_j \quad C_\xi(t_i, t_j) \quad \geq \quad 0$$

for every set $t_1, t_2, \ldots, t_n \in T$, and any set of constants $a_1, a_2, \ldots, a_n \in R^n$.

In a more familiar form, if we consider a matrix C with elements $C_{ij} = C_\xi(t_i, t_j)$ and any constant vector $a \in R^n$, then

$$C^T = C \qquad \text{and} \qquad a^T C\, a \geq 0,$$

i.e., the matrix C is symmetric and nonnegative definite.

2.5 Stationarity and Ergodicity

If the index set T of a stochastic process $\xi(t, \omega)$, $t \in T$ is taken to be the real line R, and its physical interpretation that of time, then questions about the time invariance of the probabilistic structure of the process lead to the concept of stationarity.

A stochastic process $\xi(t, \omega)$ is called strictly stationary if the finite dimensional distributions of the process have the property

$$F_{\xi, t_1, \ldots, t_n}(\lambda_1, \ldots, \lambda_n) \quad = \quad F_{\xi, t_1 + \tau, \ldots, t_n + \tau}(\lambda_1, \ldots, \lambda_n) \tag{2.19}$$

for every $\tau \in T \equiv R$ and every finite set $t_1, t_2, \ldots, t_n$.

A second-order stochastic process $\xi(t, \omega)$ is called weakly stationary (or wide sense stationary, or second-order stationary) if the mean value function $\mu_\xi(t)$ and the (auto)correlation function of the process $R_\xi(t, s)$ have the following properties:

$$\mu_\xi(t) = \mu_\xi(t + \tau) \qquad \text{and} \qquad R_\xi(t, s) = R_\xi(t + \tau, s + \tau) \quad \text{for every } t \in T.$$

This directly implies that $\mu_\xi(t) = \mu_\xi = \text{const.}$, and that

$$R_\xi(t, s) = R_\xi(t - s, 0) = R_\xi(t - s)$$

The same property follows directly for the (auto)covariance function

$$C_\xi(t,s) = R(t-s) - \mu_\xi^2 = C_\xi(t-s) \tag{2.20}$$

Every strictly stationary stochastic process is also a weakly stationary process, but the converse is not generally true. A class of stochastic processes, such that weak stationarity implies also strict stationarity, is that of Gaussian stochastic processes.

A stochastic process $\xi(t, \omega)$ is said to be Gaussian if for any set $t_1, t_2, \ldots, t_n \in T$, the corresponding random variables

$$\xi^{t_1}(\omega), \xi^{t_2}(\omega), \ldots, \xi^{t_n}(\omega)$$

have a multivariate Gaussian joint distribution.

Another important concept associated with the physical interpretation of the index set T as time is that of ergodicity. Probabilities, according to the limiting frequency approach at least, are associated with infinite ensembles of events. However, we frequently have to deal with a unique process and not an ensemble of such processes. For stationary processes whose probabilistic behavior is invariant with respect to time transformations, an infinite ensemble may be conceptually constructed from time shifts of the original process. Ensemble averages can thus be replaced by averages over the time domain.

A stochastic process $\xi(t, \omega)$ is ergodic if its probabilistic structure can be determined from a single realization $\xi^\omega(t)$. We need concern ourselves here only with the correlation function of the process. An estimate of the correlation function of a stationary process can be obtained by averaging a single realization over a time interval [-T, T]

$$R_T(\tau) = \frac{1}{2T} \int_{-T}^{+T} \xi(t+\tau)\ \xi(t)\ dt \qquad (2.21)$$

Obviously,

$$E\ \{R_T(\tau)\} = R(\tau) \qquad (2.22)$$

and if

$$\lim_{T\to\infty} R_T(\tau) = R(\tau) \quad . \qquad (2.23)$$

the process is said to be ergodic in (auto)correlation.

2.6 Linear Spaces --A Geometric Overview

A great deal of mathematical entities of a diverse nature (real numbers, vectors, functions, random variables, etc.) have a lot of structure and properties which only appear to be different. Their similarities can be brought to light if we deprive them of their particular characteristics and study them in a unified way based on their common ones. To identify such common characteristics a great deal of abstraction is necessary. Qualifying mathematical entities may be viewed as "points" in a "space" which is an abstraction of our familiar three-dimensional Euclidean space. The abstract counterparts of common geometric notions such as distance, length, angle, orthogonality, projection, etc. become then powerful tools of analysis.

We shall present here only a short account of some basic "geometric" results from linear algebra and functional analysis, culminating with the concept of a Hilbert space.

A linear space X is a collection of mathematical entities which we shall call elements, x, y, z, ..., together with two operations called addition and multiplication by a scalar, satisfying in connection with the field of reals R the following axioms:

(a) $x + y = y + x$

(b) $x + (y + z) = (x + y) + z$

(c) there exists an element $\emptyset \in X$, called the null element such that $x + \emptyset = x$ for every x in X

(d) for every $x \in X$, there exists a unique element $(-x)$ such that $x + (-x) = \emptyset$

(e) $a(bx) = (ab)x$ for all $x \in X$ and $a, b \in R$

(f) $a(x + y) = ax + ay$

(g) $(a + b)x = ax + bx$

(h) $1x = x$, $1 \in R$

The fundamental property of a linear space is that linear combinations of any finite number of its elements are also elements of the space (linearity property).

A set of elements x_i, $i = 1, 2, \ldots, n$ of a linear space is said to be linearly independent if the relation

$$\sum_{i=1}^{n} a_i \quad x_i \quad = \quad 0$$

holds only if $a_i = 0$ for all i.

A set of elements x_i of a linear space X is said to span X (or to be a spanning set of X) if every element in X can be expressed as a linear combination of the elements x_i.

A set of linearly independent elements which also spans a space X is said to be a basis for X.

The number of elements in a basis of a linear space X is called the dimension of the space. Alternatively, the dimension of a space is the maximum number of linearly independent elements in the space.

A linear space X is called a normed linear space if to each $x \in X$ corresponds a real number $\|x\|$ having the following properties:

(a) $\|x\| \geq 0$

(b) $\|x\| = 0$ if and only if $x = \emptyset$

(c) $\|ax\| = |a| \, \|x\|$ for every $a \in R$

(d) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)

$\|x\|$ is called the norm of x.

A metric space is a set of elements such that for every pair of elements a real valued function d(x, y) is defined, called the metric of the space, and having the following properties:

(a) $d(x, y) \geq 0$

(b) $d(x, y) = 0$ if and only if $x = y$

(c) $d(x, y) = d(y, x)$

(d) $d(x, y) + d(y, z) \geq d(x, z)$

The metric is an abstraction of the usual concept of distance while the norm is an abstraction of length. Every normed linear space is also a metric space with the following definition of the metric

$$d(x, y) = \|x - y\|$$

A linear space X is called an inner product space if for every pair of its elements a function $< x, y >$, called the inner product of X, is defined with the following properties:

(a) $< x + y, z > = < x, z > + < y, z >$ $\qquad x, y, z \in X$

(b) $< x, y > = < y, x >$

(c) $< ax, y > = a < x, y >$ $\qquad a \in R$

(d) $< x, x > \geq 0$ and $< x, x > = 0$ if and only if $x = \emptyset$

Every inner product space is also a normed linear space with norm defined as $\|x\| = < x, x >^{\frac{1}{2}}$ and, consequently, also a metric space.

It is essential to realize that the definition of a certain inner product (or norm, or metric) is by no means unique for a certain linear space. On the contrary, it may be possible to define a number (not necessarily finite) or inner products (or norms, or metrics) over one and the same linear space, thus giving rise to a number of different inner product spaces (or normed linear spaces, or metric spaces).

The angle θ between two nonzero elements x, y of an inner product space X is defined by means of

$$\cos \theta = \langle x, y \rangle / \|x\| \, \|y\| \qquad 0 \leq \theta \leq \pi$$

For $\theta = \pi/2$ we have $\langle x, y \rangle = 0$, and we say that x and y are mutually orthogonal ($x \perp y$).

A set of elements x_i of an inner product space X is called an orthogonal set if $\langle x_i, x_j \rangle = 0$ for $i \neq j$ and $\langle x_i, x_i \rangle \neq 0$. A similar set x_i^* is called orthonormal if it is an orthogonal set and, in addition, $\langle x_i^*, x_i^* \rangle = 1$. A basis with orthogonal (orthonormal) elements is called an orthogonal (orthonormal) basis.

Given a set of orthonormal elements x_i^* of an inner product space X and an arbitrary element $y \in X$, we call the series

$$\sum_i \langle y, x_i^* \rangle \; x_i^*$$

the Fourier series of y with respect to the set x_i^*. For an orthogonal set x_i, the Fourier series of y is

$$\sum_i \frac{\langle y, x_i \rangle}{\|x_i\|^2} \; x_i$$

Given two elements x, y of an inner product space X, we call the element $\|x\|^{-2} \langle y, x \rangle \; x$, the projection of y on x. The Fourier series of an element y with respect to an orthogonal set x_i, is therefore the sum of the projections of y on the elements x_i.

A subset M of a linear space X is called a linear subspace of X, if M is itself a linear space.

Given a sequence x_n of elements of a metric space X, we say that the sequence converges to an element x of X if $\lim_{n\to\infty} d(x_n, x) = 0$. This is convergence in metric, and since normed linear spaces are also metric spaces, we can also define convergence in norm by $\lim_{n\to\infty} \| x - x_n \| = 0$.

A sequence of elements x_i of a metric space X is called a Cauchy sequence, if for every $\epsilon > 0$ there exists an integer N_ϵ such that $d(x_m, x_n) \leq \epsilon$ for all $m, n \geq N_\epsilon$.

A metric space X is called complete if every Cauchy sequence in X has a limit also in X, i.e., if it contains the limits of all its Cauchy sequences. Since inner product and normed linear spaces are also metric spaces, we can define two new types of spaces:

A complete normed linear space is called a Banach space.

A complete inner product space is called a Hilbert space.

Let X be a metric space and S a subspace of X. The closure of S denoted by $\overline{S}$ is the set of all limits of convergent sequences of S. A subset S of the metric space X is called closed if $\overline{S} = S$, i.e., if it contains the limits of all its convergent sequences. A subset S of the metric space X is called dense in X if $\overline{S} = X$. A metric space X is separable if there exists a countable dense set in it. A separable Hilbert space contains a countable number of elements such that the subspace they span is identical to the Hilbert space.

Let H be a (separable) Hilbert space and M a closed linear subspace of H. An element $x \in H$ is said to be orthogonal to M ($x \perp M$), if it is orthogonal to every element y of M. The set $M^\perp$ of all elements of H orthogonal to M is a closed subspace of H, called the orthogonal complement of M with respect to H. One can show that for every closed M, H is the direct sum of the orthogonal subspaces M and $M^\perp$, and we denote this by $H = M \oplus M^\perp$. This means that any element x of H can be decomposed in a

unique way $x = \hat{x} + x^{\perp}$, so that $\hat{x} \in M$ and $x^{\perp} \in M^{\perp}$. $\hat{x}$ is called the projection of x on M, and we denote this by $\hat{x} = \mathcal{P}_M(x)$. We also have $x^{\perp} = x - \hat{x} = \mathcal{P}_{M^{\perp}}(x)$.

Let x_o be a fixed element of a linear space X, and let x_i, i = 1, 2, ..., n be a set of linearly independent elements of X. The set of all linear combinations of the form

$$x_0 + \sum_i a_i x_i, \qquad a_i \in R \tag{2.24}$$

is called a linear variety of dimension n. A linear variety is the space of elements $x = x_o + x'$, where x' belongs to M, the subspace of X spanned by the set x_i. A linear variety V may be viewed as a translation of a linear subspace M by a fixed element x_o, and we write symbolically $V = x_o + M$.

Let x_i be a set of n linearly independent elements of an inner product space X and c_i a set of n real constants. The set of all elements y of X satisfying

$$< y, x_i > = c_i \qquad \text{for all } i \tag{2-25}$$

is called a hyperplane P of co-dimension n.

Let y_o be a fixed element of P. For every element y of P we have $< x_i, y - y_o > = < x_i, y > - < x_i, y_o > = c_i - c_i = 0$. It follows that $y - y_o \perp M$, where M is the subspace spanned by the set x_i. Every element $y \in P$ can be written in the form $y = y_o + y'$ where $y' = y - y_o \in M^{\perp}$, and the hyperplane P can be viewed as a linear variety $P = y_o + M^{\perp}$.

A functional is a mapping with domain a linear space X and range the set of reals R. A linear functional is a functional L with the property $L(ax + by) = a\,L(x) + b\,L(y)$ where $x, y \in X$ and $a, b \in R$. If X is a normed linear space, a linear functional L on X is said to be bounded if there exists a constant $M \in R$, such that $L(x) \leq M \| x \|$ for all $x \in X$. A linear functional L over a normed linear space X is said to be continuous if for every sequence x_n of X converging in norm to $x \in X$, $L(x_n) \to L(x)$ in the usual sense. A linear functional over a normed linear space is bounded if and only if it is

continuous. Given a normed linear space X, the set of all bounded linear functionals over X is itself a normed linear space X^*, called the normed conjugate or dual space of X, with norm defined as

$$\| L \| = \sup_{\substack{x \in X \\ x \neq \emptyset}} \frac{| L(x) |}{\| x \|} \tag{2.26}$$

For any linear bounded functional L over a Hilbert space H there exists a unique element x_L of H, called the representer of L, such that $L(x) = \langle x_L, x \rangle$ for every $x \in H$. We usually denote the dual space of H by H^* and write $x^*, y^*, z^*, \ldots$ for elements of H^* with representers x, y, z in H.

Let H be a Hilbert space with inner product $\langle \cdot , \cdot \rangle$, H^* its dual, and $(\Omega, \mathcal{A}, P)$ a probability measure space. A mapping $x: \Omega \to H$ is said to be a Hilbert space-valued random variable if

$$y^*(x^\omega) = \langle x^\omega, y \rangle \tag{2-27}$$

is an ordinary random variable for every $y^* \in H^*$ (or equivalently for every $y \in H$, where y is, of course, the representer of y^*).

2.7 Best Linear Approximation and the Normal Equations

The problem of best linear approximation may be defined as follows: Given a normed linear space X and a closed linear subspace M of X, find the element $\hat{y}$ of M which best approximates a given element x of X, in the sense that

$$\| x - \hat{y} \| = \min_{y \in M} \| x - y \| \tag{2.28}$$

To solve the problem in a Hilbert space, consider the unique decomposition of x into its projections on M and $M^\perp$:

$$x = \hat{x} + x', \qquad \hat{x} = \mathcal{P}_M(x), \qquad x' = \mathcal{P}_{M^\perp}(x)$$

For any arbitrary element $y \in M$, we have, taking into account that $\hat{x} \perp x'$, $x' \perp y$,

$$\|x - y\|^2 = \|x'\|^2 + \|\hat{x} - y\|^2 \qquad (2\text{-}29)$$

The first term is fixed and the second, nonnegative; and therefore $\|x - y\|^2$ is minimized by setting $y = \hat{y} = \hat{x} = \mathcal{P}_M(x)$. The projection of x on M is the closest element to x among all elements of M in the sense of the norm of X. Furthermore, the best approximation always exists and it is unique.

If X is an inner product space and M the span of a finite set of n elements x_i, the best approximation $\hat{x}$ to an element x of X from M can be written as

$$\hat{x} = \sum_i a_i \, x_i$$

where the n coefficients a_i are to be determined. Since $\hat{x} \in M$ and $x - \hat{x} \in M^{\perp}$, we have $x - \hat{x} \perp x_i$ for all i; and $\langle x - \hat{x}, x_i \rangle = 0$, or

$$\sum_j a_j \langle x_i, x_j \rangle = \langle x_i, x \rangle \qquad (2.30)$$

These are the "normal equations," and can be written in matrix form as follows:

$$\begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \cdots & \langle x_1, x_n \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \cdots & \langle x_2, x_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_n, x_1 \rangle & \langle x_n, x_2 \rangle & \cdots & \langle x_n, x_n \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \langle x_1, x \rangle \\ \langle x_2, x \rangle \\ \vdots \\ \langle x_n, x \rangle \end{bmatrix}$$

or in compact notation

$$Ga = u \qquad (2\text{-}31)$$

where G is called the Gram matrix. If M has dimension n then the elements x_i forming a basis for M are linearly independent, and the Gram matrix G is nonsingular. A unique set of coefficients can be obtained by

$$a = G^{-1} u \qquad (2\text{-}32)$$

The distance of x from its best approximation $\hat{x}$ (interpreted as the "error of approximation") is provided by the induced metric

$$d(x, \hat{x}) = \| x - \hat{x} \|$$

It can easily shown that

$$\| x - \hat{x} \|^2 = \| x \|^2 - \| \hat{x} \|^2 = \| x \|^2 - \sum_i \sum_j a_i a_j \langle x_i, x_j \rangle =$$

$$= \| x \|^2 - a^T G a = \| x \|^2 - u^T G^{-1} u \qquad (2.33)$$

If M is of dimension less than n, then the Gram matrix is simgular; and the set of coefficients a_i cannot be uniquely defined. However, the best approximation exists and is unique; and the normal equations have infinite number of solution a_i, each of them giving rise to the same best approximation $\hat{x} = \sum_i a_i x_i$.

A related problem is that of finding the element of minimal norm among the elements of a linear variety or a hyperplane. This can be viewed as a problem of best approximating the null element $\emptyset$ from a linear variety or hyperplane.

Given an inner product space X and a linear variety $V = x_o + M$ in X, we can decompose each element y of V as

$$y = \hat{y} + y', \qquad \hat{y} = \mathcal{P}_M(y), \qquad y' = \mathcal{P}_{M^\perp}(y)$$

Since $\hat{y} \perp y'$,

$$\| y \|^2 = \| \hat{y} + y' \|^2 = \| \hat{y} \|^2 + \| y' \|^2 \qquad (2\text{-}34)$$

We can further show that y' is the same for every $y \in V$, as follows: Consider two elements $y_1 = x_0 + \bar{y}_1$, $y_2 = x_0 + \bar{y}_2$ of V, where $\bar{y}_1, \bar{y}_2 \in M$. It follows that $y_1 - y_2 = \bar{y}_1 - \bar{y}_2 \in M$. Decomposing as usual,

$$y_1 = \hat{y}_1 + y_1', \qquad y_2 = \hat{y}_2 + y_2', \qquad \hat{y}_1, \hat{y}_2 \in M, \qquad y_1', y_2' \in M^\perp,$$

we have that

$$\hat{y}_1 - \hat{y}_2 \in M, \qquad y_1 - y_2 \in M$$

and

$$y_1' - y_2' = (y_1 - y_2) - (\hat{y}_1 - \hat{y}_2) \in M$$

But since y_1', $y_2' \in M^\perp$, we also have $y_1' - y_2' \in M^\perp$. This is only possible if $y_1' - y_2' = \emptyset$, and $y_1' = y_2'$.

To minimize

$$\| y \|^2 = \| \hat{y} \|^2 + \| y' \|^2 \tag{2-35}$$

where $\| y' \|^2$ is fixed and $\| \hat{y} \|^2$ is nonnegative, we must set $\hat{y} = \emptyset$. It follows that the element of V with minimal norm is the unique element

$$y' = \mathcal{P}_{M^\perp}(y) \qquad \text{for every } y \in V.$$

In the case of a hyperplane P described by

$$< y, x_i > = c_i \qquad i = 1, 2, \ldots, n \qquad y \in P$$

we can write P as a linear variety $P = y_o + M^\perp$, where y_o is a fixed element of P and M is the span of the set x_i. The element y' of P with minimal norm is then given by the unique projection of any element $y \in P$ on $(M^\perp)^\perp = M$

$$y' = \mathcal{P}_M(y)$$

To determine y', consider any element y of P. Then $y' = \mathcal{P}_M(y) = \sum_i a_i x_i$ can be found with the help of the normal equations

$$\sum_j a_j < x_i, x_j > = < x_i, y > \tag{2-36}$$

Since $y \in P$ we have $< x_i, y > = c_i$, so that the normal equations become

$$\sum_j a_j < x_i, x_j > = c_i \tag{2-37}$$

or, in matrix notation, $Ga = c$. The elements x_i are linearly independent so that $a = G^{-1} c$ uniquely determines $y' = \sum_i a_i x_i$.

The importance of the normal equations in solving estimation problems can hardly be overemphasized. In the next chapter we shall apply them in deriving a number of apparently different algorithms.

3. ESTIMATION TECHNIQUES

3.1 Introductory Remarks

The objective of any analysis of geodetic data is to obtain optimal values for observables differing from actual observations because of measurement errors, for parameters functionally related to the above observables, and, in the most general case, functions related to observables and/or parameters. The essential question, of course, is what are the optimality criteria to be satisfied by such optimal values.

Before establishing specific optimality criteria, we must notice that, in general, the optimal values sought are associated with some corresponding "true values." Although such true values are difficult to define and even impossible to materialize in practice, we can nevertheless draw the following general outline for any reasonable optimality criterion: Optimal values should be as close as possible to true values.

Both optimal and true values may therefore be modeled as elements in an abstract mathematical space (i.e., a set of such elements), where, in addition, the concept of distance is defined enabling us to determine how "close" any two elements of the space are. But such a mathematical model is provided by the concept of a metric space where a metric or distance is defined for every pair of elements. The choice of a specific metric corresponds to the choice of a specific optimality criterion.

We shall mostly confine ourselves here to a specific (but nevertheless of wide applicability) kind of metric spaces, namely, linear complete inner product spaces (Hilbert spaces), where the metric $\rho(x, y)$ follows from the definition of the corresponding inner product through the relation

$$\rho(x,y) = \| x - y \| = \left\{ < x - y, x - y > \right\}^{\frac{1}{2}}$$

In such a mathematical setup, we shall find our problem of obtaining optimal estimates to correspond to the problem of best approximating an element of the space from the elements of some linear subspace. The solution to this problem is given by the celebrated normal equations, and computational algorithms will always result from applications of this solution. The resulting algorithms may well be drastically different, but our common approach of solution will provide us with a unified view, where similarities and dissimilarities between what is usually considered to be different estimation techniques can be sharply identified.

Such an analysis can, of course, take place at the algorithm level, where by proper algebraic manipulations one can show how different algorithms can be derived from a common starting point. Such approaches can be found, e.g., in [Liebelt, 1967, Chapter 5] and in the recent work by Krakiwsky [1975]. It is our belief, however, that little is to be gained from approaches where the connection between different algorithms is explored without reference to a unique underlying problem formulation and solution. It is hoped that the approach taken here will contribute to the understanding of varying estimation techniques, especially since the geometric character of Hilbert space techniques is more appealing intuitively, after familiarization with these mathematical tools is attained.

We have used the term estimation techniques repeatedly here, although estimation is usually connected with statistical and probability concepts completely absent from our discussion. Indeed all the probabilistic notions one needs can be condensed in the definition of the inner product involved, and they are only involved in justifying the optimality of the corresponding metric as an appropriate measure of "closeness."

A more appropriate term might have been "adjustment techniques," but unfortunately this term is already connected to some specific algorithms.

The solutions to our problems are quite independent of the specific inner product definition. However, the corresponding computational algorithms result by replacing the specific inner product into the general

normal equations solution, and are, therefore, related to both probabilistic concepts and deterministic mathematical tools from approximation theory.

3.2 Least Squares Adjustment

Among various estimation techniques, least squares adjustment is the one most widely used in geodesy since the time of its development by Gauss. There is, therefore, very little to be said about algorithms associated with least squares. Usually these algorithms are derived from variational principles, as solutions to the problem of minimizing a quadratic form. Such an approach solves the problem but has little to offer to the understanding of its mathematical context and its relation to other techniques.

Instead, we shall approach here the problem from a Hilbert space point of view, and we will show how particular algorithms can be directly derived from solutions to problems already considered in Chapter 2.

The notation will generally follow that of Uotila [1967], with some minor changes of signs, so that our results can be easily compared to the variationally derived algorithms in that work.

3.2.1 Method of Observation Equations (Adjustment by Elements)

The problem to be solved here can be described as follows: A finite set of unknown observables represented by a vector $L^* \in R^n$, is a priori known to be in a linear functional relation

$$L^* = A X \tag{3.1}$$

with a vector $X \in R^u$ representing a finite set of unknown parameters, where A is a known $n \times u$ matrix with $n > u$.

A vector L of available (known) observations L differs from L^* because of the presence of unknown observational errors V, according to

$$L = L^* + V \quad \text{or} \quad L = A X + V \tag{3.2}$$

This linear matrix equation (observation equations) is satisfied by an infinite

number of pairs of values of the unknown vectors X, V, since to every element X in R^u corresponds some element $V = L - AX$ in R^n. Among all such possible pairs of solutions X, V, we seek to find one $\hat{X}$, $\hat{V} = L - A\hat{X}$, which minimizes the quadratic form

$$V^T P V = (L - AX)^T P (L - AX) \tag{3.3}$$

where P is a given positive definite n × n matrix.

Setting $y = AX$, $\hat{y} = A\hat{X}$, we can reformulate our problem as follows: Let $\mathcal{L}$ be the n-dimensional complete inner product space (Hilbert space) with elements n × 1 real vectors and inner product

$$< f, g > = g^T P f \qquad f, g \in \mathcal{L} \tag{3.4}$$

Let A_i denote the i^{th} column of the matrix A. Obviously $L \in \mathcal{L}$, $A_i \in \mathcal{L}$ for i = 1, 2, ..., u and since

$$y = A X = \sum_{i=1}^{n} X_i A_i \tag{3.5}$$

it follows that $y \in \mathcal{A}$, where $\mathcal{A}$ = span $(A_1, A_2, \ldots, A_u)$. The least squares criterion $V^T P V = \min$ can now be written as

$$V^T P V = < V, V > = \| V \|^2 = \| L - y \|^2 = \min \tag{3.6}$$

We seek to find $\hat{y} \in \mathcal{A}$ such that

$$\| L - \hat{y} \| = \min_{y \in \mathcal{A}} \| L - y \| \tag{3.7}$$

But this is a problem of best approximating $L \in \mathcal{L}$ from the elements of the linear subspace $\mathcal{A} \subset \mathcal{L}$.

The solution is well known to be provided by the normal equations

$$\begin{bmatrix} \langle A_1, A_1 \rangle & \langle A_1, A_2 \rangle & \cdots & \langle A_1, A_u \rangle \\ \langle A_2, A_1 \rangle & \langle A_2, A_2 \rangle & \cdots & \langle A_2, A_u \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle A_u, A_1 \rangle & \langle A_u, A_2 \rangle & \cdots & \langle A_u, A_u \rangle \end{bmatrix} \begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \vdots \\ \hat{X}_u \end{bmatrix} = \begin{bmatrix} \langle L, A_1 \rangle \\ \langle L, A_2 \rangle \\ \vdots \\ \langle L, A_u \rangle \end{bmatrix} \tag{3.8}$$

To derive a computational algorithm from this solution, we make use of the inner product definition; and after replacing $\langle A_i, A_j \rangle = A_i^T P A_j = A_j^T P A_i$ and $\langle L, A_i \rangle = A_i^T P L$, we obtain

$$\begin{bmatrix} A_1^T P A_1 & A_1^T P A_2 & \cdots & A_1^T P A_u \\ A_2^T P A_1 & A_2^T P A_2 & \cdots & A_2^T P A_u \\ \vdots & \vdots & \ddots & \vdots \\ A_u^T P A_1 & A_u^T P A_2 & \cdots & A_u^T P A_u \end{bmatrix} \begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \vdots \\ \hat{X}_u \end{bmatrix} = \begin{bmatrix} A_1^T P L \\ A_2^T P L \\ \vdots \\ A_u^T P L \end{bmatrix} \tag{3.9}$$

or

$$\begin{bmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_u^T \end{bmatrix} P \begin{bmatrix} A_1 & A_2 & \cdots & A_u \end{bmatrix} \hat{X} = \begin{bmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_u^T \end{bmatrix} P L$$

or simply

$$(A^T P A) \hat{X} = A^T P L \tag{3.10}$$

If the columns of A are linearly independed, then $A^T PA$ is a nonsingular matrix so that we may obtain

$$\hat{X} = (A^T P A)^{-1} A^T P L \tag{3.11}$$

$$\hat{y} = A \hat{X} = A (A^T P A)^{-1} A^T P L \tag{3.12}$$

$$\hat{V} = L - \hat{y} = L - A (A^T P A)^{-1} A^T P L \tag{3.13}$$

a set of well-known results! Compare, e.g., with [Uotila, 1967, Section 3.1].

The solution is illustrated in Figure 3.1. The $\hat{X}_i$'s are the coordinates of the best approximation $\hat{y}$ to L, from the subspace $\mathcal{A}$, with respect to the basis $\{A_i\}$ in this subspace. We also have $\hat{V} \perp \mathcal{A}$ as it can easily be verified

$$\hat{V}^T P A = [\, \hat{V}^T P A_1, \ \hat{V}^T P A_2, \ldots, \hat{V}^T P A_u \,] =$$

$$= (L^T - L^T P A (A^T P A)^{-1} A^T) P A = 0 \tag{3.14}$$

i.e., $\hat{V}^T P A_i = 0 = \langle A_i, \hat{V} \rangle$ or $\hat{V} \perp A_i$ $i = 1, 2, \ldots, u$, and $\hat{V} \perp \mathcal{A}$.

3.2.2 Condition Equations

In this problem a set of unknown observables $L^* \in R^n$ is a priori known to satisfy a linear relation

$$B L^* = 0 \tag{3.15}$$

where B is a known $m \times n$ with $m < n$ and rank (B) = m. A corresponding set of known available observations $L \in R^n$ differs from L^* because of the presence of unknown observational errors $V \in R^n$, according to the relation

$$L = L^* + V \tag{3.16}$$

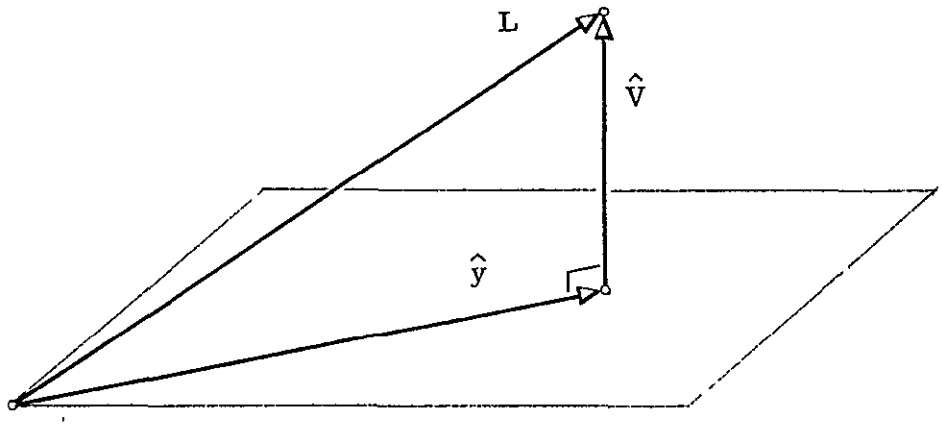

Figure 3.1 The Geometry of Least Squares Adjustment (Observation Equations)

Since B and L are known, the vector $W = BL$ is also known and in view of $BL^* = 0$, it satisfies the relation (condition equations)

$$BV = W \tag{3.17}$$

It can be shown that since rank $(B) = m < n$, the matrix equation $BV = W$ has a solution and, furthermore, an infinite number of solutions. Among all vectors V satisfying $BV = W$, we seek to find one $\hat{V}$ which minimizes the quadratic form $V^T P V$ where P is a given positive definite $n \times n$ matrix.

To reveal the geometry of the problem in a Hilbert space context, we use the transformation

$$\tilde{B} = P^{-1} B^T$$

($\tilde{B}$ is now $n \times m$), so that

$$W = BV = BP^{-1} PV = \tilde{B}^T PV \tag{3.18}$$

If $\tilde{B}_i$ denotes the i^{th} column of $\tilde{B}$, we have

$$\tilde{B}_i^T PV = W_i \tag{3.19}$$

Introducing the Hilbert space $\mathcal{L}$ of $n \times 1$ real vectors with inner product

$$\langle f, g \rangle = g^T Pf \qquad g, f \in \mathcal{L} \tag{3.20}$$

we can write the above relation in the form

$$\langle \tilde{B}_i, V \rangle = W_i \tag{3.21}$$

Since rank $(B) = m$, we also have rank $(\tilde{B}) = m$ and the columns $\tilde{B}_i$ are linearly independent. It follows that V satsifies $BV = W$ if it belongs to the set

$$H = \{V; \langle \tilde{B}_i, V \rangle = W_i \text{ for } i = 1, 2, \ldots, m\} \tag{3.22}$$

The set H can be directly identified as a hyperplane of co-dimension m in $\mathcal{L}$. If $N = \text{span } \{B_1, B_2, \ldots, B_n\}$, then the hyperplane H can be identified with a linear variety, with the help of the unique fixed element $V_0 = \mathcal{P}_N(V)$ for every $V \in H$ (see Section 2.7),

$$H = \{V;\ V = V_0 + V',\ V' \in N^{\perp}\} \tag{3.23}$$

For any V in H we have the orthogonal decomposition $V = V_0 + V'$, where $V_0 \in N$ is fixed, $V' \in N^{\perp}$, and, therefore, $V_0 \perp V'$.

Using the Pythagorean theorem, we may write

$$V^T P V = \langle V, V \rangle = \|V\|^2 = \|V_0\|^2 + \|V'\|^2 \tag{3.24}$$

Since V_0 is fixed, the above norm is minimized for $V' = 0$, and our solution to $V^T P V = \min$ is $\hat{V} = V_0$. To find explicitly this value, we consider any element $V \in H$, and we have

$$\hat{V} = V_0 = \mathcal{P}_N(V) \tag{3.25}$$

Since $N = \text{span}(\tilde{B}_1, \tilde{B}_2, \ldots, \tilde{B}_n)$, this projection is provided by the normal equations

$$\begin{bmatrix} \langle \tilde{B}_1, \tilde{B}_1 \rangle & \langle \tilde{B}_1, \tilde{B}_2 \rangle & \cdots & \langle \tilde{B}_1, \tilde{B}_n \rangle \\ \langle \tilde{B}_2, \tilde{B}_1 \rangle & \langle \tilde{B}_2, \tilde{B}_2 \rangle & \cdots & \langle \tilde{B}_2, \tilde{B}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \tilde{B}_n, \tilde{B}_1 \rangle & \langle \tilde{B}_n, \tilde{B}_2 \rangle & \cdots & \langle \tilde{B}_n, \tilde{B}_n \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \langle \tilde{B}_1, V \rangle \\ \langle \tilde{B}_2, V \rangle \\ \\ \langle \tilde{B}_n, V \rangle \end{bmatrix} \tag{3.26}$$

and

$$\hat{V} = \sum_{i=1}^{n} a_i \tilde{B}_i \tag{3.27}$$

Using the definition of inner product in $\mathcal{L}$ and the fact that $\langle \tilde{B}_i, V \rangle = \tilde{B}_i^T P V = W_i$ for any $V \in H$, we obtain

$$\begin{bmatrix} \tilde{B}_1^T \\ \tilde{B}_2^T \\ \vdots \\ \tilde{B}_n^T \end{bmatrix} P \; [\, \tilde{B}_1, \tilde{B}_2, \ldots, \tilde{B}_n \,] \; a = W, \quad \text{or} \quad (\tilde{B}^T P \tilde{B})\, a = W \tag{3.28}$$

Since $\tilde{B} = P^{-1} B^T$, and introducing $M = B P^{-1} B^T$, we get

$$B P^{-1} P P^{-1} B^T a = B P^{-1} B^T a = M a = W \tag{3.29}$$

and

$$\hat{V} = \sum_{i=1}^{n} B_i \, a_i = \tilde{B} \, a = P^{-1} B^T a \tag{3.30}$$

Since the rows of B are linearly independent, M is invertible so that

$$a = M^{-1} W \tag{3.31}$$

and

$$\hat{V} = P^{-1} B^T M^{-1} W \tag{3.32}$$

a well known result! Compare, e.g., with [Uotila, 1967, Section 3.2]. The solution is illustrated in Figures 3.2a and 3.2b.

3.2.3 Generalized Model

A more general least squares model is a combination of observation and condition equations of the form

$$W = AX + BV \tag{3.33}$$

where W is a known $n \times 1$ vector, A and B are known $n \times u$ and $n \times m$ matrices, respectively (rank (B) = n, rank (A) = u, $u < n < m$), and X and V are unknown $u \times 1$ and $m \times 1$ vectors, respectively.

Among all possible pairs X, V satisfying $AX + BV = W$, we want to find one $\hat{X}$, $\hat{V}$ such that $\hat{V}$ minimizes the quadratic form $V^T P V$, where P is

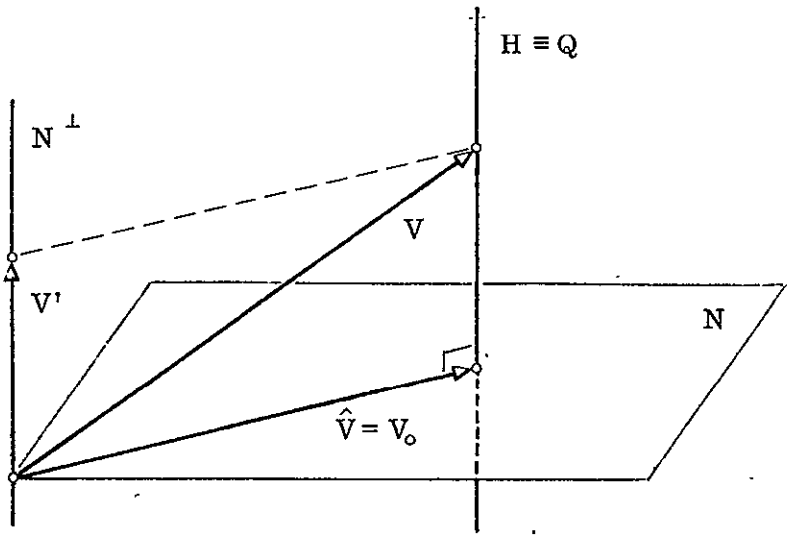

Figure 3.2 a The Geometry of Least Squares Adjustment (Condition Equations)

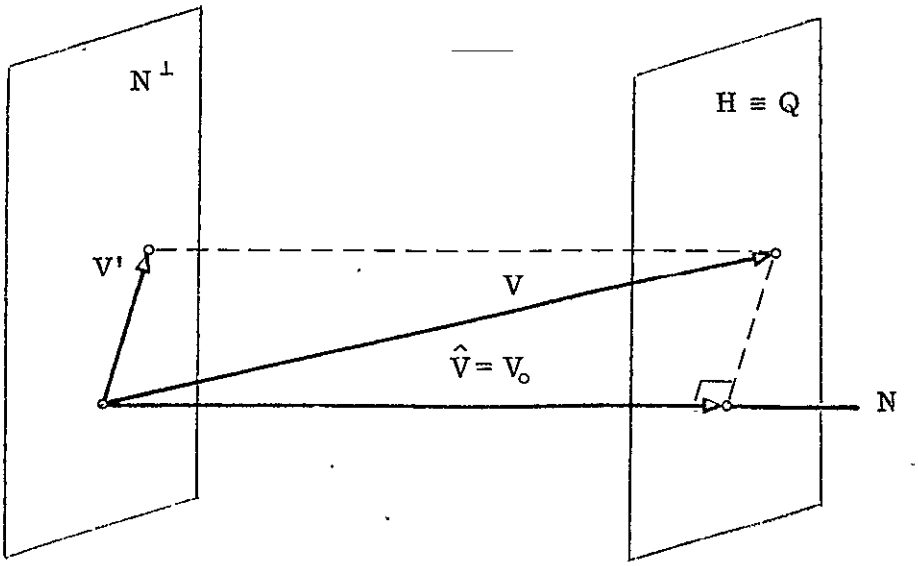

Figure 3.2 b The Geometry of Least Squares Adjustment (Condition Equations)

a given positive definite n × n matrix.

We shall solve this problem through the following device: For every fixed X we shall find the unique V_X minimizing $V^T P V$ among all V satisfying $BV = W_X$ where $W_X = W - AX$ is a fixed vector. Then by letting X vary over R^u, we shall find the unique vector $\hat{V}$ minimizing $V_X^T P V_X$ among all the previously obtained vectors V_X. In this way $\hat{V}$ will also minimize $V^T P V$ among all vectors V belonging to pairs X, V satisfying $AX + BV = W$. The only objection to such an approach is that there might exist two pairs $\hat{X}_1$, $\hat{V}$ and $\hat{X}_2$, $\hat{V}$, both satisfying $AX + BV = W$ and with $\hat{X}_1 \neq \hat{X}_2$. This possibility can be ruled out by subtracting $A\hat{X}_2 + B\hat{V} = W$ from $A\hat{X}_1 + B\hat{V} = W$ to obtain $A(\hat{X}_1 - \hat{X}_2) = 0$ and noting that $Ay = 0$ implies $y = 0$. Indeed, since rank (A) $= u$, the columns A_i of A are linearly independent and $0 = Ay = \sum_{i=1}^{n} y_i A_i$ implies $y_i = 0$ for all i, so that $y = 0$.

To proceed with the solution we fix X to obtain

$$BV = W_X \qquad \text{where } W_X = W - AX \tag{3-34}$$

This is simply a condition equations model, and $V^T P V$ is minimized by

$$V_X = P^{-1} B^T M^{-1} W_X = P^{-1} B^T M^{-1} (W - AX) \tag{3.35}$$

where

$$M = BP^{-1}B^T$$

Introducing

$$\tilde{L} = P^{-1} B^T M^{-1} W \qquad \text{and} \qquad \tilde{A} = P^{-1} B^T M^{-1} A$$

we obtain

$$V_X = \tilde{L} - A\tilde{X} \qquad \text{or} \qquad \tilde{L} = \tilde{A}X + V_X \tag{3.36}$$

This is an observation equations model and the answer to the minimization of $V_X^T P V_X$ (and consequently of $V^T P V$) is given by $\hat{V} = \tilde{L} - \tilde{A}\hat{X}$ where

$$\hat{X} = (\tilde{A}^T P \tilde{A})^{-1} \tilde{A}^T P \tilde{L} =$$

$$= (A^T M^{-1} B P^{-1} P P^{-1} B^T M^{-1} A)^{-1} A^T M^{-1} B P^{-1} P P^{-1} B^T M^{-1} W =$$

$$= (A^T M^{-1} A)^{-1} A^T M^{-1} W , \qquad (3.37)$$

and

$$\hat{V} = \tilde{L} - A \tilde{X} = P^{-1} B^T M^{-1} W - P^{-1} B^T M^{-1} A \hat{X} =$$

$$= P^{-1} B^T M^{-1} (W - A \hat{X}) \qquad (3.38)$$

These are exactly the results we obtain through variational approaches as it can be seen by a comparison with equations (123), (120) and (122) in [Uotila, 1967, p. 57].

3.2.4 Probabilisitic Justification of Least Squares

The philosophy of least squares techniques can be summarized as follows: A vector of n observables is a priori known to belong to a set of vectors satisfying a certain mathematical model. The linearity of the model leads to a linear subspace or hyperplane of R^n as the set of vectors satisfying the model equations. We shall call this set the "model space." When a vector of n observations corresponding to the vector of n observables in question is realized, it is in general found to be outside the model space because of observational errors. It is only natural then to suggest as an estimate of the observables an element of the model space which is as close as possible to the observations vector. This approach incorporates the anticipation of observational errors being small or at least not arbitrarily large. What is more arbitrary is the introduction of a certain metric, which depends on an arbitrary symmetric positive definite weight matrix P, through the relation

$$\rho(x,y) = \| x-y \| = \sqrt{\langle x-y, x-y \rangle} = \{ (x-y)^T P (x-y) \}^{\frac{1}{2}}$$

The above P-dependent inner product introduces a certain geometry and turns R^n into a Hilbert space L_P^n.

The selection of the estimate of the observables depends on the metric, i.e., on the matrix P, and is therefore arbitrary to the extent that P is arbitrary.

The question of metric optimality, i.e., of the selection of an optimal weight matrix P, finds an answer when the observational errors are modeled as random quantities with zero expectation and finite variance-covariance matrix. We shall examine each particular least squares method separately in this respect.

Observation equations. The model now becomes probabilistic. The unknown observables $L^* \in R^n$ relate to the unknown parameters $X \in R^u$ by $L^* = AX$ where A is a known $n \times u$ matrix with rank (A) = u.

The actually-obtained (known) observations L^ω relate to the unknown observation errors V^ω by $L^\omega = L^* + V^\omega$, where V^ω is considered to be an outcome (i.e., a value for some fixed ω) of a random vector $V(\omega)$ with zero mean $E\{V\} = 0$, and known covariance matrix $E\{VV^T\} = S$. Accordingly, L^ω is the outcome of a random vector $L(\omega) = AX + V(\omega)$, with unknown mean $E\{L\} = AX + E\{V\} = AX$, but known covariance matrix

$$E\{(L - AX)(L - AX)^T\} = E\{VV^T\} = S \qquad (3.39)$$

The deterministic model related to outcomes of the relevant random variables is

$$L^\omega = AX + V^\omega \qquad (3.40)$$

where L^ω, A are known, and X, V^ω are unknown. This is the usual observation equation model, and minimization of $(V^\omega)^T P V^\omega$ leads to an estimate $\hat{X}_P$ of X

$$\hat{X}_P = (A^T P A)^{-1} A^T P L^\omega \qquad (3.41)$$

where the subscript P of $\hat{X}_P$ denotes the dependence of the estimate on the weight matrix P used. By letting ω vary we obtain

$$\hat{X}_P(\omega) = (A^T P A)^{-1} A^T P L(\omega) \tag{3.42}$$

so that $\hat{X}_P = \hat{X}_P^{\omega}$ is the outcome of the corresponding random vector $\hat{X}_P(\omega)$. $\hat{X}_P(\omega)$ has unknown mean

$$E\{\hat{X}_P(\omega)\} = (A^T P A)^{-1} A^T P E\{L(\omega)\} =$$

$$= (A^T P A)^{-1} A^T P A X = X \tag{3.43}$$

but known covariance matrix

$$Q_P = \left[(A^T P A)^{-1} A^T P\right] S \left[(A^T P A)^{-1} A^T P\right]^T =$$

$$= (A^T P A)^{-1} A^T P S P A (A^T P A)^{-1} \tag{3.44}$$

We say that $\hat{X}_P^{\omega}$ is an unbiased estimate of X to denote that $\hat{X}_P^{\omega}$ is the outcome of a random variable $\hat{X}_P(\omega)$ whose expected value equals the true value of X.

A particular choice of weight matrix $P = S^{-1}$ leads to an estimate $\hat{X}_S$ of X, which is the outcome ($\hat{X}_S = \hat{X}_S^{\omega}$) of a corresponding random vector $\hat{X}_S(\omega)$ with

$$E\{\hat{X}_S(\omega)\} = X \tag{3.45}$$

and covariance matrix

$$Q_S = (A^T S^{-1} A)^{-1} A^T S^{-1} S S^{-1} A (A^T S^{-1} A)^{-1} = (A^T S^{-1} A)^{-1} \tag{3.46}$$

It can be shown that

$$Q_S \leq Q_P \qquad \text{for any qualifying P}$$

in the sense that for two square matrices A and B, $A \geq B$ if the matrix A - B is nonnegative definite. See, for example, [Deutsch, 1965, p. 62].

If now q is a scalar parameter linearly related to X

$$q = \sum_i d_i X_i = d^T X \tag{3.47}$$

we have the following estimates of q and their corresponding variances:

$$\begin{aligned} \hat{q}_P &= d^T \hat{X}_P & \sigma_P^2 &= d^T Q_P d \\ \hat{q}_S &= d^T \hat{X}_S & \sigma_S^2 &= d^T Q_S d \end{aligned} \tag{3.48}$$

Since $Q_P - Q_S$ is nonnegative definite, we have that:

$$d^T (Q_P - Q_S) d \geq 0$$

so that

$$d^T Q_P d \geq d^T Q_S d \quad \text{and} \quad \sigma_P^2 \geq \sigma_S^2 \tag{3.49}$$

This means that the choice of weight matrix $P^{-1} = S = E\{VV^T\}$ leads to a minimum variance estimate for any scalar linear function of the parameter vector X.

Another statistical property shared by any least squares estimate of a linear function of X is unbiasedness

$$E\{ \hat{X}_P(\omega) \} = X \tag{3.50}$$

$$E\{ \hat{q}_P(\omega) \} = E\{ d^T \hat{X}_P(\omega) \} = d^T E\{ \hat{X}_P(\omega) \} = d^T X = q \tag{3.51}$$

Condition equations. The deterministic model

$$B V^{\omega} = W^{\omega} \tag{3.52}$$

(B, W^{ω} known, V^{ω} unknown) refers to outcomes of corresponding random vectors $V(\omega)$, $W(\omega)$. $V(\omega)$ has zero mean $E\{V\} = 0$ and known covariance matrix $E\{VV^T\} = S$. It follows that $W(\omega)$ has zero mean $E\{W\} = BE\{V\} = 0$ and known covariance matrix $E\{WW^T\} = BSB^T$. For any choice of weight matrix $P = D^{-1}$, the least squares estimate of V^{ω} is

$$\hat{V} = \hat{V}^{\omega} = D B^T (B D B^T)^{-1} W^{\omega} \tag{3.53}$$

Letting ω vary, we obtain the related random vector

$$\hat{V}(\omega) = D B^T (B D B^T)^{-1} W(\omega) \tag{3.54}$$

with $E\{\hat{V}(\omega)\} = 0$ and covariance matrix

$$E\{\hat{V}(\omega)\hat{V}^T(\omega)\} = D B^T (B D B^T)^{-1} B S B^T (B D B^T)^{-1} B D \tag{3.55}$$

The vector W^{ω} is related to observations L^{ω} through $W^{\omega} = BL^{\omega}$ and L^{ω} corresponds to unknown observables L^* by means of $L^{\omega} = L^* + V^{\omega}$ where $BL^* = 0$. An estimate $\hat{L}^*$ of L^* can be obtained by means of

$$\hat{L}^* = \hat{L}^{*\omega} = L^{\omega} - \hat{V}^{\omega} \tag{3.56}$$

The error of the estimate $\hat{L}^*$ is defined to be

$$e = \hat{L}^* - L^* = (L^{\omega} - V^{\omega}) - (L^{\omega} - \hat{V}^{\omega}) = \hat{V}^{\omega} - V^{\omega} = e^{\omega} \tag{3.57}$$

The corresponding random variable $e(\omega) = \hat{V}(\omega) - V(\omega)$ has zero mean, and its covariance matrix Q_e is easily found to be

$$Q_e = DB^T (BDB^T)^{-1} BSB^T (BDB^T)^{-1} BD - DB^T (BDB^T)^{-1} BS - SB^T (BDB^T)^{-1} BD + S \tag{3.58}$$

Using the matrix identity from [Liebelt, 1967, p. 30, equation (1-53)],

$$ACA^T - BA^T - AB^T = (A - BC^{-1})\, C\, (A - BC^{-1})^T - BC^{-1}B^T \tag{3.59}$$

with $A \rightarrow DB^T(BDB^T)^{-1}$, $C \rightarrow BSB^T$ and $B \rightarrow SB^T$, we obtain

$$Q_e = \left[DB^T(BDB^T)^{-1} - SB^T(BSB^T)^{-1}\right] (BSB^T)^{-1} \left[DB^T(BDB^T)^{-1} - SB^T(BSB^T)^{-1}\right]^T - SB^T(BSB^T)^{-1}BS + S \tag{3.60}$$

Following the reasoning of [Liebelt, 1967, Section 5-3, p. 139], Q_e can be minimized in the sense that $y^T Q_e\, y \geq y^T (\min Q_e) y$ for any $y \in R^n$, by a choice of P such that

$$DB^T(BDB^T)^{-1} - SB^T(BSB^T)^{-1} = 0 \tag{3.61}$$

i.e., by choosing $P = S^{-1}$ or $D = P^{-1} = S$. For this choice we have an estimate of V:

$$\hat{V} = \hat{V}^{\omega} = S B^T (B S B^T)^{-1} W^{\omega} \tag{3.62}$$

and the minimum covariance matrix of the prediction error $e(\omega)$ becomes

$$Q_e = S - S B^T (B S B^T)^{-1} B S \tag{3.63}$$

Again the choice $P = S^{-1}$ related to the probabilistic structure of V justifies the least squares solution as a minimum variance of prediction error solution.

Generalized least squares model. In this particular case the introduction of probabilistic reasoning oversimplifies the derivation of the corresponding solution. The deterministic model is

$$W^{\omega} = AX + BV^{\omega} \tag{3.64}$$

with the corresponding random vector $V(\omega)$ having zero mean $E\{V\} = 0$ and known covariance matrix $E\{VV^T\} = S$. By setting $e = BV$, we obtain a model identical to the case of observation equations

$$W^{\omega} = AX + e^{\omega} \tag{3.65}$$

where the corresponding random vector $e(\omega)$ has zero mean $E\{e\} = BE\{V\} = 0$ and known covariance matrix $E\{ee^T\} = BSB^T = M$. The minimum variance solution corresponds to the choice of weight matrix $P = M^{-1}$

$$\hat{X} = \hat{X}^{\omega} = (A^T M^{-1} A)^{-1} A^T M^{-1} W^{\omega} =$$

$$= [A^T (BSB^T)^{-1} A]^{-1} A^T (BSB^T)^{-1} W^{\omega} \tag{3.66}$$

This is exactly the solution of the original model

$$W^{\omega} = A X + B V^{\omega}$$

corresponding to the weight matrix choice $P = S^{-1}$.

3.2.5 Weighted Parameters and Stochastic Processes

One particular case of least squares is when some or all of the parameters are considered to be random quantities with known expectation and covariance matrix. In the case of the generalized model we have

$$\tilde{W} = \tilde{A}\,\tilde{X} + \tilde{B}\,\tilde{V} = [A \;\; G] \begin{bmatrix} X \\ \tilde{s} \end{bmatrix} + \tilde{B}\,\tilde{V} = A X + G\,\tilde{s} + \tilde{B}\,\tilde{V} \tag{3.67}$$

with $\quad E\{\tilde{V}\} = 0, \quad E\{\tilde{V}\,\tilde{V}^T\} = C_{VV}$

$$E\{\tilde{s}\} = \bar{s}, \quad E\{(s-\bar{s})(s-\bar{s})^T\} = C_{ss}$$

setting $\tilde{s} = \bar{s} + s$ $(E\{s\} = 0)$, we obtain

$$\tilde{W} - G\,\bar{s} = A\,X + [\,G\;\tilde{B}\,] \begin{bmatrix} s \\ \tilde{V} \end{bmatrix} \tag{3.68}$$

This can be written after some obvious changes in notation as

$$W = A\,X + B\,V \tag{3.69}$$

with

$$E\{V\} = \begin{bmatrix} E\{s\} \\ E\{\tilde{V}\} \end{bmatrix} = 0,$$

$$E\{V\,V^T\} = \begin{bmatrix} E\{s\,s^T\} & E\{s\,\tilde{V}^T\} \\ E\{\tilde{V}\,s^T\} & E\{\tilde{V}\,\tilde{V}^T\} \end{bmatrix} = \begin{bmatrix} C_{ss} & C_{sV} \\ C_{Vs} & C_{VV} \end{bmatrix} = S$$

This is exactly a simple generalized least squares model with minimum variance solution

$$\hat{X} = (A^T M^{-1} A)^{-1} A^T M^{-1} W, \qquad \text{where} \quad M = B\,S\,B^T \tag{3.70}$$

and corresponding covariance matrix

$$C_{XX} = (A^T M^{-1} A)^{-1} \tag{3.71}$$

Usually the additional assumption $C_{sV} = 0$ holds, so that

$$M = [\, G \quad \tilde{B} \,] \begin{bmatrix} C_{ss} & 0 \\ 0 & C_{vv} \end{bmatrix} \begin{bmatrix} G^T \\ \tilde{B}^T \end{bmatrix} = G\, C_{ss}\, G^T + \tilde{B}\, C_{vv}\, \tilde{B}^T \tag{3.72}$$

and

$$\hat{X} = \left[A^T (G\, C_{ss}\, G^T + \tilde{B}\, C_{vv}\, \tilde{B}^T)^{-1}\, A \right]^{-1} A^T (G\, C_{ss}\, G^T + \tilde{B}\, C_{vv}\, \tilde{B}^T)^{-1}\, W \tag{3.73}$$

In the particular case that $B = G = I$, and with some changes of notation $s \to s'$, $W \to x$, $V \to n$, $C_{ss} \to C_{s's'} = C$, $C_{vv} \to C_{nn} = D$, $\bar{C} = C + D$, we obtain

$$\hat{X} = (A^T \bar{C}^{-1} A)^{-1}\, A^T \bar{C}^{-1}\, x \tag{3.74}$$

one of the results of Moritz [1972, p. 15, equation (2-35)], under an approach which he calls "least squares collocation," in connection with the model

$$x = A X + s' + n \tag{3.75}$$

Equation (2-36) in Moritz [1972] can be similarly derived applying results from classical least squares (minimum variance) methods. As to the third of the results (his equation (2-38)), we will have more to say in Section 3.3.

The case where all parameters are random reduces to the method of condition equations

$$\tilde{W} = A\, \tilde{s} + \tilde{B}\, \tilde{V}, \quad E\{\tilde{s}\} = \bar{s}, \quad E\{\tilde{V}\} = 0, \quad s = \tilde{s} - \bar{s},$$

$$E\{s\} = 0, \quad E\{s s^T\} = C_{ss}, \quad E\{\tilde{V}\tilde{V}^T\} = C_{vv} \tag{3.76}$$

We rewrite the model as

$$\tilde{W} - A\,\bar{s} = A\, s + \tilde{B}\, \tilde{V} = [\, A \quad \tilde{B} \,] \begin{bmatrix} s \\ \tilde{V} \end{bmatrix} \tag{3.77}$$

or after some obvious change of notation

$$B \quad V \;=\; W, \qquad \text{with} \qquad E\{V\} = \begin{bmatrix} E\{s\} \\ E\{\tilde{V}\} \end{bmatrix} = 0, \tag{3.78}$$

$$E\{V V^T\} = \begin{bmatrix} E\{s\,s^T\} & E\{s\,\tilde{V}^T\} \\ E\{\tilde{V}\,s^T\} & E\{\tilde{V}\,\tilde{V}^T\} \end{bmatrix} = \begin{bmatrix} C_{ss} & C_{sv} \\ C_{vs} & C_{vv} \end{bmatrix} = S$$

The solution is

$$V = \begin{bmatrix} s \\ \tilde{V} \end{bmatrix} = S\,B^T\,(B\,S\,B^T)^{-1}\,W \tag{3.79}$$

In the particular case that $C_{sV} = 0$, we obtain

$$s = C_{ss}\,A^T\,(A\,C_{ss}\,A^T + \tilde{B}\,C_{VV}\,\tilde{B}^T)^{-1}\,W \tag{3.80}$$

$$\tilde{V} = C_{VV}\,\tilde{B}^T\,(A\,C_{ss}\,A^T + \tilde{B}\,C_{VV}\,\tilde{B}^T)^{-1}\,W \tag{3.81}$$

Of particular interest is the simple case when A = B = I (observing a set of random quantities s under additive white noise n ($V \rightarrow n$)),

$$W = S + n$$

with solution

$$s = C_{ss}\,(C_{ss} + C_{nn})^{-1}\,W \tag{3.82}$$

We shall return to this equation in Section 3.3, in a somewhat different context.

Up to now we have discussed random parameters without reference to how they might arise in an actual real life situation. As we have seen, unknown random parameters $\tilde{s}$ with known expectation $\bar{s}$ and covariance matrix can be reduced to random parameters s with zero expectation and known covariance matrix.

In a least squares context (with minimum variance justification of the metric used), the distinction between zero mean random parameters s and zero mean random observational errors V is strictly verbal!

The solution algorithms we derived are exactly based in this lack of differentiation between the two. However, the distinction is very real in the process of modeling a real life situation, i.e., in the process of arriving through scientific reasoning to a least squares probabilistic model. The reasoning behind modeling observational errors as random variables with zero mean is too familiar to be repeated here, but the case of random parameters needs some discussion. A linear least squares model arises when a set of observations corresponds to a set of observables known (modeled) to be in a linear (or properly linearized) functional relation to a set of unknown parameters. Some (or even all) of these parameters may be connected to some unknown underlying function (or a number of such functions) with a certain domain T.

If the unknown function is modeled as a second-order stochastic process (usually called signal) with known mean value function and (auto)-covariance function, we may reason as follows: The underlying stochastic process is viewed as a mapping $\xi: \Omega \to \Phi$, where $\{\Omega, \mathcal{A}, P\}$ is a probability space and Φ is the space of the sample functions of the process $\xi(t, \omega)$. The connection of an unknown parameter $\tilde{s}_i$ to the underlying function is mathematically provided by the concept of a functional with domain Φ, i.e., by a mapping $l_i^*: \Phi \to R$. If the composite mapping $l_i^* \circ \xi : \Omega \to R$ is measurable with respect to the probability space $\{\Omega, \mathcal{A}, P\}$, then $\tilde{s}_i = l_i^*(\xi(t, \omega))$ is a random variable. The whole set of such random variables forms a vector $\tilde{s}$; and if the mean and covariance matrix of $\tilde{s}$ can be induced from the known mean value function and covariance function of the stochastic process $\xi(t, \omega)$, we have arrived at the models we considered at the beginning of the section.

Of particular interest is the case when the space of sample functions Φ is a Hilbert space and the functionals l_i^* belong to the dual space Φ^*, i.e.,

the set of linear bounded (continuous) functional on Φ. We will have a chance to return to this case later on for a more rigorous treatment of the problem.

3.3 Linear Least Squares Prediction

In linear least squares adjustment (observation equations), the original estimation problem is mathematically modeled as linear best approximation problem. A known element of a space is best approximated from the elements of a known subspace. However, the known element to be approximated does not appear explicitly in the solution (normal equations). Instead, only its inner product with the elements spanning the subspace of approximation need to be known.

We have already exploited this fact in the case of condition equations where we have obtained the solution by best approximating an unknown element V from the elements of a subspace N. The additional condition that V should belong to a certain hyperplane H provided us with the knowledge of the values of the inner product between V and any of the spanning elements in N.

Based upon this key observation, we shall now examine an estimation technique where an unknown element is best approximated from the elements of a known subspace, and the necessary information for the solution is contained in the a priori knowledge of the inner product values between the unknown element and the spanning elements of the subspace.

In the case of least squares adjustment techniques we started with a deterministic solution based upon an arbitrary inner product (arbitrary weight matrix), and finally the particular choice of a statistically meaningful metric (choice of $P^{-1} = E\{VV^T\}$) led to statistically meaningful results (minimum variance solution). Here we shall follow an exactly reversed path. We shall start with a statistically meaningful metric (inner product) and a minimum variance solution to finally obtain a deterministic solution with a corresponding weight matrix.

Because of this duality (deterministic-probabilistic prediction), we have chosen the title "linear least squares prediction," instead of the usual "minimum variance or minimum mean square error prediction," also following the terminology in [Doob, 1953, Chapter 12] and in [Cramer and Leadbetter, 1967, Section 5.7].

3.3.1 Probabilistic Approach (Minimum Variance or Minimum Mean Square Error Prediction)

We shall first present the method in a somewhat abstract context, and then proceed to show its connection to real life estimation problems.

The proper environment for our method is a space with elements square integrable real valued random variables (i.e., random variables $\xi(\omega)$ with $E\{|\xi(\omega)|^2\} < \infty$). If the random variables are defined on an abstract probability space $\{\Omega, \mathcal{A}, P\}$, we shall denote this space of square integrable random variables by $\mathcal{L}^2(\Omega, \mathcal{A}, P)$. This space can be turned into a Hilbert space by introducing the inner product

$$\langle x(\omega), y(\omega) \rangle = E\{x(\omega)\, y(\omega)\} \tag{3.83}$$

$\mathcal{L}^2(\Omega, \mathcal{A}, P)$ can be shown to be a linear space and the correlation of two elements can be easily shown to satisfy the defining properties of inner product. It can also be shown that $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ is a complete space (see [McGarty, 1974, p. 136]) and consequently a Hilbert space.

Given a set of random variables $x_i(\omega) \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$, $i = 1, 2, \ldots, n$, we denote their span by $M \subset \mathcal{L}^2(\Omega, \mathcal{A}, P)$. For any other random variable $y(\omega) \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$, we want to find the best approximation $\hat{y}(\omega)$ to $y(\omega)$ from M. The answer is well known to be

$$\hat{y}(\omega) = \mathcal{P}_M(y(\omega))$$

Since $\hat{y}(\omega) \in M$, and if we further assume that the random variables $x_i(\omega)$ are linearly independent, there exists a unique set of constants a_i, such that

$$\hat{y}(\omega) = \sum_{i=1}^{n} a_i \; x_i(\omega) \qquad (3.84)$$

The constants a_i are determined from the normal equations

$$\begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \dots & \langle x_1, x_n \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \dots & \langle x_2, x_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_n, x_1 \rangle & \langle x_n, x_2 \rangle & \dots & \langle x_n, x_n \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \langle x_1, y \rangle \\ \langle x_2, y \rangle \\ \vdots \\ \langle x_n, y \rangle \end{bmatrix} \qquad (3.85)$$

In matrix notation

$$C_{xx} \; a = C_{xy} \qquad (C_{xx} \text{ is } n \times n, \; C_{xy} \text{ is } n \times 1) \qquad (3.86)$$

where $[C_{xx}]_{ij} = E\{ x_i(\omega) \; x_j(\omega) \}$

and $[C_{xy}]_i = E\{ x_i(\omega) \; y(\omega) \}$

Introducing the matrix notation

$$\bar{x} = [x_1, x_2, \dots, x_n]^T$$

we have

$$C_{xx} = E\{ \bar{x} \; \bar{x}^T \}, \quad C_{xy} = E\{ \bar{x} \; y \}, \quad a = C_{xx}^{-1} \; C_{xy}$$

$$\hat{y} = \sum a_i \; x_i = a^T \; \bar{x} \qquad (3.87)$$

and $$\hat{y} = C_{xy}^T \; C_{xx}^{-1} \; \bar{x} \qquad (3.88)$$

Let us now examine how such a problem may arise in practice. Suppose we observe the values of a function s(t) at certain points t_i, I = 1, 2, ..., n, and we want to predict its value at some other point t_p. The observations and the value to be predicted correspond to functionals on some function space containing s(t) called evaluation functionals and of the form

$$l_i^*(s) = s(t_i) \qquad i = 1, 2, \ldots, n$$

and (3.89)

$$l_p^*(s) = s(t_p)$$

If the unknown function s(t) is modeled as a second-order stochastic process $s(t, \omega)$ (signal), then the evaluation functionals on the sample function space define random variables

$$l_i^*(s^\omega) = s^{t_i}(\omega) \tag{3.90}$$

The observations x_i are outcomes of the random variables $s^{t_i}(\omega)$. The outcome corresponding to $s^{t_p}(\omega)$ remains unknown and is approximated by the outcome corresponding to the projection of $s^{t_p}(\omega)$ on M = span $\{s^{t_i}(\omega)\}$

$$\hat{s}_p = \sum_i a_i x_i = a^T \bar{x} \tag{3.91}$$

where $a = C_{ss}^{-1} C_{sp}$ and $\hat{s}_p = C_{sp}^T C_{ss}^{-1} \bar{x}$

The elements of the matrices involved are

$$[C_{ss}]_{ij} = E\{ s^{t_i}(\omega) s^{t_j}(\omega) \} = r(t_i, t_j) \tag{3.92}$$

$$[C_{sp}]_i = E\{ s^{t_i}(\omega) s^{t_p}(\omega) \} = r(t_i, t_p) \tag{3.93}$$

where r(t, s) is the correlation function of the stochastic process s(t, ω). If s(t, ω) has zero mean, then r(t, s) is its covariance function. From a given process with known mean value function, we can always construct one with zero mean (see Section 2.4), and we will, therefore, assume that the stochastic processes involved are zero mean, without loss of generality.

In the more general case, our model might involve not one but a number of second-order stochastic processes of known autocovariances and cross-covariances (when correlated). The observations and prediction(s) need not necessarily correspond to values of these processes at certain points t_1, t_p. The only conditions on their nature is that the corresponding functionals l^* on the sample function space Φ of the process induce mappings $l \circ \varphi: \Omega \to R$ ($\varphi: \Omega \to \Phi$), which are measurable with respect to the probability space $\{\Omega, \mathcal{A}, P\}$ (i.e., random variables) and also belong to $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ (i.e., they have finite variances).

The Hilbert space where the approximation of $s^{t_p}(\omega)$ from the subspace M takes place need not necessarily be $\mathcal{L}^2(\Omega, \mathcal{A}, P)$. It can be replaced by a Hilbert space $\mathcal{L}^2(s(t, \omega), t \in T)$ with the same inner product and defined as the completion of the space $\mathcal{L}(s(t, \omega), t \in T)$ of all random variables $u(\omega)$ which may be written in the form

$$u(\omega) = \sum_{i=1}^{k} c_i \, s^{t_i}(\omega), \qquad c_i \in R \tag{3.74}$$

$\mathcal{L}^2(s(t, \omega), t \in T)$ contains all elements in $\mathcal{L}(s(t, \omega), t \in T)$ and the limits of sequences of such elements. For more details, see [Parzen, 1959, p. 259].

Of particular interest is the case where one of the processes involved x(t, ω) has sample functions in some Hilbert space H, and any other process y(t, ω) has sample functions also in H, related to those of x(t, ω) through a bounded linear operator $L: H \to H$

$$y^{\omega} = L\{x^{\omega}\} \tag{3.95}$$

Any bounded linear functional f^* on H has a representer $f \in H$

$$f^*(x^\omega) = \langle x^\omega, f \rangle \tag{3.96}$$

If g^* is a functional acting on $y^\omega(t)$, with representer g, then

$$g^*(y^\omega) = \langle y^\omega, g \rangle = \langle Lx^\omega, g \rangle \tag{3.97}$$

If we introduce an operator L^* called the adjoint of L (see [Taylor, 1958, Section 4.9, p. 249]) and satisfying

$$\langle Lx, y \rangle = \langle x, L^* y \rangle \tag{3.98}$$

we obtain

$$g^*(y) = \langle Lx, g \rangle = \langle x, L^* g \rangle = \langle x, h \rangle = h^*(x) \tag{3.99}$$

where

$$h = L^* \{ g \}$$

We need, therefore, deal only with the process $x(t, \omega)$ and linear bounded functionals $f^* \in H^*$, the dual space of H.

3.3.2 Deterministic Approach (Collocation)

In this section we shall examine a deterministic prediction technique introduced by Krarup [1969], in relation to the prediction of quantities related to the gravity field of the earth, and usually referred to as (exact) collocation. We shall follow Tscherning [1973] with a little more emphasis on "geometry."

Strictly speaking, collocation is a technique for finding a solution to a differential equation with insufficient boundary data. The differential equation admits an infinity of solutions and sufficient boundary data determine a unique solution among all possible ones. Insufficient boundary data restrict the candidate solutions to those satisfying the boundary conditions. Collocation is then a technique for determining a solution which is the smoothest, in some certain sense, among all solutions satisfying the boundary conditions.

Here we shall approach the problem without reference to any differential equations. We shall simply assume that we observe the values of bounded linear functionals on some unknown function f belonging to a Hilbert space H, and we want to find an estimate $\hat{f}$ of f, such that

$$\|\hat{f}\| = \min_{f \in P} \|f\| \tag{3.100}$$

where P is the set of functions in H satisfying the conditions imposed by the observations.

Let the observations d_i, $i = 1, 2, \ldots, n$ correspond to bounded linear functionals l_i^* on H ($l_i^* \in H^*$) with representers $l_i \in H$

$$d_i = l_i^*(f) = \langle f, l_i \rangle \qquad i = 1, 2, \ldots, n \tag{3.101}$$

This set of equations can be directly identified as a restriction that f should belong to a hyperplane P in H, which can alternatively be described as a linear variety, with the help of the unique element $g_o = \mathcal{P}_M(f)$, for every $f \in P$ (see Section 2.7, p. 25):

$$P = \{ g;\ g = g_o + g',\ g' \in M^\perp \} \tag{3.102}$$

where

$$M = \text{span}\{ l_i,\ i = 1, 2, \ldots, n \}$$

The situation is now similar to the case of condition equations in linear least squares adjustment, and the element $f_o \in P$ with minimum norm is provided by

$$f_o = \mathcal{P}_M(g) \tag{3.103}$$

where g is any element in P (see Figure 3.3). Since $f_o \in M$, we have

$$f_o = \sum_{i=1}^{n} a_i\ l_i \tag{3.104}$$

and the coefficients a_i are provided by the normal equations:

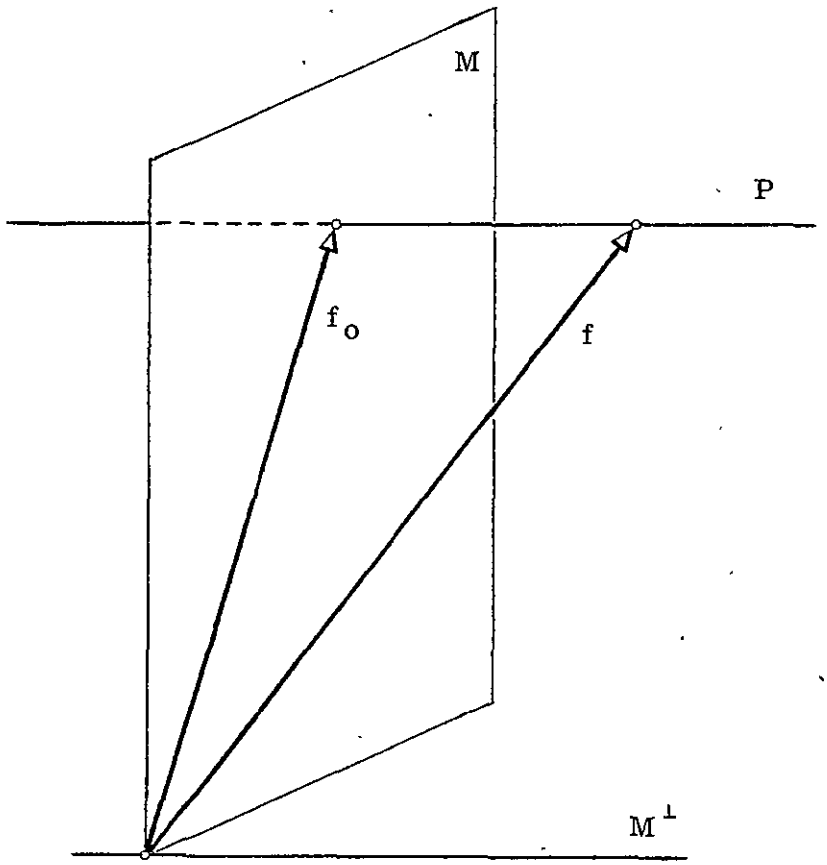

Figure 3.3 The Geometry of Collocation

$$\begin{bmatrix} \langle l_1, l_1 \rangle & \langle l_1, l_2 \rangle & \cdots & \langle l_1, l_n \rangle \\ \langle l_2, l_1 \rangle & \langle l_2, l_2 \rangle & \cdots & \langle l_2, l_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle l_n, l_1 \rangle & \langle l_n, l_2 \rangle & \cdots & \langle l_n, l_n \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \langle l_1, g \rangle \\ \langle l_2, g \rangle \\ \vdots \\ \langle l_n, g \rangle \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \tag{3.105}$$

or, in matrix notation,

$$C \; a = d, \qquad \text{with} \quad C_{ij} = \langle l_i, l_j \rangle \tag{3.106}$$

If the representers l_i are linearly independent, then C is nonsingular and

$$a = C^{-1} \; d \tag{3.107}$$

We have, therefore, estimated the unknown function f by its projection on the "data space," i.e., the space generated by the representers of the functionals corresponding to the available data.

We are next interested in estimating the value of some other possibly different bounded linear functional l_p^* on f. Its true value is

$$d_p = l_p^* (f) = \langle f, l_p \rangle \tag{3.108}$$

We can obtain an estimate using the estimate f_0 of f

$$\hat{d}_p = l_p^* (f_0) = \langle f_0, l_p \rangle \tag{3.109}$$

The error of prediction is

$$d_p - \hat{d}_p = \langle f, l_p \rangle - \langle f_0, l_p \rangle = \langle f - f_0, l_p \rangle \tag{3.110}$$

Since $f \in P$ and $f_0 \in P$, we have that $f - f_0 \in M^{\perp}$. If we decompose

$$l_P = l_{PM} + l_{PM^\perp} \tag{3.111}$$

where

$$l_{PM} = \mathcal{P}_M(l_P) \qquad \text{and} \qquad l_{PM^\perp} \in M^\perp$$

we obtain, taking into account that also $f_0 \in M$,

$$d_P - \hat{d}_P = \langle f - f_0 , l_{PM} + l_{PM^\perp} \rangle =$$

$$= \langle f - f_0, l_{PM} \rangle + \langle f, l_{PM^\perp} \rangle - \langle f_0, l_{PM^\perp} \rangle =$$

$$= \langle f, l_{PM^\perp} \rangle = l^*_{PM^\perp}(f)$$

i.e., only the part of l_P in $M^\perp$ contributes to the error. If $l_{PM^\perp} = 0$, i.e., $l_P \in M$, then $d_P - \hat{d}_P = 0$. Especially when $l_P = l_i$, then obviously $l_i \in M$, and

$$d_i - \hat{d}_i = \langle f - f_0 , l_i \rangle = 0 \tag{3.113}$$

We have, therefore, recovered the original observations, i.e., the approximation f_0 obtains the values observed. This justifies the name collocation as in the case of a differential equation solution.

An alternative way to look upon the estimate is the following

$$\hat{d}_P = \langle f_0, l_P \rangle = \langle f_0, l_{PM} + l_{PM^\perp} \rangle = \langle f_0, l_{PM} \rangle + \langle f_0, l_{PM^\perp} \rangle =$$

$$= \langle f_0, l_{PM} \rangle = \langle f_0 - f + f, \; l_{PM} \rangle =$$

$$= -\langle f - f_0, l_{PM} \rangle + \langle f, l_{PM} \rangle = \langle f, l_{PM} \rangle$$

Hence

$$\hat{d}_P = \langle f , l_{PM} \rangle = l^*_{PM}(f) \tag{3.114}$$

This means that the prediction corresponds to the true value of a functional l^*_{PM} with representer (see Figure 3.4)

$$l_{PM} = \mathcal{P}_M(l_P) \tag{3.115}$$

Since $f_0 = \sum_{i=1}^{n} a_i l_i$, we have for the prediction

$$\hat{d}_P = \langle f_0, l_P \rangle = \langle \sum_i a_i l_i , l_P \rangle = \sum_i a_i \langle l_i, l_P \rangle =$$

$$= [\, a_1 \; a_2 \; \dots \; a_n \,] \begin{bmatrix} \langle l_1, l_P \rangle \\ \vdots \\ \langle l_n, l_P \rangle \end{bmatrix} = a^T C_P = C_P^T a = C_P^T C^{-1} d \tag{3.116}$$

where

$$[\, C_P \,]_i = \langle l_i, l_P \rangle$$

To apply the method one must obviously know how to compute inner products between elements of the Hilbert space. Furthermore, the function f to be approximated is usually known to belong to a linear space which becomes a Hilbert space only after the introduction of an inner product. The choice of inner product is generally not unique and the same original linear space may give rise to different Hilbert space, different geometries, and, consequently, different choices of approximations f_0 to the original function f.

The problem of inner product choice remains open in a similar way as with the problem of choosing a weight matrix in the case of deterministic

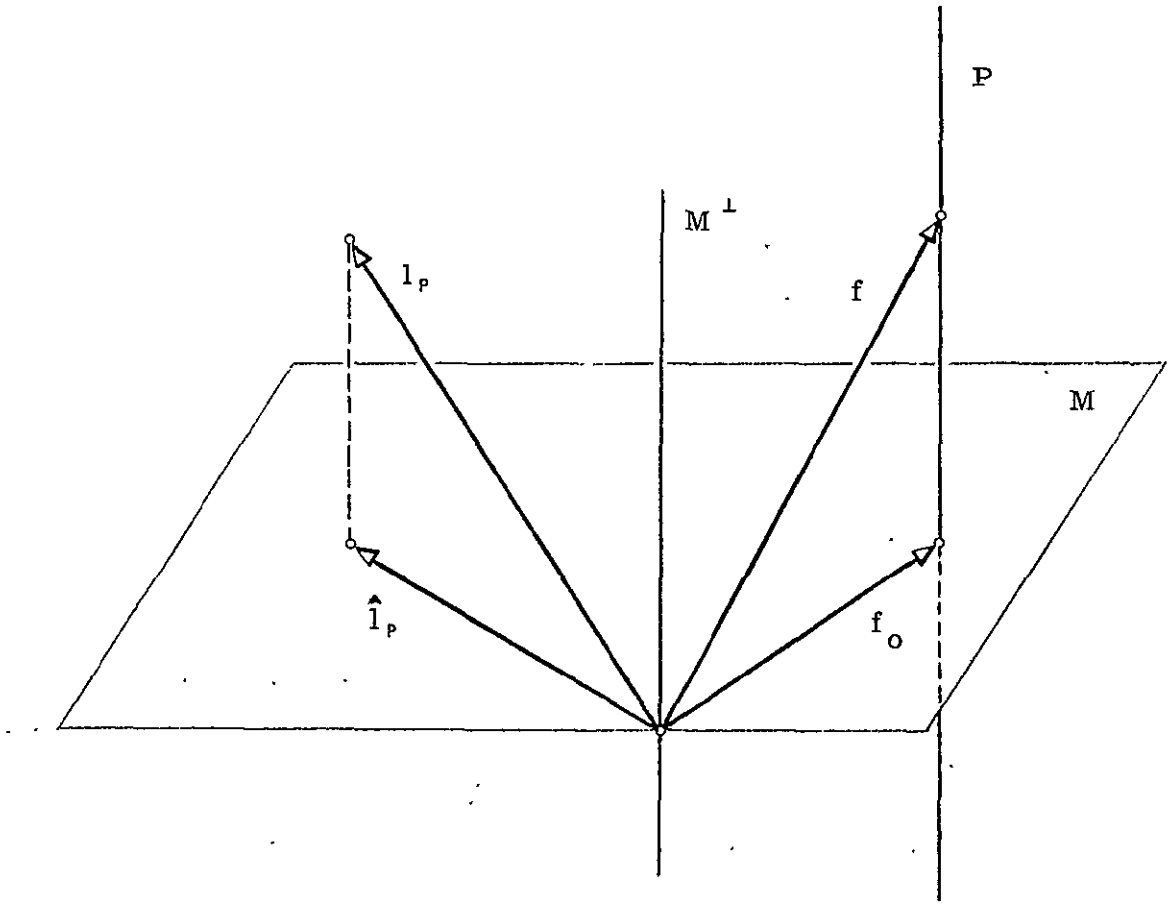

Figure 3.4 The Geometry of Collocation from an Alternative Point of View

linear least squares adjustment. Again the answer will be found by introducing probabilistic reasoning to justify metric (inner product) optimality. We shall consider this problem in the next section where a link between the probabilistic approach (minimum variance prediction) and the deterministic approach (collocation) will be established. Before we do this we need to introduce the concept of a reproducing kernel in a Hilbert space.

Let H be a Hilbert space of functions $f: T \to R$. A function $k: T \times T \to R$ is said to be a reproducing kernel $k(t, s)$ for H if

(a) $k^t(s) \in H$ for every $t \in T$ ($k^t(s) = k(t, s)$ denotes the mapping $k^t: T \to R$ for fixed t)

(b) $f(s) = \langle f, k^s \rangle$ for every $s \in T$ and every $f \in H$ (reproducing property).

It can easily be shown that $k(t, s)$ is symmetric, i.e., $k(t, s) = k(s, t)$. For a Hilbert space H to have a reproducing kernel, it is necessary and sufficient that the linear functionals (evaluation functionals)

$$l_t^*(f) = f(t) \qquad \text{for every } f \in H \text{ and } t \in T$$

are bounded, i.e., $l_t^* \in H^*$. (See [Aronszajn, 1950] and also [Lauritzen, 1973, Chapter 4]).

If $e_n(t)$, $n = 1, 2, \ldots$ is an orthonormal basis for H, then

$$k(t,s) = \sum_{n=1}^{\infty} e_n(t) \; e_n(s) \tag{3.117}$$

We shall next apply the concept of a reproducing kernel to two particular cases of collocation.

Case A: All functionals l_i^*, l_p^* involved are evaluation functionals

$$l_i^*(f) = f(t_i), \qquad l_p^*(f) = f(t_p) \tag{3.118}$$

The representer $l(s)$ of any functional $l^* \in H^*$ is related to the reproducing kernel by means of

$$l(s) = l^*(k^s) \tag{3.119}$$

For a proof, see [Shapiro, 1971, theorem 6.2.4.1, p. 85]. In particular, for evaluation functionals we have

$$l_i(s) = l_i^*(k^s) = k^s(t_i) = k^{t_i}(s) \tag{3.120}$$

It follows that, in view of the reproducing property and the symmetry of k(t, s)

$$\langle l_i, l_j \rangle = \langle k^{t_i}, k^{t_j} \rangle = k^{t_i}(t_j) = k(t_i, t_j) \tag{3.121}$$

and, similarly,

$$\langle l_i, l_p \rangle = k(t_i, t_p) \tag{3.122}$$

i.e., the elements of the matrices C, C_p are obtained by evaluating the reproducing kernel at the corresponding points

$$C_{ij} = k(t_i, t_j), \qquad [C_p]_i = k(t_i, t_p) \tag{3.123}$$

Case B: The functionals l_i^*, l_p^* involved are of the form

$$l_i^*(f) = (Lf)_{(t_i)}, \qquad l_p^*(f) = (Lf)_{(t_p)}$$

for some bounded linear operator L: H → H.

Let $\overline{l}_i^*$, $\overline{l}_p^*$ denote the evaluation functionals of the previous case. Then

$$l_i^*(f) = \langle Lf, \bar{l}_i \rangle = \langle f, L^* \bar{l}_i \rangle \qquad (3.124)$$

where we have made use of the definition of the adjoint operator L^* of L. It follows that l_i^* has representer

$$l_i = L^* \bar{l}_i = L^* \bar{l}_i^*(k^s) = L^* k^{t_i} \qquad (3.125)$$

Making use of the definition of the adjoint and the reproducing property, we obtain

$$\langle l_i, l_j \rangle = \langle L^* k^{t_i}, L^* k^{t_j} \rangle = \langle L L^* k^{t_i}, k^{t_j} \rangle =$$

$$= (L L^* k^{t_i})_{(t_j)} = (L L^* k)_{(t_i, t_j)} \qquad (3.126)$$

In a similar way

$$\langle l_i, l_p \rangle = (L L^* k)_{(t_i, t_p)} \qquad (3.127)$$

($L L^* k$ is a shorthand notation for $L_t [L_s^* k(t, s)]$, where subscripts of operators denote the variable on which the operators act.)

The elements of the matrices C and C_p are obtained by applying the operator on the reproducing kernel, according to a "law of propagation" of the reproducing kernel. The similarity to the law of propagation of covariance functions of stochastic processes under linear transformations provides a "hint" to a stronger connection to be revealed in the next section.

3.3.3 Relation Between Probabilistic Approach and Collocation

In the two previous sections, we examined two ways of modeling and solving the problem of prediction of quantities related to an unknown function from observations related to the same function. At a first glance the two

approaches seem radically different. In the probabilistic approach, the unknown function is modeled as a second-order stochastic process $s(t, \omega)$ and the prediction problem becomes an approximation problem in the Hilbert space $\mathcal{L}^2(s(t, \omega), t \in T) \subset \mathcal{L}^2(\Omega, \mathcal{A}, P)$. In the deterministic approach, the unknown function is modeled as an unknown element f(t) of a Hilbert space H of functions $x: T \to R$, and the prediction is dominated by the metric induced by the inner product in H. Knowledge of the structure of the unknown function (e.g., continuity, differentiability) usually gives rise to a number of different inner product choices and, consequently, different Hilbert spaces H where the prediction takes place.

The problem of inner product choice is equivalent to the problem of covariance function choice in the probabilistic approach. It will be shown here how a connection between a certain covariance function and a corresponding inner product may reveal the equivalence of the two approaches. Our main tools for this purpose are going to be the Moore-Aronszajn-Loeve theorem and the Karhunen-Loeve expansion of stochastic processes.

The Karhunen-Loeve expansion is usually given for intervals of the real line, i.e., for stochastic processes $\xi(t, \omega)$, $t \in T$ with $T = [a, b] \subset R$. A formal exposition can be found in [Papoulis, 1965, p. 457], and a more rigorous treatment with an exposition of the conditions $\xi(t, \omega)$ must satisfy for the expansion to hold can be found in [Gikhman and Skorokhod, 1969, p. 188]. We shall give here a more general exposition of the Karhunen-Loeve expansion, without restrictions on the specific nature of T, for the case where $\xi(t, \omega)$ has sample functions in a Hilbert space and can be viewed as a Hilbert-valued random variable. We shall follow [Lauritzen, 1973, Chapter 6] and [Rozanov, 1968, Chapter I, Section 3] with a minor shift of emphasis from the dual space H^* to the Hilbert space H itself, and without introducing the restruction of $\xi(t, \omega)$ being Gaussian.

A second-order stochastic process $\xi(t, \omega)$ ($\xi: T \times \Omega \to R$) with sample functions $\xi^{\omega}(t)$ in a Hilbert space H of functions $x: T \to R$ can be alternatively viewed as a Hilbert space-valued random variable $\xi: \Omega \to H$.

Assuming without loss of generality that $E\{\xi(t, \omega)\} = 0$, the covariance function of $\xi(t, \omega)$ is a mapping $r: T \times T \to R$, defined as $r(t, s) = E\{\xi^t \xi^s\}$. Now, if we view $\xi(t, \omega)$ as a Hilbert-valued random variable, we have

$$r(t,s) = E\{ l_t^*(\xi^\omega)\ l_s^*(\xi^\omega) \} \tag{3.128}$$

where $l_t^*, l_s^* \in H^*$ are evaluation functionals

$$l_t^*(x) = x(t) \qquad \text{for every } x \in H$$

Let $H_T^* \subset H^*$ denote the set of all evaluation functionals in H^* and define a mapping $R^*: H_T^* \times H_T^* \to R$ as

$$R^*(l_t^*, l_s^*) = r(t,s) \tag{3.129}$$

This mapping can be extended to the whole H^* ($R^*: H^* \times H^* \to R$) by setting

$$R^*(l_i^*, l_j^*) = E\{ l_i^*(\xi^\omega)\ l_j^*(\xi^\omega) \} \quad \text{for every } l_i^*, l_j^* \in H^* \tag{3.130}$$

Since each $l^* \in H^*$ has a unique representer $l \in H$, we can define a mapping $\tilde{R}: H \times H \to R$ as

$$\tilde{R}(l_i, l_j) = R^*(l_i^*, l_j^*) \text{ for every } l_i, l_j \in H \tag{3.131}$$

$\tilde{R}$ can be represented by an operator $\mathcal{R}: H \to H$ by means of

$$\tilde{R}(l_i, l_j) = \langle \mathcal{R} l_i, l_j \rangle_H \tag{3.132}$$

Such a bounded linear operator has the following properties:

(a) $\mathcal{R}$ is symmetric (self adjoint):

$$\langle \mathcal{R}x, y \rangle_H = E\{ x^*(\xi^\omega)\ y^*(\xi^\omega) \} = E\{ y^*(\xi^\omega)\ x^*(\xi^\omega) \} =$$

$$= \langle \mathcal{R} y, x \rangle_H = \langle x, \mathcal{R} y \rangle_H \tag{3.133}$$

(See [Taylor, 1958, p. 324].)

(b) $\mathcal{R}$ is positive (or positive semidefinite, or nonnegative), since it is symmetric and

$$< \mathcal{R}\, x, x >_H \;=\; E\left\{ \left[x^*(\xi^\omega) \right]^2 \right\} \;\geq\; 0 \qquad \text{for every } x \in H. \tag{3.134}$$

(See [Dunford and Schwartz, 1963, p. 906].)

(c) $\mathcal{R}$ is a Hilbert-Schmidt operator. For a definition of the term, see [Dunford and Schwartz, 1963, p. 1010]. A proof that $\mathcal{R}$ is Hilbert-Schmidt is given in [Grenander, 1963, p. 129], based upon corollary 3, p. 1011 of [Dunford and Schwartz, 1963].

(d) $\mathcal{R}$ is compact (completely continuous). For a definition of the term, see [Taylor, 1958, p. 274] or [Kolmogorov and Fomin, 1970, p. 239]. This property follows directly from the fact that $\mathcal{R}$ is Hilbert-Schmidt (see [Dunford and Schwartz, 1963, p. 1012, theorem 6]).

(e) $\mathcal{R}$ has finite trace, i.e., if $\{e_n\}$ is a complete orthonormal sequence in H [Grenander, 1963, p. 129]:

$$\sum_n < \mathcal{R} e_n, e_n >_H \;=\; \sum_n E\{ [< e_n, \xi^\omega >_H]^2 \} \;=\; E\{ \| \xi^\omega \|_H^2 \} \leq \infty \tag{3.135}$$

where ξ is restricted to the class of second-order Hilbert space-valued random variables satisfying the last inequality.

Such linear bounded operators which are symmetric, positive, and have finite trace are called S-operators in the relevant literature [Prokhorov, 1956, p. 172; Bharucha-Reid, 1972, p. 48; Grenander, 1963, p. 129; Parthasarathy, 1967, p. 154].

Since $\mathcal{R}$ is symmetric and compact, we have the following theorem based on corollary on p. 251 of [Kolmogorov and Fomin, 1970] (see also [Taylor, 1958, p. 336, theorem 6.4-B; and Bharucha-Reid, 1972, p. 48]):

Theorem: $\mathcal{R}$ has a sequence of eigenfunctions $\{e_n\}$ with corresponding eigenvalues $\{\lambda_n\}$, such that $\{e_n\}$ is a complete orthonormal sequence

in H, and

$$\mathcal{R}\, x = \sum_n \lambda_n \, \langle x, e_n \rangle_H \, e_n \qquad \text{for every} \quad x \in H \tag{3.136}$$

It follows from property (e) above that

$$\sum_n \langle \mathcal{R} e_n, e_n \rangle_H = \sum_n \langle \lambda_n e_n, e_n \rangle_H = \sum_n \lambda_n \tag{3.137}$$

and

$$\lambda_n \to 0$$

Since $\mathcal{R}$ is positive, we also have

$$\langle \mathcal{R} e_n, e_n \rangle_H = \langle \lambda_n e_n, e_n \rangle_H = \lambda_n \geq 0 \tag{3.138}$$

We shall refer to $\mathcal{R}$ from now on as the "covariance operator" of ξ. Strictly speaking, $\mathcal{R}$ as defined here is "correlation operator," and becomes a covariance operator under the assumption that $m = E\{\xi\} = \emptyset$, where the mean element $m \in H$ of ξ is defined by means of

$$x^*(m) = E\{ x^*(\xi^\omega) \} \qquad \text{for every} \quad x^* \in H^* \tag{3.139}$$

If $m \neq \emptyset$, we can introduce a covariance operator $\mathcal{C}$ by means of

$$\langle \mathcal{C}\, x, y \rangle_H = E\{ x^*(\xi^\omega - m) \; y^*(\xi^\omega - m) \}, \qquad x, y \in H \tag{3.140}$$

The representer l_t of the evaluation functional

$$l_t^*(x) = x(t), \qquad x \in H,$$

has a Fourier expansion in H

$$l_t = \sum_n e_n(t) \quad e_n \tag{3.141}$$

since for every $x \in H$ we have

$$l_t^*(x) = \langle x, l_t \rangle_H = \langle x, \sum_n e_n(t)\ e_n \rangle_H =$$

$$= \sum_n \langle x, e_n \rangle_H \quad e_n(t) = x(t) \tag{3.142}$$

Applying the above theorem (equation (3-136)), we obtain

$$\mathcal{R}\, l_t = \sum_n \lambda_n \langle l_t, e_n \rangle_H \quad e_n = \sum_n \lambda_n\, e_n(t)\ e_n \tag{3.143}$$

It follows that

$$r(t,s) = \langle \mathcal{R}\, l_t, l_s \rangle_H = \langle \sum_n \lambda_n\, e_n(t)\ e_n, \sum_m e_m(s)\ e_m \rangle_H$$

$$= \sum_n \sum_m \lambda_n\ e_n(t)\ e_m(s) \langle e_n, e_m \rangle_H$$

$$= \sum_n \sum_m \lambda_n\, e_n(t)\ e_m(s)\ \delta_{nm} = \sum_n \lambda_n\, e_n(t)\ e_n(s) \tag{3.144}$$

Since $\mathcal{R}$ is positive we have $\lambda_n \geq 0$, and setting $\lambda_n = \sigma_n^2$ we obtain

$$r(t,s) = \sum_n \sigma_n^2\ e_n(t)\ e_n(s) \tag{3.145}$$

For any fixed ω, $\xi^{\omega} \in H$ and has a Fourier expansion

$$\xi^{\omega} = \sum_{n} < \xi^{\omega}, e_n >_H e_n = \sum_{n} \xi_n(\omega) e_n \qquad (3.146)$$

The functional $d_n^*(\xi^{\omega}) = \xi_n(\omega)$ has representer $d_n = e_n$ since

$$< \xi, e_n >_H = \xi_n = d_n^*(\xi) \qquad (3.147)$$

and, therefore,

$$E\{ \xi_n(\omega) \xi_m(\omega) \} = < \mathcal{R} e_n, e_m >_H \qquad (3.148)$$

Since e_n is an eigenfunction of $\mathcal{R}$, we have

$$\mathcal{R} e_n = \lambda_n e_n \qquad (3.149)$$

and consequently

$$E\{ \xi_i \xi_j \} = < \lambda_i e_i, e_j >_H = \lambda_i < e_i, e_j >_H = \sigma_i^2 \delta_{ij} \qquad (3.150)$$

The expansion

$$\xi(t, \omega) = \sum_{n} \xi_n(\omega) e_n(t) \qquad (3.151)$$

with mutually orthogonal coefficients

$$E\{ \xi_i \xi_j \} = 0 \qquad \text{for } i \neq j \qquad (3.152)$$

and

$$E\{ \xi_n^2 \} = \lambda_n = \sigma_n^2 \qquad (3.153)$$

is called the Karhunen-Loeve expansion of $\xi(t, \omega)$.

We can now restrict ourselves to the study of positive definite covariance operators $\mathcal{R}$, such that

$$< \mathcal{R} x, x >_H > 0 \qquad \text{for every } x \in H \text{ and } x \neq \emptyset$$

Indeed, if $\mathcal{R}$ is simply nonnegative, the Hilbert space-valued random variable ξ can be written in the form

$$\xi = \xi_A + \xi_B \tag{3.154}$$

where

$$\xi_A = \sum_{n,\ \lambda_n > 0} \xi_n e_n \tag{3.155}$$

and

$$\xi_B = \sum_{n,\ \lambda_n = 0} \xi_n e_n \tag{3.156}$$

It can easily be shown that

$$E\{ x^*(\xi^\omega)\ y^*(\xi^\omega) \} = E\{ x^*(\xi_A^\omega)\ y^*(\xi_A^\omega) \} \tag{3.157}$$

This means that ξ and ξ_A are indistinguishable with respect to their second moments, relevant to prediction problems and that ξ can be replaced by ξ_A. We can also replace the original Hilbert space H with a new Hilbert space H', being the span of the eigenfunctions e_n of $\mathcal{R}$ corresponding to strictly positive eigenvalues $\lambda_n > 0$. The new covariance operator $\mathcal{R}'$ of ξ_A is the restriction of $\mathcal{R}$ in H' and is positive definite. With this in mind, we shall assume from now on that $\mathcal{R}$ is positive definite and $\lambda_n = \sigma_n^2 > 0$.

From the Moore-Aronszajn-Loeve theorem [Parzen, 1961, p. 965], we know that the covariance function r(t, s) generates a unique Hilbert space H(k) of which k(t, s) = r(t, s) is the reproducing kernel.

We shall next determine the structure (inner product) of H(k) and its relation to the Hilbert space H.

If ϵ_n is an orthonormal system in H(k) and since k(t, s) is the reproducing kernel, we must have

$$k(t,s) = \sum_n \sigma_n^2 \, e_n(t) \, e_n(s) = \sum_n \epsilon_n(t) \, \epsilon_n(s) \qquad (3.158)$$

We can therefore obtain an orthonormal system in H(k) by setting

$$\epsilon_n = \sigma_n e_n \qquad n = 1, 2, \ldots \qquad (3.159)$$

Let f, g $\in$ H have Fourier expansions

$$f = \sum_n \langle f, e_n \rangle_H \, e_n = \sum_n f_n \, e_n, \qquad g = \sum_n g_n \, e_n$$

so that

$$\langle f, g \rangle_H = \sum_n f_n \, g_n \qquad (3.160)$$

If, in addition, f, g $\in$ H(k), we have

$$\langle f, g \rangle_{H(k)} = \left\langle \sum_n f_n \, e_n, \sum_k g_k \, e_k \right\rangle_{H(k)} =$$

$$= \sum_n \sum_k f_n \, g_k \, \langle \sigma_n^{-1} \epsilon_n, \sigma_k^{-1} \epsilon_k \rangle_{H(k)} =$$

$$= \sum_n \sum_k \sigma_n^{-1} \sigma_k^{-1} \, f_n \, g_k \, \delta_{nk} = \sum_n \sigma_n^{-2} \, f_n \, g_n \qquad (3.161)$$

We have thus established a definition of inner product for elements of H(k) that also belong to H

$$< f, g >_{H(k)} = \sum_n \sigma_n^{-2} < f, e_n >_H < g, e_n >_H \qquad (3.162)$$

We will next show that minimum variance (probabilistic) prediction in the sense of Section 3.3.1 is equivalent to deterministic prediction in a Hilbert space in the sense of Section 3.3.2. In both cases the prediction is given as a linear combination of the vector of observations d, of the form

$$\hat{d}_P = C_P^T C^{-1} d \qquad (3.163)$$

The difference is that in the probabilistic approach

$$C_{ij} = E\{ l_i^*(\xi^\omega)\, l_j^*(\xi^\omega) \}, \quad [C_P]_i = E\{ l_i^*(\xi^\omega)\, l_P^*(\xi^\omega) \} \qquad (3.164)$$

where l_i^*, l_j^* are observation functionals and l_P^* the prediction functional, while in the deterministic approach

$$C_{ij} = < \bar{l}_i, \bar{l}_j >_{\bar{H}}, \qquad [C_P]_i = < \bar{l}_i, \bar{l}_P >_{\bar{H}} \qquad (3.165)$$

where $\bar{l}_i$, $\bar{l}_j$, $\bar{l}_P$ are the representers of the l_i^*, l_j^*, l_P^* functionals in some Hilbert space $\bar{H}$.

If $\bar{H}$ is taken to be the Hilbert space H(k) with reproducing kernel the covariance function r(t, s) of the stochastic process $\xi(t, \omega)$ involved in the probabilistic approach, it will be shown that

$$E\{ l_i^*(\xi^\omega)\, l_j^*(\xi^\omega) \} = < \bar{l}_i, \bar{l}_j >_{\bar{H}} \qquad (3.166)$$

and consequently the results of both prediction approaches are identical!

To establish the above equality, let $x^*, y^* \in H^*$ be two arbitrary functionals in H^* with representers x, y in H and x^k, y^k in H(k), defined as

$$x^*(f) = < f, x >_H = < f, x^k >_{H(k)} \qquad (3.167)$$

$$y^*(f) = < f, y >_H = < f, y^k >_{H(k)} \qquad (3.168)$$

for every $f \in H$ and also for $f \in H(k)$.

Let the Fourier expansions of f, x, y, x^k, y^k in H (assuming x^k, $y^k \in H$ also) be

$$f = \sum_n f_n \, e_n , \qquad x = \sum_n x_n \, e_n , \qquad y = \sum_n y_n \, e_n ,$$

$$x^k = \sum_n x_n^k \, e_n , \qquad y^k = \sum_n y_n^k \, e_n$$

Then, using the definition of inner product in H(k), we have

$$< f, x >_H = \sum_n f_n \, x_n = < f, x^k >_{H(k)} = \sum_n \sigma_n^{-2} \, f_n \, x_n^k \qquad (3.169)$$

and similarly

$$\sum_n f_n \, y_n = \sum_n \sigma_n^{-2} \, f_n \, y_n^k \qquad (3.170)$$

It follows that

$$x_n^k = \sigma_n^2 \, x_n , \qquad y_n^k = \sigma_n^2 \, y_n \qquad (3.171)$$

and, consequently,

$$< x^k, y^k >_{H(k)} = \sum_n \sigma_n^{-2} \, x_n^k \, y_n^k = \sum_n \sigma_n^{-2} \, (\sigma_n^2 \, x_n)(\sigma_n^2 \, y_n) =$$

$$= \sum_n \sigma_n^2 \, x_n \, y_n \qquad (3.172)$$

On the other hand,

$$E\{ x^*(\xi^\omega)\ y^*(\xi^\omega) \} = E\{ <\xi^\omega, x>_H\ <\xi^\omega, y>_H \} =$$

$$= E\{ (\sum_n x_n\ \xi_n)\ (\sum_k y_k\ \xi_k) \} =$$

$$= \sum_n \sum_k x_n\ y_k\ E\{ \xi_n\ \xi_k \} = \sum_n \sum_k x_n\ y_k\ \sigma_n^2\ \delta_{nk} =$$

$$= \sum_n \sigma_n^2\ x_n\ y_n \qquad (3.173)$$

The desired equality follows directly:

$$E\{ x^*(\xi)\ y^*(\xi) \} = <x^k, y^k>_{H(k)} = \sum_n \sigma_n^2\ x_n\ y_n \qquad (3.174)$$

Let l^2 denote the space of square summable sequences $f_1, f_2, \ldots,$ with

$$\sum_{n=1}^{\infty} f_n^2 < \infty \qquad (3.175)$$

Then l^2 is isometric to the original Hilbert space H, by the Riesz-Fischer theorem [Kolmogorov and Fomin, 1970, p. 153].

The space l_k^2 of sequences $f_1, f_2, \ldots$ with

$$\sum_{n=1}^{\infty} \sigma_n^{-2}\ f_n^2 < \infty \qquad (3.176)$$

is isometric to the Hilbert space H(k).

Let us formally consider sequences as infinite vectors

$$\bar{f} = [\, f_1, f_2, \ldots \,]^T, \qquad \bar{g} = [\, g_1, g_2, \ldots \,]^T \tag{3.177}$$

and also the infinite dimensional matrix $\bar{P}$ with $p_n = \sigma_n^{-2}$

$$\bar{P} = \begin{bmatrix} p_1 & 0 & & \\ 0 & p_2 & & 0 \\ & & \ddots & \\ & 0 & & \end{bmatrix} \tag{3.178}$$

We can now formally write the inner products in H and H(k) as

$$< f, g >_H = \bar{g}^T \bar{f}, \qquad < f, g >_{H(k)} = \bar{g}^T \bar{P} \bar{f}. \tag{3.179}$$

This establishes a sort of analogy with inner products in finite dimensional spaces appearing in linear least squares adjustment techniques. An arbitrary choice of $\bar{P}$ gives a "weighted least squares" deterministic solution. An optimal choice can be made when probabilistic reasoning is used and $\bar{P}$ is taken to be the inverse of an infinite dimensional matrix $\bar{S}$ with elements $\bar{S}_{ij} = \sigma_i^2 \delta_{ij}$. $\bar{S}$ can be identified as the covariance matrix of the infinite vector

$$\xi = [\xi_1, \xi_2, \ldots]^T$$

of the Fourier coefficients of the corresponding stochastic process $\xi(t, \omega)$

$$\bar{S} = E\{ \bar{\xi}\ \bar{\xi}^T \} \tag{3.180}$$

3.4 Minimum Error Bound Prediction

As shown in the previous section, collocation solutions become probabilistically meaningful when the inner product is linked to the known covariance function of the stochastic process serving as a model of the underlying unknown function. In case such a probabilistic reasoning is not justified, or even when we just have no knowledge of the covariance function, a prediction can still be carried out with a more or less arbitrary choice of inner product, corresponding to a "model covariance function." The resulting variance of the prediction error is only a "model variance" and should not be used in drawing statistical inferences on the prediction. When probabilistic reasoning does apply and the true covariance function is known, then the prediction error variance answers the question: "How good is the prediction?" In a strictly deterministic context, a similar answer must be found if the prediction is to be of any use at all. We shall try to answer this question here.

In a deterministic solution we cannot use variances (even though "model variances" can be computed), but a concept believed to be more legitimate is that of a bound of the prediction error. The error certainly remains unknown in general, but we might be able to ascertain that it does not exceed a certain bound in absolute value.

Let us first see what a "model variance" of the prediction error corresponds to. Probabilistic prediction using a model covariance function $k(t, s)$ corresponds to deterministic prediction (collocation) in a Hilbert space $H(k)$ with $k(t, s)$ as its reproducing kernel. The model variance of the prediction error $\epsilon_p = l_p^*(\xi^\omega) - \hat{l}_p^*(\xi^\omega)$ is given by

$$\sigma_p^2 = E\{ [l_p^*(\xi^\omega) - \hat{l}_p^*(\xi^\omega)]^2 \} \qquad (3.181)$$

where l_p^* is the functional corresponding to the prediction and $\hat{l}_p^*$ is its projection on the data space M^* (the span of the functionals corresponding to the observations).

As we have already established in the previous section

$$\sigma_p^2 = E\{[l_p^*(\xi^\omega) - \hat{l}_p^*(\xi^\omega)]^2\} = \langle l_p - \hat{l}_p, l_p - \hat{l}_p \rangle_{H(k)} = \| l_p - \hat{l}_p \|_k^2 \tag{3.182}$$

where l_p, $\hat{l}_p$ are the representers of l_p^*, $\hat{l}_p^*$ in H(k).

Suppose that in general we are given a set of observations corresponding to functionals l_i^* on an unknown deterministic element $\xi \in H(k)$, with representers l_i in H(k); and we want to predict the value of a functional l_p^* with representer l_p in H(k). Knowledge of the values of the observations

$$d_i = l_i^*(\xi) = \langle \xi, l_i \rangle_{H(k)} \tag{3.183}$$

allows us to find the projection ξ_0 of ξ on the data space $M = \text{span}(l_i,\ i = 1, 2, \ldots, n)$. If x^* is an arbitrary functional with representer x, we cannot in general compute $x^*(\xi)$ except when $x \in M$. In this case, using the decomposition

$$\xi = \xi_0 + \xi', \qquad \xi_0 \in M, \qquad \xi' \in M^\perp \tag{3.184}$$

we obtain

$$x^*(\xi') = \langle \xi', x \rangle_{H(k)} = 0 \tag{3.185}$$

since $\xi' \perp x$, and

$$x^*(\xi) = x^*(\xi_0 + \xi') = x^*(\xi_0) \tag{3.186}$$

We can compute $x^*(\xi_0)$, since ξ_0 is known.

This leads us to seek a functional x^* with representer x in M, so that we can compute

$$\hat{d}_p = x^*(\xi) \tag{3.187}$$

as an approximation to the true but unknown value

$$d_p = l_p^*(\xi) \tag{3.188}$$

For any $x \in M$, the prediction error is

$$\epsilon_p = d_p - \hat{d}_p = l_p^*(\xi) - x^*(\xi) = \langle l_p, \xi \rangle_{H(k)} - \langle x, \xi \rangle_{H(k)} =$$

$$= \langle l_p - x, \xi \rangle_{H(k)} \tag{3.189}$$

Using the Schwarz inequality [Davis, 1963, p. 159], we obtain

$$| \epsilon_p |^2 = | \langle l_p - x, \xi \rangle_{H(k)} |^2 \leq \| l_p - x \|_k^2 \quad \| \xi \|_k^2 \tag{3.190}$$

and

$$| \epsilon_p | \leq \| l_p - x \|_k \quad \| \xi \|_k \tag{3.191}$$

We have thus found a bound for the absolute value of the error of prediction. An obvious criterion of prediction optimality is to try to make this bound as small as possible. ξ is unknown but fixed and consequently $\| \xi \|_k$ is fixed. It remains to minimize $\| l_p - x \|_k$, i.e., to find an element $\hat{x} \in M$, which satisfies

$$\| l_p - \hat{x} \|_k = \min_{x \in M} \| l_p - x \|_k \tag{3.192}$$

The solution is well known to be

$$\hat{x} = \hat{l}_p = \mathcal{P}_M(l_p) \tag{3.193}$$

This choice differs only in motivation from the collocation approach where ξ is approximated by $\xi_0 = \mathcal{P}_M(\xi)$, and then the value of any functional $l_p^*(\xi)$ is approximated by $l_p^*(\xi_0)$. This follows from the fact already established in Section 3.3.2, i.e.,

$$l_p^*(\xi_0) = \langle l_p, \mathcal{P}_M(\xi) \rangle_{H(k)} = \langle \mathcal{P}_M(l_p), \xi \rangle_{H(k)} =$$

$$= \langle \xi, \hat{l}_p \rangle_{H(k)} = \hat{l}_p^*(\xi) \qquad (3.194)$$

This approach explains why and in what sense the approximation ξ_0 to ξ provides optimal estimates $l_p^*(\xi_0)$ of quantities $l_p^*(\xi)$. The optimality criterion is minimization of the error bound

$$\| l_p - \hat{l}_p \|_k \quad \| \xi \|_k$$

If in some way it is possible to obtain a bound for the norm of ξ

$$\| \xi \|_k \leq M_k \qquad (3.195)$$

and computing

$$\sigma_p^2 \doteq \| l_p - \hat{l}_p \|_k^2 \qquad (3.196)$$

we have a bound

$$| \epsilon_p | \leq m_k \doteq \sigma_p M_k \qquad (3.197)$$

When, therefore, a "model covariance function" is used, the computed "model variances" of prediction errors must be multiplied by the number M_k^2 to obtain the square of the bound for the absolute value of the prediction error.

Obviously M_k depends on the inner product in H(k) and consequently on the used model covariance k(t, s).

We shall call M_k the "covariance model error bound number." We shall later return to the problem of finding M_k in the case of predictions related to the gravity field of the earth.

3.5 Kalman-Bucy Filtering

A linear filter is a transformation L of functions x, y: $T \to R$, $T \subseteq R$, with two properties: linearity,

$$L\{\alpha x + \beta y\} = \alpha L\{x\} + \beta L\{y\} \tag{3.198}$$

and "time" invariance, if

$$\begin{aligned} y(t) &= (Lx)_{(t)} \text{ and} \\ x'(t) &= x(t+\tau) = (U_\tau x)_{(t)} \end{aligned} \tag{3.199}$$

then

$$y' = Lx' = LU_\tau x = U_\tau y = U_\tau Lx \tag{3.200}$$

and $LU_\tau = U_\tau L$, i.e., the filter commutes with the "shift" transformation U_τ.

We shall consider here linear filters having a time domain representation in the form of a convolution integral

$$y(t) = (Lx)_{(t)} = \int_{-\infty}^{+\infty} h(u)\ x(t-u)\ du \tag{3.201}$$

The filter is completely determined by the kernel h(u), called the impulse response function of the filter.

Linear filters may also be formally considered to transform stochastic processes $x(t, \omega)$, $y(t, \omega)$. A more rigorous discussion of some particular filters will be undertaken in Chapter 5. In the following discussion, we consider only stochastic processes; and we can therefore suppress the variable ω without danger of confusion.

Given an observed outcome of a stochastic process y(t) related to an unknown stochastic process x(t) to be estimated, the filtering problem is to find an impulse response function h(u), such that the output of the related filter with input y(t)

$$\hat{x}(t) = (L\,y)_{(t)} = \int_{-\infty}^{+\infty} h(u)\;\; y(t-u)\; du \qquad (3.202)$$

best approximates the signal x(t), in the sense of minimizing

$$E\,\{[x(t) - \hat{x}(t)]^2\} \qquad \text{for every} \quad t \in T. \qquad (3.203)$$

Since the observed function y(t) cannot be known for future values, we must restrict the admissible impulse response functions to those with $h(u) = 0$ for $u < 0$, for the related filter to be physically realizable.

For weakly stationary processes x(t), y(t), with

$$\begin{aligned} E\{\, y(t)\;\; y(t+\tau)\,\} &= \psi_{yy}(\tau) \\ E\{\, x(t)\;\; y(t+\tau)\,\} &= \psi_{xy}(\tau) \end{aligned} \qquad (3.204)$$

the answer to the minimization of $E\,\{[x(t) - \hat{x}(t)]^2\}$ is given by the Wiener-Hopf equation

$$\psi_{xy}(\tau) = \int_{0}^{\infty} h(s)\;\; \psi_{yy}(\tau - s)\; ds\;. \qquad (3.205)$$

The concept of a process y(t) continuously observed over all its past history is clearly a mathematical idealization. Consider instead a vector y of a finite number of observations

$$y_i = y(t_i) \qquad i = 1, 2, \ldots, n \qquad (3.206)$$

and the estimation of the value of x(t) at a certain epoch t_p, $\hat{x}_p = \hat{x}(t_p)$. The filter equation

$$\hat{x}_p = \int_{0}^{\infty} h(u)\;\; y(t_p - u)\; du \qquad (3.207)$$

should be replaced by a summation

$$\hat{x}_p = \sum_i h(t_p - t_i)\, y(t_i) = \sum_i h_i\, y_i = h^T y \qquad (3.208)$$

In a similar fashion, the Wiener-Hopf equation should be replaced by

$$\psi_{xy}(t_p - t_j) = \sum_i h(t_p - t_i)\ \psi_{yy}(t_i - t_j) \qquad (3.209)$$

or in matrix notation

$$C_{yy}\ h = C_{yx} \qquad (3.210)$$

where

$$\begin{aligned} [C_{yy}]_{ij} &= \psi_{yy}(t_j - t_i) = E\{ y(t_i)\ y(t_j) \} \\ [C_{yx}]_i &= \psi_{xy}(t_p - t_i) = E\{ y(t_i)\ x(t_p) \} \end{aligned} \qquad (3.211)$$

and finally

$$\hat{x}_p = C_{yx}^T\ C_{yy}^{-1}\ y \qquad (3.212)$$

This is exactly the formula for linear least squares (minimum variance) prediction. We can consider the filtering problem and the Wiener-Hopf equation as a generalization of minimum variance prediction for continuous observations.

Wiener [1949] and Kolmogorov solved in the early 1940's the problem of filtering a signal additively corrupted by noise, under the assumptions of stationarity, ergodicity, and knowledge of the entire past of the observed process. Wiener's results were expressed in the frequency domain and could not be directly extended to the nonstationary case. The work of Kalman and Bucy extended Wiener's results to the nonstationary case and

a process observed over a finite time interval only [Kalman, 1960; Kalman and Bucy, 1961].

Kalman and Bucy considered the vector process to be estimated to be the state x(t) of a dynamical system described by a linear differential equation

$$\frac{dx(t)}{dt} = F(t)\ x(t) + G(t)\ u(t), \qquad x(t_o) = x_o \tag{3.213}$$

and an observed process

$$z(t) = H(t)\ x(t) + v(t) \tag{3.214}$$

where u(t) and v(t) are white noise zero mean stochastic processes with covariance matrices

$$\begin{aligned} E\{\, u(t)\ u^T(s)\,\} &= Q(t)\ \delta(t-s) \\ E\{\, v(t)\ v^T(s)\,\} &= R(t)\ \delta(t-s) \\ E\{\, u(t)\ v^T(s)\,\} &= 0 \end{aligned} \tag{3.215}$$

where $\delta(t-s)$ is the Dirac delta function. x_o is a random vector with

$$E\{x_o\} = m_o, \quad E\{(x_o - m_o)(x_o - m_o)^T\} = Q_o, \quad E\{x_o v^T(s)\} = 0. \tag{3.216}$$

The solution of the filtering problem $\hat{x}(t)$ is the minimum variance estimate of x(t) based upon the observed process z(t) over the interval $[t_o, t]$ and is given by the solution of the following differential equation. (The most compact notation x_t, F_t, etc. is introduced in place of x(t), F(t), ...)

$$\frac{d\hat{x}_t}{dt} = F_t\ \hat{x}_t + K_t\,(z_t - H_t\ \hat{x}_t) \qquad \hat{x}(t_o) = m_o \tag{3.217}$$

where

$$K_t = P_t H_t^T R_t^{-1}, \qquad P_t = E\{(x_t - \hat{x}_t)(x_t - \hat{x}_t)^T\} \quad (3.218)$$

and the error covariance matrix P_t is the solution of

$$\frac{dP_t}{dt} = F_t P_t + P_t F_t^T - P_t H_t^T R_t^{-1} H_t P_t + G_t Q_t G_t^T, \quad P_o = Q_o. \quad (3.219)$$

The continuous observations case is hardly of any interest in geodetic problems where observations are discrete and finite in number. With further reference to [Bucy and Joseph, 1968; and Sage and Melsa, 1971] for more details, we turn our attention to the case of discrete observations,

$$y(t_i) = H(t_i)\, x(t_i) + v_i, \quad i=1,2,..,n; \qquad E\{v_i v_j^T\} = R_i\, \delta_{ij}. \quad (3.220)$$

A solution for minimum variance estimates $\hat{x}(t_i)$ of $x(t_i)$ can be obtained with the use of least squares adjustment techniques, if the means and covariances of the random variables $x(t_i)$ were known. Estimates $\hat{x}(t)$ for epochs other than the observation epochs t_i can be obtained from $\hat{x}(t_i)$ by least squares prediction techniques if the mean and the covariance function of the signal $x(t)$ were known. The required first- and second-order statistics of $x(t)$ can be obtained from those of $u(t)$ and x_o. The solution to the state differential equation

$$\frac{dx_t}{dt} = F_t x_t + G_t u_t \quad (3.221)$$

is of the form

$$x_t = \Phi(t, t_o)\, x_o + \int_{t_o}^{t} \Phi(t, \xi)\, G_\xi\, u_\xi\, d\xi. \quad (3.222)$$

The definition of the above integral raises some problems that we shall examine in Section 5.2. $\Phi(t, s)$ is the state transition matrix obtained by solving the differential equation

$$\frac{d}{dt} \Phi(t,t_o) = F(t) \Phi(t,t_o) , \qquad \Phi(t_o,t_o) = I \tag{3.223}$$

and using the transition property

$$\Phi(t,t_o) = \Phi(t,s) \Phi(s,t_o) \tag{3.224}$$

The required mean and covariance functions are

$$m_t = E\{x_t\} = \Phi(t,t_o) m_o \tag{3.225}$$

$$C_{xx}(t,s) = E\{(x_t - m_t)(x_s - m_s)^T\} =$$

$$= \Phi(t,t_o) Q_o \Phi^T(s,t_o) + \int_{t_o}^{\min(t,s)} \Phi(t,\xi) G_\xi Q_\xi G_\xi^T \Phi^T(s,\xi) d\xi \tag{3.226}$$

Suhc a global approach (in the sense of processing all observations together) shows the relation to least squares techniques but has great computational disadvantages due to the necessity of inverting large matrices. The alternative is a sequential solution where $\hat{x}(t_i \mid t_i)$, the minimum variance estimate of $x(t_i)$ based upon past observations $y(t_j)$ ($j = 1, 2, \ldots, i$), is obtained from the similarly defined estimate $\hat{x}(t_{i-1} \mid t_{i-1})$. It is possible to obtain a global solution for $\hat{x}(t_i \mid t_i)$ using observations up to epoch t_i only, and a similar global solution for $\hat{x}(t_{i-1} \mid t_{i-1})$, and then show the relation between the two. However, this involves an enormous algebraic effort, and it is much easier to derive directly the one-step solution for obtaining $\hat{x}(t_i \mid t_i)$ from $\hat{x}(t_{i-1} \mid t_{i-1})$.

First a discrete state model can be obtained

$$x_{i+1} = \Phi(i+1, i)\ x_i + w_{i+1} \tag{3.227}$$

by setting

$$x_i = x(t_i) \tag{3.228}$$

$$\Phi(i, j) = \Phi(t_i, t_j), \text{ and} \tag{3.229}$$

$$w_i = \int_{t_{i-1}}^{t_i} \Phi(t_i, \xi)\ G(\xi)\ u(\xi)\ d\xi \ . \tag{3.230}$$

w_i is a white sequence with $E\{w_i\} = 0$ and $E\{w_i\ w_j^T\} = \bar{Q}_i\ \delta_{ij}$ where

$$\bar{Q}_i = \int_{t_{i-1}}^{t_i} \Phi(t_i, \xi)\ G(\xi)\ Q(\xi)\ G^T(\xi)\ \Phi^T(t_i, \xi)\ d\xi \ . \tag{3.231}$$

A somewhat more general discrete state model is of the form

$$x_{i+1} = \Phi(i+1, i)\ x_i + \Gamma_i\ w_{i+1} \tag{3.232}$$

$\hat{x}(i+1 \mid i+1) = \hat{x}(t_{i+1} \mid t_{i+1})$ can be obtained from $\hat{x}(i \mid i)$, the covariance matrix $P(i \mid i)$ of the error in $\hat{x}(i \mid i)$, and the observation

$$y_{i+1} = H_{i+1}\ x_{i+1} + v_{i+1} \tag{3.233}$$

with the help of the following algorithm [Jazwinski, 1970, p. 270]:

$$\hat{x}(i+1|i) = \Phi(i+1,i)\ \hat{x}(i|i)$$

$$P(i+1|i) = \Phi(i+1,i)\ P(i|i)\ \Phi^T(i+1,i) + \Gamma_i\ Q_{i+1}\ \Gamma_i^T$$

$$K_{i+1} = P(i+1|i)\ H_{i+1}^T\ [\ H_{i+1}\ P(i|i)\ H_{i+1}^T + R_{i+1}\ ]^{-1}$$

$$\hat{x}(i+1|i+1) = \hat{x}(i+1|i) + K_{i+1}\ [\ y_{i+1} - H_{i+1}\ \hat{x}(i+1|i)\ ]$$

$$P(i+1|i+1) = [\ I - K_{i+1}\ H_{i+1}\ ]\ P(i+1|i) =$$

$$= [I - K_{i+1}\ H_{i+1}]\ P(i+1|i)\ [I - K_{i+1}\ H_{i+1}]^T + K_{i+1}\ R_{i+1}\ K_{i+1}^T$$

(3.234)

We shall next derive this algorithm with the use of the least squares adjustment (condition equations) technique.

Consider that an a priori unbiased estimate $\bar{x}_i$ of x_i is available, and C_i is the a priori known covariance matrix of the error $x_i - \bar{x}_i$. An a priori unbiased estimate $\bar{x}_{i+1}$ of x_{i+1} and the a priori covariance matrix C_{i+1} of the error $\delta x_{i+1} = x_{i+1} - \bar{x}_{i+1}$ can be obtained by simple propagation,

$$\bar{x}_{i+1} = \Phi(i+1,i)\ \bar{x}_i$$

$$C_{i+1} = \Phi(i+1,i)\ C_i\ \Phi^T(i+1,i) + \Gamma_i\ \bar{Q}_{i+1}\ \Gamma_i^T \qquad (3.235)$$

The observation equations can be rewritten in the form

$$l_{i+1} = y_{i+1} - H_{i+1}\ \bar{x}_{i+1} = H_{i+1}\ \delta x_{i+1} + v_{i+1} = [\ H_{i+1}\ I\ ] \begin{bmatrix} \delta x_{i+1} \\ v_{i+1} \end{bmatrix} \qquad (3.236)$$

where l_{i+1} is known and δx_{i+1} has zero mean and covariance C_{i+1}. Our model is of the condition equation type $BV = W$ (see Section 3.2.2) and the weight matrix of the zero mean vector

$$V = \begin{bmatrix} \delta x_{i+1} \\ v_{i+1} \end{bmatrix} \quad \text{is} \quad P = \begin{bmatrix} C_{i+1}^{-1} & 0 \\ 0 & R_{i+1}^{-1} \end{bmatrix} \tag{3.237}$$

The solution is of the form

$$\hat{V} = P^{-1} B^T M^{-1} W \tag{3.238}$$

with $M = BP^{-1}B^T$. Applying this to our case, we obtain

$$M = [\, H_{i+1} \;\; I \,] \begin{bmatrix} C_{i+1} & 0 \\ 0 & R_{i+1} \end{bmatrix} \begin{bmatrix} H_{i+1}^T \\ I \end{bmatrix} = H_{i+1} C_{i+1} H_{i+1}^T + R_{i+1} \tag{3.239}$$

$$\hat{V} = \begin{bmatrix} \delta \hat{x}_{i+1} \\ \hat{v}_{i+1} \end{bmatrix} = \begin{bmatrix} C_{i+1} & 0 \\ 0 & R_{i+1} \end{bmatrix} M^{-1} \, l_{i+1} \tag{3.240}$$

Setting $K_{i+1} = C_{i+1} H_{i+1}^T M^{-1}$, we obtain

$$\delta \hat{x}_{i+1} = K_{i+1} \, l_{i+1} = K_{i+1} \, (\, y_{i+1} - H_{i+1} \, \bar{x}_{i+1} \,) \tag{3.241}$$

$$\hat{x}(i+1, i+1) = \bar{x}_{i+1} + \delta \hat{x}_{i+1} = \bar{x}_{i+1} + K_{i+1} \, (y_{i+1} - H_{i+1} \, \bar{x}_{i+1}) =$$

$$= K_{i+1} \, y_{i+1} + (\, I - K_{i+1} \, H_{i+1} \,) \, \bar{x}_{i+1} \tag{3.242}$$

The covariance of y_{i+1} if R_{i+1}, and by simple propagation the covariance of the error in $\hat{x}(i+1 \mid i+1)$ becomes

$$P(i+1|i+1) = K_{i+1} R_{i+1} K_{i+1}^T + \\ + [I - K_{i+1} H_{i+1}] C_{i+1} [I - K_{i+1} H_{i+1}]^T \tag{3.243}$$

Identifying $\bar{x}_i$, $\bar{x}_{i+1}$, C_i, C_{i+1} with $\hat{x}(i \mid i)$, $\hat{x}(i+1 \mid i)$, $P(i \mid i)$ and $P(i+1 \mid i)$, respectively, the algorithm follows directly.

This solves the filtering problem of finding an estimate $\hat{x}(i \mid i)$ of x_i based upon past observations only. After $\hat{x}(n|n)$ is obtained, the estimates $\hat{x}(i|i)$ must be updated for the effect of future observations y_{i+1}, y_{i+2}, ..., y_n to obtain $\hat{x}(i|n)$. This is the smoothing problem, and finding $\hat{x}(t|n)$ for any other epoch t is the prediction problem. Refer to [Liebelt, 1967, Section 6-8] for smoothing algorithms.

Our main point here has been to show that for the case of discrete observations Kalman-Bucy filtering techniques are equivalent to the familiar least squares (minimum variance) adjustment methods.

4. DETERMINISTIC AND STOCHASTIC MODELS OF GEODETIC PROCESSES

4.1 Introductory Remarks

Physical processes related to geodetic work can be loosely divided into three categories. The first two correspond to the traditional objectives of geodetic research: Determination of the gravitational field and the shape of the earth.

The first category thus includes the gravity field of the earth and processes that result from transformations of the gravity field such as gravity anomalies, geoid undulations, etc.

The second category includes processes related to changes of the earth's geometric shape with time (station drifts, earth tides, etc.) as well as to changes of the earth's position in inertial space (precession-nutation, polar motion, variations in rotational velocity, etc.). We shall collectively call such processes "earth motion processes." By similarity one might include in this category processes related to the motion of other celestial bodies (lunar theory and librations, planetary motions, etc.) which might be involved in geodetic work.

The third category includes "noise processes," i.e., processes which although of no direct interest (to the geodesist at least) still appear in experiments directed towards the determination of geodetic parameters and processes. Having in mind current techniques for obtaining geodetic data, we might mention atmospheric refraction, nongravitational accelerations acting on artificial satellites, and, of course, observational noises associated with the observing instruments themselves.

The role that modeling of physical processes plays in the estimation of geodetic parameters is directly proportional to the accuracy of the available

observational techniques. In the presence of large observational inaccuracies, simple models, or even complete omission of the effect of the process, may be of no significant consequence on data analysis. In fact, advances in observational accuracy made possible the discovery or identification of processes whose existence had been preestablished by theory.

Recent advances in observational techniques necessitate the use of more sophisticated models if the objective of a "centimeter level geodesy" is to be realized.

Among models for processes which appear to be of critical importance in view of present day observational techniques, we shall outline here the two which we consider most important--the gravity field and the rotation of the earth.

4.2 The Gravity Field of the Earth

4.2.1 The Model

The gravity field of the earth is usually divided into two parts: a reference field (normal gravity potential) and the anomalous or disturbing potential. The disturbing potential T is known to belong to the class of functions harmonic outside the surface of the earth and regular at infinity (disregarding or including the effect of the atmosphere in the reference field).

It is also known from Runge's theorem (see [Krarup, 1969]) that T may be approximated arbitrarily well, in a certain sense, from members of the class of functions harmonic outside a sphere contained in the earth's interior (Bjerhammar sphere), and also regular at infinity. Such functions may be expanded into a series of spherical harmonics

$$f(P) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{nm} \, e_{nm}(P) \qquad (4.1)$$

where

$$e_{nm}(P) = \left(\frac{R}{r_P}\right)^{n+1} \sqrt{2(2n+1)\,\frac{(n-|m|)!}{(n+|m|)!}} \begin{cases} \cos|m|\lambda_P & m \geq 0 \\ \sin|m|\lambda_P & m < 0 \end{cases} P_{n|m|}(\cos\theta_P) \qquad (4.2)$$

where r_P, λ_P, θ_P are the spherical coordinates of the point P, R is the radius of the Bjerhammar sphere, and $P_{nm}(\cos\theta_P)$ are associated Legendre polynomials. In view of the relation

$$\frac{1}{4\pi R^2} \int_{S_R} e_{nm}(P)\; e_{kl}(P)\; d\sigma_P = \delta_{nk}\,\delta_{ml} \qquad (4.3)$$

(S_R denotes the surface of the sphere of radius R), we can consider a Hilbert space of potentials H, with inner product

$$\langle f, g \rangle = \frac{1}{4\pi R^2} \int_{S_R} f(P)\; g(P)\; d\sigma_P, \qquad f, g \in H \qquad (4.4)$$

and with e_{nm} (n = 0, 1, ...; m = -n, ..., -1, 0, 1, ..., n) as an orthonormal basis. (For a more rigorous discussion see [Lauritzen, 1973, Chapter 2].) Two functions with Fourier expansions

$$f(P) = \sum_{n,\,m} f_{nm}\, e_{nm}(P) \qquad \text{and} \qquad g(P) = \sum_{n,\,m} g_{nm}\, e_{nm}(P) \qquad (4.5)$$

have inner product

$$\langle f, g \rangle_H = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{nm}\, g_{nm} \qquad (4.6)$$

In view of Dirichlet's principle [Heiskanen and Moritz, 1967, p. 18], we can consider the restrictions of potentials on the surface of the Bjerhammar sphere, constituting a Hilbert space H with the above inner product.

The disturbing potential T(P) can now be modeled as a random field $T(P, \omega)$ with sample functions $T^{\omega}(P)$ in H or, alternatively, as a Hilbert-valued random variable with values in H. The Fourier coefficients T_{nm} of the expansion of T in H are random variables

$$T(P, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} T_{nm}(\omega)\ e_{nm}(P) \tag{4.7}$$

An additional restriction is that $T(P, \omega)$ should be a statistically homogenous (isotropic) random field, i.e., that for any set of points P_i, $i = 1, 2, \ldots, k$, the joint probability distribution of the random variables $T(P_i, \omega)$ is identical to that of the random variables $T(Q_i, \omega)$, where

$$\overline{r}_{Q_i} = M\,\overline{r}_{P_i} \tag{4.8}$$

$\overline{r}_{Q_i}$, $\overline{r}_{P_i}$ are position vectors of points Q_i, P_i, and M is a matrix representation of any rotation on the sphere ($MM^T = I$, $\det M = 1$). Obukhov [1947] has shown that for the random field $T(P, \omega)$ to be isotropic, we must have

$$E\{T_{nm}\, T_{kl}\} = \sigma_n^2\, \delta_{nk}\, \delta_{ml} \tag{4.9}$$

This implies that the admissible covariance functions for isotropy are those represented by a covariance operator $\mathscr{R}$: $H \to H$, having $e_{nm}(P)$ as its eigenfunctions and with corresponding eigenvalues σ_{nm}^2 independent of m

$$\mathscr{R} e_{nm}(P) = \sigma_n^2\ e_{nm}(P) \tag{4.10}$$

Any appropriate sequence of "coefficient variances" σ_n^2 (i.e., such that $T(P, \omega)$ is a second-order random field) defines a corresponding covariance function

$$r(P, Q) = \sum_n \sigma_n^2 \sum_m e_{nm}(P)\, e_{nm}(Q) = \sum_n \sigma_n^2\, (2n+1)\, P_n(\cos\psi_{PQ}) \tag{4.11}$$

$$r(P, Q) = \sum_n c_n P_n(\cos \psi_{PQ}) = r(\psi_{PQ}) \qquad (4.12)$$

We have used here the addition formula for spherical harmonics [Müller, 1966, p. 10], and we have introduced the "degree variances" c_n [Heiskanen and Moritz, 1967, p. 257]. ψ_{PQ} is the angular distance of the two points P and Q.

Lauritzen [1973] has proved that such a random field, under the restriction of Gaussianity, is not ergodic; and therefore its covariance function cannot be found through sampling by taking averages over the sphere. In his words [p. 80]:

> .... This means that, even if we knew gravity all over the earth, we would not be able to find the true value of the covariance function. ... Somehow the problem is not suited for statistical treatment. ...

However, one can still use a "model covariance" r(P, Q) and interpret the algorithm as deterministic prediction (collocation) in a Hilbert space H(k) with r(P, Q) as its reproducing kernel. We have already shown how such an approach can be motivated from a "minimum error bound" point of view. The application of this concept to prediction problems related to the gravity field of the earth leads to criteria for the optimality of the model covariance function and thus opens the way for the solution of the very important problem of the choice of norm (inner product) for the Hilbert space H(k) of potentials.

4.2.2 Application of Minimum Error Bound Prediction in Gravimetric Geodesy and the Optimum Norm Choice Problem

Suppose that we want to predict some quantity related to the earth's disturbing potential T from observations also related to T. We shall symbolically denote this problem by the triplet (D, p, k) where D stands for the set of observation functionals, p for the prediction functional, and k for the covariance-reproducing kernel of the Hilbert space H(k) in which the

prediction is taking place. The solution algorithm leads to a model variance σ^2 of the prediction error. We have found that the prediction error is bounded by

$$|\epsilon| \leq \sigma \|T\|_k \tag{4.13}$$

where $\|T\|_k$ is the norm of T in H(k)

$$\|T\|_k = \sum_{n,m} \sigma_n^{-2} T_{nm}^2 \tag{4.14}$$

T_{nm} are the coefficients of the Fourier expansion of T not in H(k) but in the Hilbert space H with orthonormal basis e_{nm}. σ_n^2 are the coefficients in the expansion of the reproducing kernel of H(k) (model covariance function)

$$r(P, Q) = k(P, Q) = \sum_{n,m} \sigma_n^2 e_{nm}(P) e_{nm}(Q) = \sum_n \sigma_n^2 (2n+1) P_n(\cos \psi_{PQ}) \tag{4.15}$$

If we could find a bound for each coefficient

$$|T_{nm}| \leq B_{nm} \tag{4.16}$$

then a bound M_k of $\|T\|_k$ can be found:

$$\|T\|_k^2 \leq M_k^2 = \sum_{n,m} \sigma_n^{-2} B_{nm}^2 \tag{4.17}$$

A proper place to look for bounds of the potential coefficients is the density function ρ of the mass of the earth which, in the first place, gives rise to the potential itself.

This has already been done by Cholshevnikov [1965 and 1968]. The bounds refer to the total harmonic potential of the earth:

$$V(Q) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} V_{nm} e_{nm}(Q) \tag{4.18}$$

In our notation these bounds are (see also [Payne, 1973]):

$$|V_{no}| \leq \frac{4\pi G\rho_{max}}{(n+1)\sqrt{2n+1}}\left(\frac{a_{max}}{R}\right)^{n+1} \tag{4.19}$$

$$\left\{\begin{matrix}|V_{n,m}|\\ |V_{n,-m}|\end{matrix}\right\} \leq \frac{8\pi G\rho_{max}}{(n+1)\sqrt{2n+1}}\left(\frac{a_{max}}{R}\right)^{n+1}\sqrt{\frac{(n-m)!}{(n+m)!}} \qquad m>0 \tag{4.20}$$

where ρ_{max} is the maximum density of the earth, G is the constant of gravitation, R is the radius of the Bjerhammar sphere, and a_{max} is the maximum distance of the earth surface from the geocenter. Cholshevnikov also gives bounds better than those above which, however, depend on the maximum variation of density with longitude (λ), $v_{\lambda,\,max}$, and with colatitude (θ), $v_{t,\,max}$ ($t=\cos\theta$) as follows:

$$|V_{no}| \leq \frac{4G(22)^{\frac{1}{4}}(\pi)^{\frac{1}{2}}\quad(2v_{t,\,max}+\rho_{max})}{(2n+1)(n+1)\quad\sqrt{(2n+1)(n-1)}}\left(\frac{a_{max}}{R}\right)^{n+1} \tag{4.21}$$

$$\left\{\begin{matrix}|V_{n,m}|\\ |V_{n,-m}|\end{matrix}\right\} \leq \frac{8Gv_{\lambda,\,max}}{\sqrt{2n+1}\;(n+1)\;m}\left(\frac{a_{max}}{R}\right)^{n+1}\sqrt{\frac{(n-m)!}{(n+m)!}} \qquad m>0 \tag{4.22}$$

Using the known coefficients of the harmonic part of the reference potential, we can compute bounds for every coefficient T_{nm}, find the bound M_k of the disturbing potential norm, and finally the bound of the prediction error

$$|\epsilon| \leq m_k = \sigma M_k \tag{4.23}$$

This bound is independent of the observations and neglects the information about T contained in the data

$$d_i = \langle T, l_i^k \rangle_{H(k)} \qquad i=1, 2, \ldots, n \tag{4.24}$$

To take this information into account we decompose T as

$$T = \mathcal{P}_M(T) + \mathcal{P}_{M^\perp}(T) = T_0 + T' \tag{4.25}$$

where M is the span of the observation functionals (data space) and $M^\perp$ its orthogonal complement in H(k).

The prediction error now is

$$\epsilon = \langle l_p^k - \hat{l}_p^k, T_0 + T' \rangle_{H(k)} = \langle l_p^k - \hat{l}_p^k, T' \rangle_{H(k)} \tag{4.26}$$

since $l_p^k - \hat{l}_p^k \perp T_0$; and

$$|\epsilon| = |\langle l_p^k - \hat{l}_p^k, T' \rangle_{H(k)}| \leq \| l_p^k - \hat{l}_p^k \|_k \, \| T' \|_k = \sigma \| T' \|_k \tag{4.27}$$

where l_p^k is the representer of the prediction in H(k), and $\hat{l}_p^k$ is its projection on the data space M. This is a better bound, since from the Pythagorean theorem

$$\| T \|_k^2 = \| T_0 \|_k^2 + \| T' \|_k^2 \quad \text{and} \quad \| T' \|_k \leq \| T \|_k \tag{4.28}$$

The new bound for the prediction error becomes

$$|\epsilon| \leq m_k = \sigma \sqrt{M_k^2 - \| T_0 \|_k^2} \tag{4.29}$$

The error bound, of course, depends on the choice of kernel in H(k), i.e., on the choice of a model covariance function.

—If k_1 and k_2 are two kernels giving rise to two error bounds m_{k_1}, m_{k_2}, we say that kernel k_1 is better than kernel k_2 for predicting a certain quantity from a certain data set if $m_{k_1} < m_{k_2}$.

We have thus found a tool for comparing kernels, and the definition of the optimal reproducing kernel (or optimal inner product or optimal model covariance function) follows directly:

—We say that a kernel k_0 is the best kernel for predicting a certain quantity from a certain set of data if the corresponding error bound m_0 satisfies

$$m_0(D, p, k_0) = \min_{k \in \mathcal{K}} m(D, p, k) \tag{4.30}$$

where m(D, p, k) denotes the dependence of the error bound on the observations D, the prediction p, and the kernel k. $\mathcal{K}$ denotes the class of all "permissible" kernels. The class $\mathcal{K}$ is difficult to define. An obvious necessary condition on members of $\mathcal{K}$ is that the corresponding bound of the disturbing potential norm is finite

$$M_k^2 = \sum_{n=0}^{\infty} \sigma_n^{-2} \sum_{m=-n}^{n} B_{nm} < \infty \tag{4.31}$$

It is more reasonable to look for the optimal kernel in a class of kernels convenient for computations. Such kernels must be given by closed expressions rather than by an infinite sequence σ_n^2. One could even try to find a closed expression including a finite number of parameters, such that the family of kernels corresponding to different sets of parameter values is "broad" enough in some sense. Then the optimal set of parameters could be found as the one giving the smaller prediction error bound. However, the dependence of the bound on the kernel is quite complex:

$$m(D, p, k) = \sigma(D, p, k) \{ M^2(k) - \| T_0 \|^2 (D, k) \}^{\frac{1}{2}} \tag{4.32}$$

where M^2 depends on the kernel, $\| T_0 \|^2$ on the kernel and the observations, and σ on kernel, observations, and prediction. More specifically

$$\sigma^2 = C_0 - C_p^T C^{-1} C_p, \qquad C_0 = < l_p^k, l_p^k >_{H(k)} \tag{4.33}$$

with

$$(C_p)_i = < l_i^k, l_p^k >_{H(k)}, \qquad C_{ij} = < l_i^k, l_j^k >_{H(k)} \tag{4.34}$$

where l_i^k, l_p^k are the representers of the observation and prediction functionals in H(k). We also have

$$T_0 = \sum_{i=1}^{n} a_i \, l_i^k \tag{4.35}$$

where the coefficients a_i are found from the solution of the normal equations (see Section 3.3.2)

$$a = C^{-1} d \tag{4.36}$$

It follows that

$$\| T_o \|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i j_i \langle l_i^k, l_j^k \rangle_{H(k)} = a^T C a = d^T C^{-1} d \tag{4.37}$$

The final equation for the prediction error bound is

$$m = \{ C_o - C_p^T C^{-1} C_p \}^{\frac{1}{2}} \left\{ \sum_{n,m} \sigma_n^{-2} B_{nm} - d^T C^{-1} d \right\}^{\frac{1}{2}} \tag{4.38}$$

Another point to be made is that one is usually interested not in one but in a number of predictions, and the idea of applying a different optimal kernel for each particular prediction is not appealing, in view of the computational effort involved. In this case, one might instead introduce a risk function, e.g., a nonnegative function $L(\overline{m})$ of the vector of prediction error bounds $\overline{m}$, with the following properties [Jazwinski, 1970, p. 147]: $L(0) = 0$, $L(u_1) \geq L(u_2) \geq 0$ for $|u_1| \geq |u_2|$, where $|u| = (u^T u)^{\frac{1}{2}}$. The optimal kernel may now be defined as the one minimizing the risk function rather than any of each individual bounds.

The objective of our discussion here has been not to give any final answers, but rather to provide guidelines and motivation for more work on the question of norm optimality in gravimetric collocation. An interesting and motivating discussion on this problem can be found in [Eeg and Krarup, 1975, especially Section 5].

Our approach is novel and, what is more important, completely independent of any probabilistic reasoning. The findings of Lauritzen (nonergodicity) stand in support of a purely deterministic approach such as ours and have actually motivated our work. Our criterion of prediction accuracy (error bound m) shares the same nice asymptotic behavior with probabilistic techniques: If the prediction functional approaches the data space M (in the sense that $\| l_p^k - \mathcal{P}_M (l_p^k) \|_k \to 0$), then the variance of the

prediction error tends to zero ($\sigma^2 \to 0$), and also $m \to 0$ since σ is one of the product terms in m.

In a probabilistic technique, the variance σ^2 is independent of the actual outcomes of the observations! It only depends on what we have observed and what we want to predict. In our approach m depends on the outcomes of the observations through $\| T_o \|$, the norm of the known component of the potential T (its projection on the data space). The closer T is to the data space, the larger $\| T_o \|$ becomes, and the smaller the bound m of the prediction error. The main defect of our approach is the neglect of the effect of observational noise. An extension is obviously needed, either in a deterministic sense introducing bounds for the observational errors too, or by means of some combination of deterministic and probabilistic concepts.

4.3 The Rotation of the Earth

4.3.1 Choice of Reference Frame and Parameterization of Earth Rotation

In this Section we are concerned with the modeling of the motion of the earth with respect to an inertial system within the framework of classical mechanics. Depending on the type of available observations, the inertial system is realized with the help of the dynamics of the solar system, the stars (taking care of proper motions), or extragalactic radio sources. The motion of the earth can be mathematically described by an infinite set of functions

$$X_i(t), \qquad Y_i(t), \qquad Z_i(t)$$

for all points i of the earth with inertial coordinates X_i, Y_i, Z_i. In reality only the motion of points on the surface of the earth is directly observable, and practical reasons confine us to a finite system of points i = 1, 2, ..., n (network stations). The vector valued function

$$\overline{X}(t) = [X_1\ Y_1\ Z_1\ X_2\ Y_2\ Z_2 \ \ldots\ X_n\ Y_n\ Z_n]^T(t) \qquad (4.39)$$

gives the state of the point network for each epoch t and provides a description of its motion.

Any arbitrary time dependent transformation matrix M(t) (i.e., such that $M(t)\,M^{T}(t) = I$ and $\det M(t) = 1$ for every t), together with a vector valued function $\bar{\delta}(t)$, defines a new moving reference frame x, y, z. The coordinates of the network points in the new system are given by

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}_{(t)} = M(t) \left\{ \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix}_{(t)} + \bar{\delta}(t) \right\} \tag{4.40}$$

or

$$\bar{r}_i(t) = M(t)\,\{\bar{R}_i(t) + \bar{\delta}(t)\} \tag{4.41}$$

We can group all possible such frames into classes through the following definition: Two given reference frames defined by

$$\begin{aligned} \bar{r}_1(t) &= M_1(t)\,\{\bar{R}(t) + \bar{\delta}_1(t)\} \\ \bar{r}_2(t) &= M_2(t)\,\{\bar{R}(t) + \bar{\delta}_2(t)\} \end{aligned} \tag{4.42}$$

are said to be equivalent or rigidly connected if there exists a constant transformation matrix M_{12} and a constant vector δ_{12} such that

$$\bar{r}_2(t) = M_{12}\,\{\bar{r}_1(t) + \bar{\delta}_{12}\} \tag{4.43}$$

If the earth, or just the network of points in question, were rigid, it would have been possible to find a class of equivalent reference frames such that the coordinates of the points with respect to any of them would be time independent. For a nonrigid network, a reasonable choice is a class of equivalent frames such that the relative motions of the points are minimized in some certain sense, for example,

$$\sum_i \left(\int_{t_1}^{t_2} \sqrt{\left(\frac{dx_i}{dt}\right)^2 + \left(\frac{dy_i}{dt}\right)^2 + \left(\frac{dz_i}{dt}\right)^2}\, dt \right) = \min \tag{4.44}$$

We can, furthermore, eliminate translatory motions by moving the origin of the inertial frame to the geocenter, taking into account the corresponding effects on the observations (e.g., secular and annual aberration, annual parallactic displacement, relativistic effects). We can also identify a particular frame from the optimal class with a frame of the same class with origin at the geocenter and axes parallel to the former. By identification here we simply mean that the rotation of both systems are described by the same set of parameters, and we can therefore take advantage of the simplifications in the equations of rotation when referred to a geocentric frame.

Alternative choices for a frame fixed, or rather attached, to a network of points on the earth can be based on physically meaningful geometric characteristics of the earth such as the geocenter and the principal axes of inertia. Such choices become relevant only when available observations are sensitive to such a physically appealing choice of frame as, for example, in the case of satellite observations in orbits governed by the gravity field of the earth. It is well known [Heiskanen and Moritz, 1967, p. 62] that the geocenter and principal axes of inertia are connected to the first- and second-degree spherical harmonic coefficients in the expansion of the attraction potential of the earth.

Even for strictly geometric observations involving both "earth" points and "inertial" points (stars, quasars, points on the moon, etc.), the directions of the principal axes of inertia appear implicitly in the equations of the motion of the earth. However, the little sensitivity of present observations to such natural geometric characteristics of the earth and the uncertainty present in the relevant equations of motion seem to justify the use of an "arbitrary" frame, especially since relative motions of network points can now be estimated to a comparatively high degree of accuracy.

The problem of connection between an arbitrary frame and a physically meaningful one can be treated separately when sufficient observations for this purpose are available.

Our choice of an arbitrary "earth-fixed" frame from a properly defined optimal class of equivalent frames (more precisely its geocentric parallel from the same class) coincides with the concept of "geographic axes," "attached in a prescribed way to the observatories," as discussed in [Munk and MacDonald, 1960, p. 11].

The rotational time dependent transformation between the arbitrary-geographic geocentric frame and the quasi-inertial geocentric frame can be described with the help of three parameters defining the transformation matrix M(t). Among possible choices, a traditional one is that of the Eulerian angles φ, θ, ψ [Goldstein, 1950, p. 107]

$$M(t) = R_3(\psi(t))\ R_1(\theta(t))\ R_3(\varphi(t)) \tag{4.45}$$

where R_1, R_3 are rotation matrices about the x and z axes respectively, and M(t) transforms inertial into earth-fixed vectors as follows

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}_{(t)} = M(t) \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix}_{(t)} \tag{4.46}$$

4.3.2 The Dynamics of the Rotation of the Earth, The Liouville Equation

The instantaneous rotation vector of the earth $\bar{\omega}$ with respect to the earth-fixed system is connected to the Eulerian angles through the geometric Euler's equations [MacMillan, 1960, p. 185]

$$\begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} = \begin{bmatrix} \sin\theta\ \sin\psi & \cos\psi & 0 \\ \sin\theta\ \cos\psi & -\sin\psi & 0 \\ \cos\theta & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{\varphi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} \tag{4.47}$$

or

$$\dot{\bar{\omega}} = \underline{S}_{(\bar{e})} \frac{d\bar{e}}{dt} = \underline{S}_{(\bar{e})} \dot{\bar{e}}, \qquad \bar{e} = [\varphi\ \theta\ \psi]^T \tag{4.48}$$

The rotational motion of the earth is governed by the Liouville equation [Munk and MacDonald, 1960, p. 9]

$$\frac{d}{dt}\left\{\underline{C}_{(t)}\,\overline{\omega}_{(t)} + \overline{h}_{(t)}\right\} + [\overline{\omega}\wedge]_{(t)}\left\{\underline{C}_{(t)}\,\overline{\omega}_{(t)} + \overline{h}_{(t)}\right\} = \overline{L}_{(t)} \qquad (4.49)$$

where

$$[\omega\wedge] = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \qquad (4.50)$$

and

$$\underline{C} = \begin{bmatrix} A & -F & -E \\ -F & B & -D \\ -E & -D & C \end{bmatrix} = \int_{\text{earth}} \begin{bmatrix} y^2+z^2 & -xy & -xz \\ -xy & x^2+z^2 & -yz \\ -xz & -yz & x^2+y^2 \end{bmatrix} dm \qquad (4.51)$$

is the matrix of moments (A, B, C) and products (D, E, F) of inertia (matrix representation of the inertia tensor C_{ij}).

$$\overline{h} = \int_{\text{earth}} [\overline{r}\wedge]\frac{d\overline{r}}{dt}\,dm \qquad (\overline{r} = [x\ y\ z]^T) \qquad (4.52)$$

is the relative angular momentum vector, and $\overline{L}$ is the vector of torques exerted on the earth.

Replacing $\overline{\omega}$ from Euler's geometric equations into the Liouville equation, we obtain the following second-order nonlinear differential equation

$$\underline{C}\,\underline{S}_{(\overline{e})}\ddot{\overline{e}} + \left\{\underline{C}\,\dot{\underline{S}}_{(\overline{e})} + \dot{\underline{C}}\,\underline{S}_{(\overline{e})}\right\}\dot{\overline{e}} + \dot{\overline{h}} + \left[(\underline{S}_{(\overline{e})}\dot{\overline{e}})\wedge\right]\left\{\underline{C}\,\underline{S}_{(\overline{e})}\dot{\overline{e}} + \overline{h}\right\} = \overline{L} \qquad (4.53)$$

To find an analytical solution $\overline{e}_{(t)}$ to the above deterministic equation for known $\underline{C}(t)$, $\overline{h}(t)$, $\overline{L}(t)$ and initial conditions $\overline{e}(t_0)$, $\dot{\overline{e}}(t_0)$ appears to be an impossible task without the help of some simplifying approximations.

We can write Liouville's equation symbolically as

$$\mathcal{L}\{\bar{e}(t)\} = \bar{L}(t) \tag{4.54}$$

where $\mathcal{L}$ stands for the relevant differential operator. In reality, the functions $\underline{C}(t)$, $\bar{h}(t)$, $\bar{L}(t)$ are only partially known, and they should be rather modeled as stochastic functions $\underline{C}(t, \omega)$, $\bar{h}(t, \omega)$, $\bar{L}(t, \omega)$. Liouville's equation now becomes a random equation

$$\mathcal{L}_{(\omega)}\{\bar{e}(t)\} = \bar{L}(t, \omega) \tag{4.55}$$

where $\mathcal{L}_{(\omega)}$ is a random operator. (See [Bharucha-Reid, 1972, p. 71] for relevant definition and discussion.) The proper mathematical theory for the treatment of such a random equation is probabilistic functional analysis. The solution is a stochastic process $\bar{e}(t, \omega)$ whose distribution depends on the (possibly random) initial conditions $\bar{e}(t_0)$, $\dot{\bar{e}}(t_0)$ and the distributions of the random functions $\underline{C}(t, \omega)$, $\bar{h}(t, \omega)$, $\bar{L}(t, \omega)$. The introduction of probabilistic concepts can only increase the difficulties in solving the above equation.

Alternatively, if we set $\bar{y}(t) = [\bar{e}(t)\ \dot{\bar{e}}(t)]^T$, we can rewrite the Liouville equation in the form

$$\frac{d\bar{y}(t)}{dt} = \begin{bmatrix} \dot{\bar{e}} \\ -(\underline{C}\,\underline{S})^{-1}\left\{[\underline{C}\,\dot{\underline{S}} + \dot{\underline{C}}\,\underline{S}]\dot{\bar{e}} + \dot{\bar{h}} + [(\underline{S}\dot{\bar{e}})\wedge](\underline{C}\,\underline{S}\bar{e} + \bar{h}) + \bar{L}\right\} \end{bmatrix} \tag{4.56}$$

or shortly

$$\frac{d\bar{y}(t)}{dt} = \bar{f}(\bar{y}(t), t) \tag{4.57}$$

The stochastic analogue of this differential equation is

$$\frac{d\bar{y}(t, \omega)}{dt} = \bar{f}(\bar{y}(t, \omega), t, \omega) \tag{4.58}$$

Upon integration we obtain a random nonlinear integral equation of the Volterra type [Bharucha-Reid, 1972, p. 187]:

$$\bar{y}(t, \omega) = \bar{y}(t_o, \omega) + \int_{t_o}^{t} \bar{f}(\bar{y}(s, \omega), s, \omega)\, ds \qquad (4.59)$$

Because of the complexity of the function $\bar{f}$, both the stochastic differential equation and its integral counterpart fall outside the types extensively studied by mathematicians. We shall therefore have to introduce some linearization giving rise to a sufficiently accurate approximate equation of a simpler form. To do this we need to introduce a more convenient set of parameters than $\bar{e}$, $\dot{\bar{e}}$ describing the rotational state of the earth.

4.3.3 Alternative Parameterization of the Rotation of the Earth

In view of Euler's geometric equations, an alternative set of state parameters is $[\bar{e}^T \bar{\omega}^T]^T$. The rotation vector $\bar{\omega}$ can be transformed into the same vector $\bar{\Omega}_I$ with respect to the inertial frame, with the help of the transformation

$$\bar{\Omega}_I(t) = R_3(\psi(t))\, R_1(\theta(t))\, R_3(\varphi(t))\, \bar{\omega}(t) = \underline{R}(\bar{e}(t))\, \bar{\omega}(t) \qquad (4.60)$$

We can now introduce an alternative parameterization which conforms to the traditional separation of the earth's rotation into three parts:

(a) Variation of the direction of the vector $\bar{\Omega}_I$ with respect to the inertial frame (precession-nutation).

(b) Variation of the direction of the vector $\bar{\omega}$ with respect to the earth-fixed frame (polar motion).

(c) Variation in the magnitude of the rotation vector, i.e., variation in the angular velocity of the rotation of the earth (length-of-day variations).

The matrix $\underline{M}(t)$ of transformation from the inertial to the earth-fixed system can be written in the form

$$\underline{M}(t) = R_2\left(-\tan^{-1}\frac{\omega_1}{\omega_2}\right) R_1\left(\tan^{-1}\frac{\omega_2}{\sqrt{\omega_1^2+\omega_3^2}}\right) R_3\left(\Theta_o + \int_{t_o}^{t} \Omega(\tau)\, d\tau\right)$$

$$R_1\left(-\tan^{-1}\frac{\Omega_{I2}}{\sqrt{\Omega_{I1}^2+\Omega_{I2}^2}}\right) R_2\left(\tan^{-1}\frac{\Omega_{I1}}{\Omega_{I3}}\right) \qquad (4.61)$$

This representation differs from the classical one [Mueller, 1969]:

$$\underline{M}(t) = R_2(-x_p)\, R_1(-y_p)\, R_3(\mathrm{GAST})\, R_1(-\epsilon - \Delta\epsilon)\, R_3(-\Delta\psi)\, R_1(\epsilon)$$
$$R_3(-z)\, R_2(\theta)\, R_3(-\zeta_o) \qquad (4.62)$$

by the fact that no differentiation is made between precession and nutation, and the R_3 part in the precession-nutation transformation has been included into the initial epoch angle Θ_o of the diurnal rotation.

The earth-fixed system can be chosen to be sufficiently close to the instantaneous rotation axis (at least for quite a long time interval where secular polar motion causes no problem) so that

$$\tan^{-1}\frac{\omega_1}{\omega_3} = \frac{\omega_1}{\omega_3} = \frac{\omega_1}{\Omega}$$

$$\tan^{-1}\frac{\omega_2}{\sqrt{\omega_1^2+\omega_3^2}} = \frac{\omega_2}{\sqrt{\omega_1^2+\omega_3^2}} = \frac{\omega_2}{\omega_3} = \frac{\omega_2}{\Omega} \qquad (4.63)$$

where

$$\Omega = \|\overline{\omega}\| = \|\overline{\Omega}_I\| = (\omega_1^2+\omega_2^2+\omega_3^2)^{\frac{1}{2}} \qquad (4.64)$$

The same approximation is not valid for $\overline{\Omega}_I$ because of its variation with respect to the inertial frame.

Introducing a new geocentric moving frame $\tilde{X}$, $\tilde{Y}$, $\tilde{Z}$ connected to the inertial one X, Y, Z through the precession-nutation theory transformation

$$\begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \end{bmatrix} = R_1(-\epsilon - \Delta\epsilon)\ R_3(-\Delta\psi)\ R_1(\epsilon)\ R_3(-z)\ R_2(\theta)\ R_3(-\zeta_o) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} =$$

$$= N(t)\ P(t) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{4.65}$$

the rotation vector $\overline{\Omega}$ with respect to $\tilde{X}$, $\tilde{Y}$, $\tilde{Z}$ is near the Z axis. With the help of the same approximations as for $\overline{\omega}$, we obtain

$$M(t) = Q\left(\frac{\omega_1}{\Omega}, \frac{\omega_2}{\Omega}\right)\ R_3\left(\Theta_o + \int_{t_o}^{t} \Omega(s)\, ds\right) Q^T\left(\frac{\Omega_1}{\Omega}, \frac{\Omega_2}{\Omega}\right)\ N(t)\ P(t) \tag{4.66}$$

where for small angles p, q

$$Q(p, q) = R_2(-p)\ R_1(q) = \begin{bmatrix} 1 & 0 & p \\ 0 & 1 & q \\ -p & -q & 1 \end{bmatrix} \tag{4.67}$$

4.3.4 The Linearized Liouville Equation

Since the rotation vector $\overline{\omega}$ is close to the z axis, it can be approximated to the first order by a vector $\overline{\omega}_o = [0 \quad 0 \quad \Omega_R]^T$, where Ω_R is an approximate constant reference value for the angular velocity of rotation $\Omega(t)$. We can introduce a vector $\overline{m}$ of small quantities defined as

$$\overline{\omega}(t) = \overline{\omega}_o + \delta\overline{\omega} \begin{bmatrix} 0 \\ 0 \\ \Omega_R \end{bmatrix} + \Omega_R\, \overline{m}(t) \tag{4.68}$$

In a similar way, the inertia tensor (matrix) can be approximated by a constant matrix

$$\underline{C}_o = \begin{bmatrix} A & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & C \end{bmatrix} \tag{4.69}$$

and a small correction matrix $\underline{c}(t)$ can be defined as

$$\underline{C}(t) = \underline{C}_o + \underline{c}(t) \tag{4.70}$$

Munk and MacDonald [1960, p. 38, Section 6.1] show that with the help of the above approximations and neglecting terms of products and squares of the small dimensionless quantities c_{ij}/C, m_i, and $h_i/(\Omega_R C)$, the Liouville equation can be written in the following linearized form:

$$\dot{\overline{m}}(t) = \sigma_r \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \overline{m}(t) + \begin{bmatrix} \sigma_r \varphi_2(t) \\ -\sigma_r \varphi_1(t) \\ \dot{\varphi}_3(t) \end{bmatrix} = \sigma_r \underline{P}\,\overline{m}(t) + \overline{f}^*(t) \tag{4.71}$$

where $\sigma_r = \dfrac{C - A}{A}\,\Omega_R$ is the Eulerian frequency (period $\dfrac{2\pi}{\sigma_r} \approx 10$ months) and

$$\overline{f}^* = \begin{bmatrix} \sigma_r \varphi_2 \\ -\sigma_r \varphi_1 \\ \dot{\varphi}_3 \end{bmatrix} = \begin{bmatrix} A^{-1}(\Omega_R c_{23} - \dot{c}_{13} + h_2 - \Omega_R^{-1} \dot{h}_1 + \Omega_R^{-1} L_1) \\ A^{-1}(-\Omega_R c_{13} - \dot{c}_{23} - h_1 - \Omega_R^{-1} \dot{h}_2 + \Omega_R^{-1} L_2) \\ C^{-1}(\quad - \dot{c}_{33} \quad - \Omega_R^{-1} \dot{h}_3 + \Omega_R^{-1} L_3) \end{bmatrix} =$$

$$= \underline{C}_o^{-1} \{ \Omega_R \underline{P}\,\overline{c} - \dot{\overline{c}} + \underline{P}\overline{h} - \Omega_R^{-1} \dot{\overline{h}} + \Omega_R^{-1} \overline{L} \} \tag{4.72}$$

with $\overline{c} = [c_{13} \;\; c_{23} \;\; c_{33}]^T$.

Matrix notation is more convenient for computations, but we will also use complex notation for its analytical advantages. In complex notation, the first two of the Liouville equations become

$$i \frac{\dot{\tilde{m}}}{\sigma_r} + \tilde{m} = \tilde{\varphi} \tag{4.73}$$

with

$$\begin{aligned} \tilde{m} &= m_1 + i m_2 \\ \tilde{\varphi} &= \varphi_1 + i \varphi_2 \end{aligned} \tag{4.74}$$

Setting $\tilde{f}^* = f_1^* + i f_2^*$, we have $\tilde{\varphi} = \frac{i}{\sigma_r} \tilde{f}^*$ and

$$\dot{\tilde{m}} = i \sigma_r \tilde{m} + \tilde{f}^* \tag{4.75}$$

We shall call $\bar{f}^*$ the total excitation function in discrepancy with Munk and MacDonald who call $\tilde{\varphi}$ the excitation function.

Of primary importance is the part of the excitation function $\bar{f}_D$, due to the rotational deformation of the earth. Rotational deformation gives rise to the following changes in the products of inertia:

$$\begin{aligned} c_{13} &= (C-A)\frac{k}{k_f} m_1 = A \Omega_R^{-1} \sigma_r \frac{k}{k_f} m_1 \\ c_{23} &= A \Omega_R^{-1} \sigma_r \frac{k}{k_f} m_2 \end{aligned} \tag{4.76}$$

where k is one of the Love numbers and k_f the corresponding "fluid" Love number [Munk and MacDonald, 1960, Chapter 5]. Using these product of inertia changes, we obtain the excitation due to rotational deformation

$$\bar{f}_D = \sigma_r \frac{k}{k_f} \begin{bmatrix} m_2 \\ -m_1 \\ 0 \end{bmatrix} = - \sigma_r \frac{k}{k_f} \underline{P}\, \bar{m} \tag{4.77}$$

Setting $\bar{f}^* = \bar{f}_D + \bar{f}$, where $\bar{f}$ is the remaining, or simply the excitation, function, we obtain

$$\dot{\bar{m}} = \sigma_r \underline{P}\, \bar{m} - \sigma_r \frac{k}{k_f} \underline{P}\, \bar{m} + \bar{f} = \sigma_r \left(1 - \frac{k}{k_f}\right) \underline{P}\, \bar{m} + \bar{f} \tag{4.78}$$

$$\dot{\bar{m}} = \sigma_o \underline{P}\, \bar{m} + f \tag{4.79}$$

where $\sigma_o = \sigma_r (1 - k/k_f)$ is the Chandler frequency (period $\frac{2\pi}{\sigma_o} \approx 14$ months). In complex notation we have, after setting $\tilde{f} = f_1 + i f_2$,

$$\dot{\tilde{m}} = i \sigma_o \tilde{m} + \tilde{f} \tag{4.80}$$

The solution of the above equation with $f = 0$ gives rise to a circular motion of the rotation axis with frequency σ_o (Chandler wobble). To explain the

broadening of the spectral peak corresponding to the Chandler frequency, it has been suggested that the earth should be regarded as a damped linear oscillator tuned to the Chandler frequency and irregularly excited [Rochester, 1970, p. 9].

To account for damping, a new linear term can be added to the linearized Liouville equation

$$\dot{\overline{m}} = \sigma_o \underline{P}\,\overline{m} - \frac{1}{\tau}\underline{D}\,\overline{m} + \overline{f} \qquad (4.81)$$

with

$$\underline{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad (4.82)$$

where τ is the relaxation time associated with the damping. In complex notation we can introduce the concept of a complex frequency $\sigma_c = \sigma_o + i/\tau$ to obtain [Smylie et al., 1973, p. 395],

$$\dot{\tilde{m}} = i\,\sigma_c\,\tilde{m} + \tilde{f} \qquad (4.83)$$

4.3.5 Solution of the Linearized Liouville Equation

We are concerned here with the solution of the linearized Liouville equation without damping

$$\dot{\overline{m}}(t) = \sigma_o \underline{P}\,\overline{m}(t) + \overline{f}(t) \qquad (4.84)$$

The last of the above three equations has the obvious solution

$$m_3(t) = m_3(t_o) + \int_{t_o}^{t} f_3(s)\,ds \qquad (4.85)$$

while the first two can be written in complex form

$$\dot{\tilde{m}}(t) = i\,\sigma_o\,\tilde{m}(t) + \tilde{f}(t) \qquad (4.86)$$

with general solution [Munk and MacDonald, 1960, p. 46, equation (6.7.1)]

$$\tilde{m}(t) = e^{i\sigma_o t}\,\tilde{m}_o + \int_{-\infty}^{t} f(\tau)\,d\tau \tag{4.87}$$

Setting $t = t_o$ and solving for $\tilde{m}_o$ we obtain

$$\tilde{m}_o = e^{-i\sigma_o t_o}\,\tilde{m}(t_o) - \int_{-\infty}^{t_o} e^{-i\sigma_o \tau}\, f(\tau)\,d\tau \tag{4.88}$$

Setting this value in the general solution, we have

$$\tilde{m}(t) = e^{i\sigma_o (t-t_o)}\,\tilde{m}(t_o) + \int_{t_o}^{t} e^{i\sigma_o (t-\tau)}\, f(\tau)\,d\tau \tag{4.89}$$

Separating real and imaginary parts and combining with the solution for $m_3(t)$ into matrix notation, we finally have

$$\overline{m}(t) = R_3[-\sigma_o (t-t_o)]\,\overline{m}(t_o) + \int_{t_o}^{t} R_3[-\sigma_o (t-\tau)]\,\overline{f}(\tau)\,d\tau \tag{4.90}$$

It is well known [McGarty, 1974, Section 2.2] that the solution to a linear differential equation

$$\frac{dx(t)}{dt} = A_{(t)}\,x(t) + y(t) \tag{4.91}$$

is of the form

$$x(t) = \Phi(t, t_o)\,x(t_o) + \int_{t_o}^{t} \Phi(t, \tau)\,y(\tau)\,d\tau \tag{4.92}$$

where $\Phi(t, s)$, called the state transition matrix, is the solution of the differential equation

$$\frac{d}{dt}\,\Phi(t,\ t_o) \;=\; A(t)\ \Phi(t,\ t_o) \tag{4.93}$$

with initial condition $\Phi(t_o,\ t_o) = I$, the identity matrix.

In our case it is easy to show that

$$\Phi(t,\ s) \;=\; R_3\left[-\sigma_o\,(t-s)\right] \tag{4.94}$$

Differentiating with respect to t, we obtain

$$\frac{d}{dt}\,\Phi(t,\ t_o) \;=\; \frac{d}{dt}\,R_3\left[-\sigma_o\,(t-t_o)\right] \;=\; \sigma_o\,\underline{P}\,R_3\left[-\sigma_o\,(t-t_o)\right] \tag{4.95}$$

and also the initial condition is satisfied, since

$$\Phi(t,\ t) \;=\; R_3\left[-\sigma_o\,(t-t)\right] \;=\; R_3(O) \;=\; I \tag{4.96}$$

for any t. $R_3\left[-\sigma_o\,(t-s)\right]$ also satisfies the transition property

$$\Phi(t_1,\ t_2)\,\Phi(t_2,\ t_3) \;=\; R_3\left[-\sigma_o\,(t_1-t_2)\right]\ R_3\left[-\sigma_o\,(t_2-t_3)\right] \;=$$

$$=\; R_3\left[-\sigma_o\,(t_1-t_2+t_2-t_3)\right] \;=\; R_3\left[-\sigma_o\,(t_1-t_3)\right] \;=\; \Phi(t_1,\ t_3) \tag{4.97}$$

and the property

$$\Phi^{-1}(t,\ s) \;=\; R_3^{-1}\left[-\sigma_o\,(t-s)\right] \;=\; R_3\left[-\sigma_o\,(s-t)\right] \;=$$

$$=\; \Phi(s,\ t) \;=\; \Phi^{T}(t,\ s) \tag{4.98}$$

4.3.6 Stochastic Solutions of the Linearized Liouville Equation

Because of the uncertainty about the exact form of the excitation function, it is not possible to obtain a solution for $\overline{m}(t)$ in terms of a finite number of parameters to be estimated from sufficient observational data. Instead, a stochastic model for $\overline{f}(t)$ is more appropriate to account for existing uncertainties.

We shall assume that the excitation function is modeled as a second-order stochastic process $f(t,\ \omega)$ with known mean value function

$$\overline{\mu}_f(t) \;=\; E\left\{\overline{f}(t,\ \omega)\right\} \tag{4.99}$$

and correlation matrix

$$R_{ff}(t, s) = E\{\overline{f}(t, \omega)\,\overline{f}^{T}(s, \omega)\} \tag{4.100}$$

The solution is also a stochastic process $\overline{m}(t, \omega)$ with mean value function

$$\overline{\mu}_m(t) = R_3[-\sigma_0(t-t_0)]\,\overline{m}(t_0) + \int_{t_0}^{t} R_3[-\sigma_0(t-\tau)]\,\overline{\mu}_f(\tau)\,d\tau \tag{4.101}$$

and correlation matrix $R_{mm}(t, s)$. To avoid complicated formulas we introduce

$$\begin{aligned} \overline{z}(t) &= R_3[-\sigma_0(t-t_0)]\,\overline{m}(t_0) \\ \overline{y}(t, \omega) &= \overline{m}(t, \omega) - \overline{z}(t) \end{aligned} \tag{4.102}$$

with

$$\overline{\mu}_y(t) = E\{\overline{y}(t, \omega)\} = \int_{t_0}^{t} R_3[-\sigma_0(t-\tau)]\,\overline{\mu}_f(\tau)\,d\tau \tag{4.103}$$

We now have

$$\overline{\mu}_m(t) = \overline{z}(t) + \overline{\mu}_y(t) \tag{4.104}$$

and

$$\begin{aligned} R_{mm}(t, s) &= \overline{z}(t)\,\overline{z}^{T}(s) + \overline{\mu}_y(t)\,\overline{z}^{T}(s) + \\ &+ \overline{z}(t)\,\overline{\mu}_y^{T}(s) + R_{yy} \end{aligned} \tag{4.105}$$

where

$$R_{yy}(t, s) = \int_{\xi=t_0}^{t}\int_{\zeta=t_0}^{s} R_3[\sigma_0(\xi-t)]\,R_{ff}(\xi, \zeta)\,R_3[\sigma_0(s-\zeta)]\,d\xi\,d\zeta \tag{4.106}$$

The covariance matrix of $m(t, \omega)$ can easily be found to be

$$C_{mm}(t, s) = R_{mm}(t, s) - \overline{\mu}_m(t)\,\overline{\mu}_m^{T}(s) = R_{yy}(t, s) - \overline{\mu}_y(t)\,\overline{\mu}_y^{T}(s) \tag{4.107}$$

We have assumed here that the initial value $\overline{m}(t_o)$ is deterministic, i.e., either completely known or completely unknown (a parameter to be determined from observations). If some a priori estimate of $\overline{m}(t_o)$ is available, it can be modeled as a vector random variable with mean $\overline{\mu}_o$ and correlation matrix $\underline{R}_o$. Under the additional assumption that $\overline{m}(t_o, \omega)$ is uncorrelated to $\overline{f}(t, \omega)$, so that $E\{\overline{m}(t_o, \omega)\ \overline{f}(t, \omega)\} = \underline{0}$ for every t, we obtain

$$\overline{\mu}_m(t) = R_3[-\sigma_o(t-t_o)]\,\overline{\mu}_o + \int_{t_o}^{t} R_3[-\sigma_o(t-\tau)]\ \mu_f(\tau)\,d\tau =$$

$$= \overline{\mu}_z(t) + \overline{\mu}_y(t) \qquad (4.108)$$

$$R_{mm}(t, s) = R_3[-\sigma_o(t-t_o)]\ \underline{R}_o\ R_3[\sigma_o(s-t_o)] +$$

$$+ \int_{\xi=t_o}^{t}\int_{\zeta=t_o}^{s} R_3[-\sigma_o(t-\xi)]\ R_{ff}(\xi,\zeta)\ R_3[\sigma_o(s-\zeta)]\,d\xi\,d\zeta \qquad (4.109)$$

and

$$C_{mm}(t, s) = R_{mm}(t, s) - \overline{\mu}_m(t)\ \overline{\mu}_m^T(s) \qquad (4.110)$$

From the mean and correlation of $\overline{m}(t, \omega)$, the corresponding mean and correlation of the rotation vector $\overline{\omega}(t, \omega)$ can be found with the help of the the transformation $\overline{\omega} = \overline{\omega}_o + \Omega_R\,\overline{m}$

$$\overline{\mu}_\omega(t) = \begin{bmatrix} 0 \\ 0 \\ \Omega_R \end{bmatrix} + \Omega_R\ \overline{\mu}_m(t) \qquad (4.111)$$

and

$$R_{\omega\omega}(t, s) = \Omega_R^2 \begin{bmatrix} 0 & 0 & \mu_{m_1}(s) \\ 0 & 0 & \mu_{m_2}(s) \\ \mu_{m_1}(t) & \mu_{m_2}(t) & 1+\mu_{m_3}(t)\ \mu_{m_3}(s) \end{bmatrix} + \Omega_R^2\ R_{mm}(t, s)$$

(4.112)

We have parameterized the rotation of the earth in terms of ω_1/Ω, ω_2/Ω, Ω, Ω_1/Ω, Ω_2/Ω. These parameters can be sufficiently approximated by $\omega_1/\Omega_R = m_1$, $\omega_2/\Omega_R = m_2$, $\omega_3 = \Omega_R(1+m_3)$, Ω_1/Ω_R, Ω_2/Ω_R respectively. Setting $M_1 = \Omega_1/\Omega_R$, $M_2 = \Omega_2/\Omega_R$, the rotation transformation becomes

$$M(t) = Q(m_1, m_2)\, R_3 \left[\Theta_0 + \Omega_R (t-t_0) + \Omega_R \int_{t_0}^{t} m_3(\tau)\, d\tau\right]$$

$$Q^T(M_1, M_2)\, N(t)\, P(t) \qquad (4.113)$$

In addition to m_1, m_2, m_3, we also need the statistics of M_1 and M_2. As an intermediate step we need to obtain the mean and covariance function of the rotation vector with respect to the inertial frame $\overline{\Omega_I}$ from the mean and covariance of the vector $\overline{\omega}$. Unfortunately, this requires the solution of the nonlinear stochastic differential equation

$$\frac{d\,\overline{e}(t)}{dt} = F(\overline{e}, \overline{\omega}) = S^{-1}(\overline{e})\,\overline{\omega} \qquad (4.114)$$

Even approximate solutions involve computationally tedious numerical integrations. We can use the mean value function $\overline{\mu}_\omega$ of $\overline{\omega}$ to obtain through numerical integration an approximate solution $\overline{e}_0$ of the deterministic equation

$$\frac{d\,\overline{e}_0}{dt} = S^{-1}(\overline{e}_0)\,\overline{\mu}_\omega \qquad (4.115)$$

A linearization is now possible through Taylor series expansion about $\overline{e}_0$, $\overline{\mu}_\omega$ and neglect of second- and higher-order terms. Introducing $\delta\overline{e} = \overline{e} - \overline{e}_0$, $\delta\overline{\omega} = \overline{\omega} - \overline{\mu}_\omega$, we obtain

$$\frac{d\,\delta\overline{e}(t)}{dt} = A(t)\,\delta\overline{e}(t) + B(t)\,\delta\overline{\omega}(t) \qquad (4.116)$$

where

$$A(t) = \left.\frac{\partial \dot{\bar{e}}}{\partial \bar{e}}\right|_{\bar{e}_o,\, \bar{\mu}_\omega} = \frac{\partial [S^{-1}(\bar{e}_o)]}{\partial \bar{e}_o}\, \bar{\mu}_\omega \tag{4.117}$$

and

$$B(t) = \left.\frac{\partial \dot{\bar{e}}}{\partial \omega}\right|_{\bar{e}_o} = S^{-1}(e_o) \tag{4.118}$$

Numerical integration of

$$\frac{d}{dt}\,\Phi(t,\, t_o) = A(t)\;\Phi(t,\, t_o) \tag{4.119}$$

with

$$\Phi(t_o,\, t_o) = I \tag{4.120}$$

yields the state transition matrix $\Phi(t, s)$, which can be used to find the mean $\bar{\mu}$ and correlation $K(t, s)$ of $\delta\bar{e}(t)$

$$\bar{\mu}(t) = \Phi(t, t_o)\,\bar{\mu}_{\delta\bar{e}}(t_o) + \int_{t_o}^{t} \Phi(t, \tau)\, B(\tau)\,\bar{\mu}_{\delta\omega}(\tau)\, d\tau \tag{4.121}$$

$$K(t, s) = \Phi(t, t_o)\;E\{\delta\bar{e}(t_o)\,\delta\bar{e}^{T}(t_o)\}\;\Phi^{T}(s,\, t_o) + \int_{\xi=t_o}^{t}\int_{\zeta=t_o}^{s} \Phi(t, \xi)\, B(\xi)\, E\{\delta\bar{\omega}(\xi)\,\delta\bar{\omega}^{T}(\zeta)\}\, B^{T}(\zeta)\,\Phi^{T}(s, \zeta)\, d\xi\, d\zeta \tag{4.122}$$

A similar linearization of the equation

$$\bar{\Omega}_I = R(\bar{e})\,\bar{\omega} = R_3(\psi)\;R_1(\theta)\;R_3(\varphi)\;\bar{\omega} \tag{4.123}$$

leads to approximate statistics for Ω_I with the use of the known mean and covariance functions of $\bar{e}$ and $\bar{\omega}$, and finally to those of $\bar{\Omega}$ and M_1, M_2. Such a laborious propagation might be of some value despite the approximate linearizations involved if the true mean and covariance function of the excitation function were known. However, this is not the case at present. To obtain such "true" statistics one should identify all physical processes

contributing to the total excitation function, and also estimate their mean and covariance functions. The present uncertainties about mechanisms with effects on the earth's rotation [Rochester, 1973] limit the feasibility of such an approach. The alternative is to use a simple model covariance function for the excitation function, which is computationally tractable and hopefully approximates the true one. In this case the propagation from the statistics of m_1, m_2, m_3 to those of M_1, M_2, simply offers model consistency which might be questionable in view of the approximations involved in the linearizations. An independent simple covariance model for M_1, M_2 may serve our purpose and also considerably reduce the computational effort involved.

Some more or less obvious properties that one might postulate on the excitation function are zero mean and stationarity of the covariance function.

A zero mean function gives rise to a circular motion of the pole with the Chandler frequency, corresponding to the "expected" behavior of the polar motion. Other irregularities superimposed on the Chandler wobble are accounted for by the fluctuations of the actual excitation function about its mean value function. However, the linearized Liouville equations holds for an arbitrary reference frame, while changes in reference frame cause changes in the excitation function. With this in mind, the zero mean condition on the excitation function must obviously hold for only one particular reference frame. Since

$$\underline{C}_0 \, \overline{f} = \Omega_R \, \underline{P} \, \overline{c} + \dot{\overline{c}} + \underline{P} \, \overline{h} - \frac{1}{\Omega_R} \dot{\overline{h}} + \frac{1}{\Omega_R} \overline{L} \qquad (4.124)$$

and, assuming,

$$E\{\dot{\overline{c}}\} = E\{\overline{h}\} = E\{\dot{\overline{h}}\} = E\{\overline{L}\} = \overline{0} \qquad (4.125)$$

the condition $E\{\overline{f}\} = \overline{0}$ implies

$$E\{\overline{c}\} = \overline{0}, \quad \text{i.e.,} \quad E\{c_{i3}\} = 0, \; i = 1, 2, 3 \qquad (4.126)$$

For this to be true it is necessary that the direction of the z-axis in the earth-fixed frame coincide with the "expected" direction of the corresponding principal axis of inertia z_I. If such an estimate is not available, the introduction of two additional parameters of rotation from the z_I to the z-axis

$$\bar{X} = R_2(\mu)\ R_1(-\lambda)\ \bar{X}_I \approx \begin{bmatrix} 1 & 0 & -\mu \\ 0 & 1 & -\lambda \\ \mu & \lambda & 1 \end{bmatrix} \bar{X}_I \qquad (4.127)$$

leads to a new excitation function $\bar{f}$ (the old one being $\bar{f}_I$ with $E\{\bar{f}_I\} = \bar{0}$) with

$$E\{\bar{f}\} = \sigma_o \begin{bmatrix} \lambda \\ -\mu \\ 0 \end{bmatrix} \qquad (4.128)$$

The frequency σ_o in place of the original one $\sigma_r = A^{-1}(C-A)\,\Omega_R$ accounts for the compensation of the difference in the rotational deformation part of the total excitation function. Use of the above mean value function results in the following mean value of the solution

$$E\{\bar{m}(t)\} = \delta + R_3[-\sigma_o(t-t_o)]\ [E\{\bar{m}(t_o)\} - \delta] \qquad (4.129)$$

where

$$\delta = [\mu\ \lambda\ 0]^T \qquad (4.130)$$

It can be seen by plotting the solution on the xy-plane (see Figure 4.1) that the difference is only in a translation corresponding to the change of direction of the z-axis.

The mean value function, except from the above effect due to the deviation of the z-axis from the principal axis of inertia, may also include periodic terms already identified from past experience in polar motion data analysis (annual, semiannual terms, etc.).

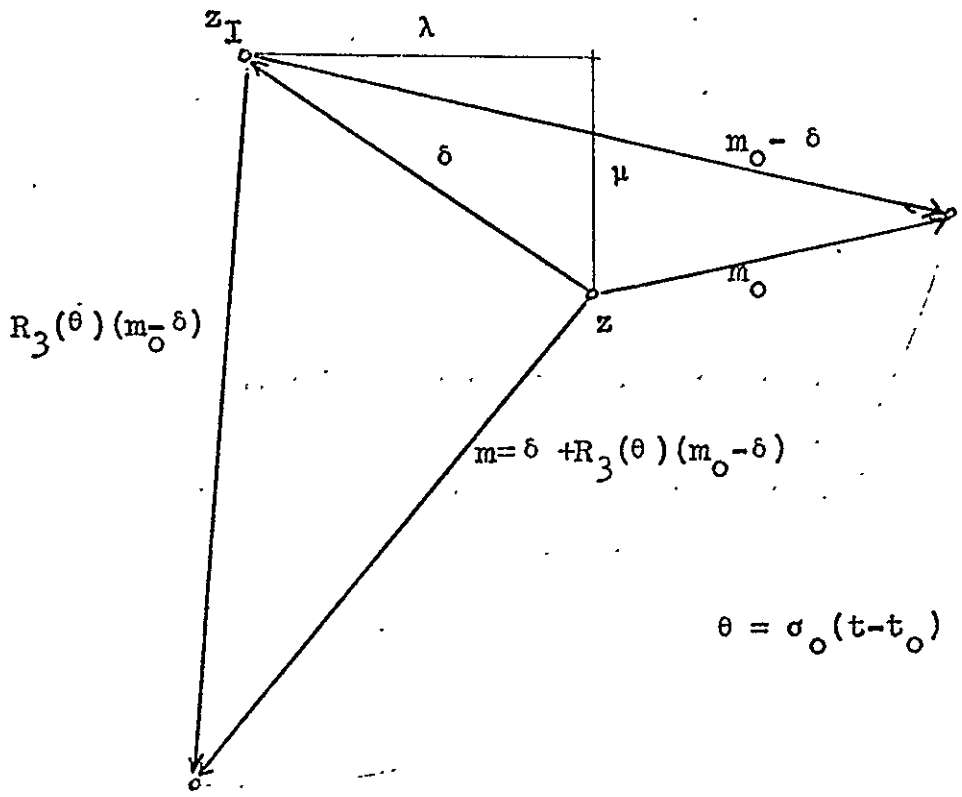

Figure 4.1 Effect of the Change of Origin in the Liouville Equation

The stationarity assumption of the covariance function corresponds to the concept of regularity in the physical processes giving rise to earth rotation, at least for periods of time which are short on a geological time scale. This regularity is essential in introducing the concept of ergodicity, which legitimizes the very use of probabilistic models.

4.3.7 Solution of the Linearized Liouville Equation with Damping

To include the effect of damping, a linear term is added to the linearized Liouville equation

$$\dot{\overline{m}}(t) = \sigma_o \underline{P}\, \overline{m}(t) - \frac{1}{\tau}\, \underline{D}\, \overline{m}(t) + \overline{f}(t) \tag{4.131}$$

with

$$\underline{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.132}$$

The only difference appears in the first two equations which in complex notation read

$$\dot{\widetilde{m}}(t) = i\, \widetilde{\sigma}_c\, \widetilde{m}(t) + \widetilde{f}(t) \tag{4.133}$$

with

$$\widetilde{\sigma}_c = \sigma_o + \frac{i}{\tau} \tag{4.134}$$

The general deterministic solution is

$$\widetilde{m}(t) = e^{i\widetilde{\sigma}_c t}\left[\widetilde{m}_o + \int_{-\infty}^{t} \widetilde{f}(\xi)\, e^{-i\widetilde{\sigma}_c \xi}\, d\xi\right] \tag{4.135}$$

Setting $t = t_o$, solving for $\widetilde{m}_o$, and replacing it in the above equation, we finally obtain

$$\tilde{m}(t) = e^{i\tilde{\sigma}_c (t-t_o)} \tilde{m}(t_o) + \int_{t_o}^{t} e^{i\tilde{\sigma}_c (t-\xi)} \tilde{f}(\xi)\, d\xi =$$

$$= e^{i\sigma_o (t-t_o)}\, e^{-(t-t_o)/\tau}\, \tilde{m}(t_o) + \int_{t_o}^{t} e^{i\sigma_o (t-\xi)}$$

$$e^{-(t-\xi)/\tau}\, \tilde{f}(\xi)\, d\xi \qquad (4.136)$$

Combining with the unchanged solution of $m_3(t)$ in matrix notation we have

$$\bar{m}(t) = R_3[-\sigma_o (t-t_o)]\; Q(t-t_o)\; \bar{m}(t_o) + \int_{t_o}^{t} R_3[-\sigma_o (t-\xi)]\; Q(t-\xi)\; \bar{f}(\xi)\, d\xi \qquad (4.137)$$

where

$$Q(a) = \begin{bmatrix} e^{-a/\tau} & 0 & 0 \\ 0 & e^{-a/\tau} & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (4.138)$$

is the "damping" matrix with the property

$$Q(a)\; R_3(\theta) = R_3(\theta)\; Q(a) \qquad (4.139)$$

The corresponding state transition matrix obviously is

$$\Phi(t, s) = R_3[-\sigma_o (t-s)]\; Q(t-s) \qquad (4.140)$$

In the absence of any excitation, the corresponding free motion (on the xy-plane) is a contracting spiral whose center is the inertia symmetry axis [Rudnick, 1956, p. 137]. Since the Chandler wobble is known to be maintained, the mean value function of the excitation must be such that

$$\tilde{\mu}_m(t) = E\{\tilde{m}(t)\} = e^{i\sigma_o (t-t_o)}\, \tilde{a} \qquad (4.141)$$

This implies that

$$\frac{d}{dt}\tilde{\mu}_m = i\sigma_o e^{i\sigma_o(t-t_o)}\tilde{a} = i\sigma_c\tilde{\mu}_m + \tilde{\mu}_f \qquad (4.142)$$

and, therefore,

$$\tilde{\mu}_f(t) = E\{\tilde{f}(t)\} = \frac{1}{\tau} e^{i\sigma_o(t-t_o)}\tilde{a} \qquad (4.143)$$

The above condition must hold for proposed models of the excitation function.

The damping-excitation hypothesis, put forward as an explanation of the broadening of the spectral peak corresponding to the Chandler frequency, poses to this day two outstanding problems in polar motion analysis. The first problem is the nature of the damping or dissipation mechanism, manifested in our equations by the uncertainty in the value of the decay time τ. The second problem, the nature of the excitation of the Chandler wobble, has been the matter of much controversy, mostly centered about the role of seismic activity and initiated by the work of Mansinha and Smylie [1967]. A detailed account of the problem can be found in [Dahlen, 1971].

In general, the excitation function may be considered to consist of two parts

$$\bar{f} = \bar{f}_c + \bar{f}_p \qquad (4.144)$$

a continuous one $\bar{f}_c$ and a discontinuous one $\bar{f}_p$, the latter associated with abrupt changes of the earth's inertia tensor caused by earthquakes.

The discontinuous excitation function is of the form

$$\begin{aligned} f_p &= \underline{C}_o^{-1}\{\Omega_R \underline{P}\,\bar{c} - \dot{\bar{c}}\} \\ \bar{c} &= [c_{13}\ c_{23}\ c_{33}]^T \end{aligned} \qquad (4.145)$$

The first two components in complex notation are

$$\begin{aligned} \tilde{f}_p &= f_{p1} + f_{p2}\,i = A^{-1}\{i\Omega_R\tilde{c} + \dot{\tilde{c}}\} \\ \tilde{c} &= c_{13} + i\,c_{23} \end{aligned} \qquad (4.146)$$

If $\Delta\bar{c}_j$ ($\Delta\tilde{c}_j$) denotes the change in $\bar{c}$ caused by the j^{th} earthquake after some initial epoch t_o, occurring at epoch t_j, we have [Mansinha and Smylie, 1967,

p. 4733; Dahlen, 1971, p. 162]

$$\bar{c}(t) = \sum_j \Delta\bar{c}_j \; u(t-t_j)$$
$$\dot{\bar{c}}(t) = \sum_j \Delta\bar{c}_j \; \delta(t-t_j) \qquad (4.147)$$

where

$$u(t-t_j) = \begin{cases} 0 & t < t_j \\ 1 & t \geq t_j \end{cases} \quad \text{is the Heaviside step function}$$

$$\delta(t-t_j) = \begin{cases} 0 & t \neq t_j \\ 1 & t = t_j \end{cases} \quad \text{is the Dirac delta function}$$

The excitation function becomes

$$\tilde{f}_P(\xi) = -\frac{i\,\Omega_R}{A}\sum_j \Delta\tilde{c}_j \; u(\xi-t_j) - \frac{1}{A}\sum_j \Delta\tilde{c}_j \; (\xi-t_j) \qquad (4.148)$$

Inserting $\tilde{f}_P$ into the solution of the Liouville equation with damping, we arrive after formal integration to

$$\tilde{m}(t) = e^{i\tilde{\sigma}_c(t-t_o)}\,\tilde{m}(t_o) + \frac{\Omega}{A\tilde{\sigma}_c}\sum_j \Delta\tilde{c}_j - \frac{(\Omega+\tilde{\sigma}_c)}{A\tilde{\sigma}_c}\sum_j \Delta\tilde{c}_j \; e^{i\tilde{\sigma}_c(t-t_j)} \qquad (4.149)$$

Separating real from imaginary parts and combining with the solution of the third component

$$m_3 = m_3(t_o) + \int_{t_o}^{t} f_3(\xi)\,d\xi =$$
$$= m_3(t_o) + \int_{t_o}^{t}\left[-\frac{1}{C}\sum_j \Delta c_{33_j} \; \delta(\xi-t_j)\right] d\xi =$$
$$m_3(t_o) - \frac{1}{C}\sum_j \Delta c_{33_j} \qquad (4.150)$$

In matrix notation

$$\bar{m}(t) = R_3[-\sigma_o(t-t_o)]\ Q(t-t_o)\ \bar{m}(t_o) + \sum_j \underline{B}\,\Delta\bar{c}_j + \sum_j \underline{G}\ R_3[-\sigma_o(t-t_j)]\ Q(t-t_j)\ \Delta\bar{c}_j \qquad (4.151)$$

where

$$\underline{G} = \frac{1}{A(\sigma_o^2 + \tau^{-2})} \begin{bmatrix} \Omega\sigma_o + \sigma_o^2 + \tau^{-2} & \Omega/\tau & 0 \\ -\Omega/\tau & \Omega\sigma_o + \sigma_o^2 + \tau^{-2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad (4.152)$$

$$\underline{B} = \frac{\Omega}{\sigma_o^2 + \tau^{-2}} \begin{bmatrix} \sigma_o/A & 1/\tau A & 0 \\ -1/\tau A & \sigma_o/A & 0 \\ 0 & 0 & -1/C \end{bmatrix} \qquad (4.153)$$

(Note that all matrices Q, $\underline{B}$, $\underline{G}$, $R_3(\theta)$ commute.)

The above equation is a solution to the deterministic problem when the occurrence epochs and the effect of the earthquakes on the inertia tensor of the earth are known. It can also be viewed as the solution corresponding to a sample function in the stochastic case where uncertainty is present in both the occurrence epochs t_j and the inertia tensor changes $\Delta\bar{c}_j$. A stochastic model can be constructed with the help of two stochastic processes, the (homogenous) Poisson count process and the filtered Poisson process. Refer to [Parzen, 1962, Chapter 4; Snyder, 1975, Chapters 2 and 4] for the relevant rigorous definitions.

The Poisson count process N(t) refers to the number of earthquakes occurring in the interval $[t_o, t]$ ($N(t_o) = 0$ w.p.1), and it is characterized by the probability

$$P\{N_{(t)} - N_{(s)} = m\} = \frac{e^{-\lambda(t-s)}\lambda^m (t-s)^m}{m!} \qquad t > s > t_o \qquad (4.154)$$

where $\lambda > 0$ is called the intensity (or mean rate) of the process, in view of

$$E\{N(t)\} = E\{N(t) - N(t_o)\} = \lambda(t - t_o) \tag{4.155}$$

A filtered Poisson process is a stochastic process of the form

$$\overline{X}(t) = \sum_{j=1}^{N(t)} \overline{g}(t, t_j, \overline{Y}_j) \tag{4.156}$$

where $\overline{Y}_j$ is a sequence of independent random vectors identically distributed with a random vector $\overline{Y}$.

The part of the damping-excitation solution $\overline{m}(t)$ of the Liouville equation, independent of the initial state $\overline{m}(t_o)$

$$\overline{X}(t) = \sum_{j=1}^{N(t)} \{\underline{B} + \underline{G} R_3 - \sigma_o(t - t_j)\ Q(t - t_j)\}\ \Delta\overline{c}_j \tag{4.157}$$

is a filtered Poisson process with

$$\overline{g}(t, t_j, \Delta\overline{c}_j) = \{\underline{B} + \underline{G} R_3 - \sigma_o(t - t_j)\ Q(t - t_j)\}\ \Delta\overline{c}_j \tag{4.158}$$

provided that earthquakes occur with a probability distribution such as described by the Poisson count process and the corresponding inertia tensor changes $\Delta\overline{c}_j$ are identically distributed.

If $E\{\Delta\overline{c}_j\} = E\{\Delta\overline{c}\}$ and $E\{\Delta\overline{c}_j\ \Delta\overline{c}_j^T\} = E\{\Delta\overline{c}\ \Delta\overline{c}^T\}$ are known, the mean value and correlation of $\overline{X}(t)$ can be found with the help of

$$E\{\overline{X}(t)\} = \lambda \int_{t_o}^{t} E\{\overline{g}(t, \tau, \Delta\overline{c})\ d\tau \tag{4.159}$$

and

$$E\{\overline{X}(t)\,\overline{X}^T(s)\} = \lambda \int_{t_o}^{\min(t, s)} E\{\overline{g}(t, \tau, \Delta\overline{c})\,\overline{g}^T(s, \tau, \Delta\overline{c})\}\ d\tau \tag{4.160}$$

A usual simplification is to replace the epochs t_j as described by N(t), with a sequence of equidistant epochs, such that

$$t_{j+1} - t_j = \Delta t = \lambda^{-1} \tag{4.161}$$

in which case the filtered Poisson process is replaced by a random walk in R^3 [Mansinha and Smylie, 1967, p. 4739].

5. ADAPTIVE ESTIMATION

5.1 Introductory Remarks

Perhaps the most intriguing problem in least squares estimation, is the determination of weights. Both least squares adjustment and prediction are based upon the a priori knowledge of means and covariances of random parameters and stochastic processes (signals). The presence of functions with uncertainties significant compared to the level of observational noise, necessitates the inclusion of their effect in modeling. One way to circumvent this problem is to include only the values of the function at observation points (epochs) as unknowns and to secure enough observations to estimate all the unknowns. This is actually done in geometric methods of satellite geodesy, where simultaneity of observations increases the number of observations without a similar increase in unknowns (satellite positions). However, this is not always possible and treating each function value as an independent unknown generally results in overparameterization.

One way to avoid this problem is to represent the unknown function(s) in terms of a finite number of parameters, using polynomials, trigonometric series, step functions, etc. Although with a sufficient number of terms it is possible to approximate a wide class of functions arbitrarily well with the help of such representations, the fact that the function to be approximated is unknown poses some serious problems. One has to determine the number of terms to be included; and even then some of the coefficients of the terms included might be very poorly determined from the available observations, thus giving rise to singularities and computational problems.

The alternative approach is to model unknown functions as stochastic processes such that their probability distributions determine the "likely" behaviour of the unknown functions. When the functions to be modeled are not directly observable, statistical sampling techniques for obtaining estimates of their means and covariances are not applicable. In this case proper mean and covariance models have to be determined from the only available source of information, the observations themselves. We say in this case that the stochastic model is "adapted" to the observational data and the resulting adaptive estimation techniques are the subject of this chapter.

5.2. Review of the Continuous-Discrete Kalman-Bucy Filter

The problem in question is the estimation of the state vector $X(t)$ of a dynamical system whose evolution in time is described by the (continuous) linear differential equation

$$\frac{d}{dt} X(t) = A(t)\, X(t) + G(t)\, u(t, \omega) \qquad (5.1)$$

from a finite number of observations (discrete observations)

$$y(t_i) = H_i\, X(t_i) + n(t_i, \omega) \qquad i=1,2,..,m \qquad (5.2)$$

$A(t)$, $G(t)$, H_i are known matrices, the forcing term $u(t, \omega)$ is a given vector stochastic process, and $n_i(\omega) = n(t_i, \omega)$ is a given sequence of random vectors.

The term filtering refers to the estimation of $X(t_m)$ and the term smoothing to the estimation of $X(t_i)$, $i=1,2,..,m-1$, from all the observations. The more general term prediction refers to the estimation of $X(t)$ for any epoch t, not necessarily coinciding with any of the observation epochs t_i.

If the random variables in $X(t_1, \omega)$, $n_1(\omega)$ are second-order with known means and covariances, the filtering-smoothing problem can be solved with the help of classical least squares adjustment techniques (condition equations) and with a minimum variance criterion for estimate optimality.

If a set of completely unknown parameters is involved, the generalized least squares method is to be used.

The prediction problem can be separately solved afterwards with the help of the estimates of $X(t_1)$ obtained, and the mean and covariance function of $X(t)$, as a minimum variance least squares prediction problem.

The problem has therefore been reduced to one of obtaining first and second order statistics for $X(t, \omega)$ from those of $u(t, \omega)$ and the initial value $X(t_o, \omega)$ (in case it is modeled as a vector of random variables and not as a vector of unknown parameters).

Usual assumptions are that $u(t)$, $X(t_o)$, n_1 are mutually independent and that n_1 is a sequence of Gaussian independent random vectors.

If $u(t)$ is a known stochastic process, so is $G(t)\,u(t)$ and we can drop the coefficient matrix $G(t)$ without any loss of generality. The state differential equation can now be formally integrated with the help of the state transition matrix $\Phi(t, s)$, to obtain

$$X(t, \omega) = \Phi(t, t_o)\; X(t_o, \omega) + \int_{t_o}^{t} \Phi(t, \xi)\; u(\xi, \omega)\; d\xi \tag{5.3}$$

Since $u(t, \omega)$ is a stochastic process the integral

$$y(t, \omega) = \int_{t_o}^{t} \Phi(t, \xi)\; u(\xi, \omega)\; d\xi \tag{5.4}$$

cannot be trivially defined as a Riemann integral in the general case. The most trivial case appears when $u(t, \omega)$ belongs to the restricted class of processes with Riemann integrable sample functions. We may relax the integrability condition for sample functions $u^{\omega}(t)$ with $\omega \in A \subset \Omega$, when $P(A) = 0$. In this case we say that almost all sample functions are Riemann integrable, or that $u(t, \omega)$ is Riemann integrable w.p.1 (with probability one). A more wide class of stochastic processes may be obtained by requiring that the above integral exists in the mean square sense (see [Jazwinski, 1970, p. 66] for definition). Mean square Riemann integrability of $u(t, \omega)$ is equivalent to Riemann integrability of its mean, correlation and covariance functions [Jazwinski, 1970, theorem 3.7 and corollary 1, pp. 66 & 67]. In this case the propagation from first and second order statistics of $u(t, \omega)$ to those of $y(t, \omega)$ can be carried out with the help of the (deterministic-ordinary) Riemann integrals.

$$E\{y(t, \omega)\} = \int_{t_o}^{t} \Phi(t, \xi) \; E\{u(\xi, \omega)\} \; d\xi \tag{5.5}$$

$$E\{y(t,\omega)\, y^{T}(s,\omega)\} = \int_{\xi=t_o}^{t} \int_{\zeta=t_o}^{s} \Phi(t,\xi) \; E\{u(\xi,\omega)\, u^{T}(\zeta,\omega)\} \; \Phi^{T}(s,\zeta) \; d\xi \; d\zeta \tag{5.6}$$

A stochastic process attracting considerable attention because of its applicability to engineering problems is the Gaussian white noise process, with zero mean, and covariance matrix

$$E\{u(t, \omega)\, u^{T}(s, \omega)\} = Q(t) \; \delta(t-s) \tag{5.7}$$

where $\delta(t-s)$ is the Dirac delta function. The white noise process has

everywhere discontinuous sample functions and fails to be Riemann mean square integrable. There is no shortcoming on the part of mathematical theory here, for white noise is not a physically realizable process, but only an idealization introduced mostly for the simplifying integration properties of the Dirac delta function. It is possible to formally represent $u(t, \omega)$ as the derivative of the Wiener process $W(t, \omega)$ [McGarty, 1974, p. 80]

$$u(t, \omega) = \frac{d}{dt} W(t, \omega) \tag{5.8}$$

In view of this relation one might attempt to define an integral of the form

$$\int_a^b G(t)\ u(t, \omega)\ dt$$

as a Stieltjes integral with the help of the Wiener process

$$\int_a^b G(t)\ dW(t, \omega)$$

This does not solve the problem because although almost all of the sample functions $W^{\omega}(t)$ are uniformly continuous, they are not of bounded variation [McGarty, 1974, Section 3.3]. Among various definitions of the above integral, the one more widely used is that due to K. Ito. We shall refer to [Jazwinski, 1970, Chapter 4] for details, restricting ourselves to the formal rule for obtaining the covariance of the process

$$y(t, \omega) = \int_{t_o}^{t} \Phi(t, \xi)\ dW(\xi, \omega) \quad : \tag{5.9}$$

$$E\{y(t,\omega)\ y^T(s,\omega)\} = \int_{t_o}^{\min(t,s)} \Phi(t,\xi)\ Q(\xi)\ \Phi^T(s,\xi)\ d\xi \qquad (5.10)$$

This rule agrees with the definition of $y(t, \omega)$ as an Ito stochastic integral and may also be formally derived from the covariance $Q(t)\ \delta(t-s)$ of the corresponding white noise process using the integration properties of the Dirac delta function.

We have succeeded in writing the solution of the state differential equation for $X(t)$ driven by white noise, in the form

$$X(t) = \Phi(t, t_o)\ X(t_o) + y(t) \qquad (5.11)$$

From the known first and second order statistics of $X(t_o)$ and $y(t)$, those of $X(t)$ can be easily derived and used to solve the filtering-smoothing and prediction problem, either globally by standard least squares adjustment and prediction techniques or, most often, by means of a sequential reformulation of the solution algorithm. The solution is similar when $u(t, \omega)$ is a process with Riemann integrable sample functions or simply a Riemann mean square integrable process.

5.3 Modeling of Stochastic Processes

In real life problems, when the state of a dynamical system (such as the orbit of an artificial satellite or the rotation of the earth) is generated by a differential equation, the forcing term $u(t)$ is in most cases an unknown function. The modeling of $u(t)$ as a stochastic process $u(t, \omega)$ presupposes that although $u(t)$ is not precisely known, it is not completely unknown either. This situation of relative uncertainty about $u(t)$ is manifested in the probability distribution of the stochastic process model $u(t,\omega)$ which roughly tells us what $u(t)$ is "likely" to be, by means of a description of

its "average" behaviour.

Since we are not concerned with an ensemble of functions $\{u_i(t)\}$ but rather with a unique function $u(t)$, the averaging has to take place over different parts of the function $u(t)$ itself. This presupposes that the function possesses some regularity properties, such that different pieces (i.e., the function over different time intervals) can be considered as samples with the same "statistical" behaviour, and their comparison yields the "average" behaviour of the function. These intuitive concepts are expressed within the probability theory framework by the ergodicity property of the stochastic process $u(t,\omega)$ serving as a model for the real (and in a sense deterministic) function $u(t)$. Ergodicity presupposes stationarity, and before such a more or less restrictive assumption is imposed, one should make certain that the process is irreducible, in the sense that it cannot be expressed as a transformation of some other original process. This simply means that a stationary - ergodic model should be used for unmodeled accelerations and the excitation function, rather than the orbit of the satellite or the rotation of the earth.

If the function $u(t)$ can be directly observed, the problem is a statistical one of determining distributions (means and covariances are sufficient for our purpose) from samples. However, we are primarily concerned here with the case when the function $u(t)$ cannot be directly observed, or it is practically impossible to do so, a situation common in geodetic work. In this case one can only construct an empirical stochastic model with the help of available mathematical tools. Before constructing such a model one must first wonder whether the necessary tools are available in the first place. To ask the question in a different way, suppose that a stochastic process $u(t,\omega)$ exists which is an appropriate model for the unknown function. What are our chances of empirically arriving at such a model ?

Indeed, the mathematical literature is almost exclusively devoted to the study of a number of elementary processes (white noise, Wiener,

Poisson) and processes derived from transformations of such processes. The answer is to be found in connection with the inverse problem, already studied by mathematicians: Given an arbitrary stochastic process, is it possible to express it in terms of some simple elementary processes ? This is a problem of representation that has been given considerable attention. A collection of papers on this subject can be found in [Ephremides and Thomas, 1973].

For a wide class of stochastic processes called purely nondeterministic or linearly regural (see [Cramer, 1964, p. 170; or Cramer, 1971, p. 7] for definition) it is possible to obtain a representation (Cramer-Hida canonical representation) of the form

$$X(t,\omega) = \sum_{n=1}^{N} \int_{-\infty}^{t} g_n(t,s)\ dz_n(s,\omega) \qquad (5.12)$$

where $z_n(s,\omega)$ are N mutually independent stochastic processes with orthogonal increments. In simple engineering terms, X(t) can be considered as the sum of the outputs of N linear filters with inputs white noise processes $\frac{dz_n}{dt}$. The smallest number N for which such a representation exists is called the (spectral) multiplicity of the process X(t). The class of second-order stationary processes we are concerned with here has spectral multiplicity one [Cramer, 1965, p. 218]. This ascertains the possibility of constructing a wide class of stationary models with the help of the white noise process.

Another aspect of modeling is the necessity of parametrization of the mean and covariance functions of the empirical model. The optimal values of such parameters must be determined from the observations available, since the non-observability of the function modeled excludes the use of sampling techniques.

Empirical models are much in use in time series analysis techniques,

when a function $u(t)$ is observed at equidistant epochs t_i ($t_i - t_{i-1} = \Delta t$ for all i). If $u(t)$ is modeled as a stochastic process $u(t,\omega)$, the sequence of random variables $u_i(\omega) = u(t_i,\omega)$ is a new discrete parameter stochastic process. Three usually used models are constructed with the help of a discrete parameter white noise process, i.e., a sequence of independent identically distributed random variables n_i with

$$E\{n_i\} = 0 \qquad \text{and} \qquad E\{n_i\, n_j\} = \sigma^2\, \delta_{ij}$$

These are [Koopmans, 1974, Chapter 7] the moving average model:

$$u_k = \sum_{j=-m}^{n} a_j\, n_{k-j} \tag{5.13}$$

the autoregressive model:

$$u_k + \sum_{j=1}^{q} b_j\, u_{k-j} = n_k, \qquad E\{u_l\, n_k\} = 0 \quad \text{for } l<k, \tag{5.14}$$

and the mixed autoregressive - moving average model:

$$\sum_{j=0}^{q} d_j\, u_{k-j} = \sum_{i=0}^{p} c_i\, n_{k-i} \qquad (\text{usually } d_0 = c_0 = 1). \tag{5.15}$$

These are finite parameter models and the first two can be extended to infinite by letting $n, m, q \to \infty$ and introducing the conditions

$$\sum_{-\infty}^{+\infty} a_j^2 < \infty, \qquad \sum_{1}^{+\infty} b_j^2 < \infty$$

Of particular importance is the one-sided moving average model

$$u_k = \sum_{j=0}^{\infty} a_j \; n_{k-j} \tag{5.16}$$

which possesses the property of physical casuality since the state u_k of the process at epoch t_k depends only on past values n_{k-j} of the exciting white noise process, and not on future ones. It can be shown that the solution to a finite autoregressive model is an one-sided (infinite) moving average [Koopmans, 1974, Section 7.3].

A purely nondeterministic second-order stationary stochastic process $X(t)$ satisfying certain conditions (see [Cramer, 1965, p. 219] for details) has multiplicity one, and canonical representation of the form

$$X(t) = \int_{-\infty}^{t} g(t-s) \; dz(s) \tag{5.17}$$

An approximation of the above integral by a summation leads to

$$X_k = X(t_k) = \sum_{i=-\infty}^{k} g(t_k - t_i) \; [\, z(t_i) - z(t_{i-1}) \,] = \sum_{i=-\infty}^{k} g_{k-i} \; n_i \tag{5.18}$$

We have set $z(t_i) - z(t_{i-1}) = n_i$, with $E\{ n_i \} = 0$ and $E\{ n_i n_j \} = \sigma^2 \delta_{ij}$, in view of $E\{ dz \} = 0$, $E\{ |dz(t)|^2 \} = dt$ $(z(t_i) - z(t_{i-1}) \approx dz)$ and the fact that $z(t)$ has orthogonal increments. A simple change of the summation index from i to $j = k - i$ gives

$$X_k = \sum_{j=0}^{\infty} g_j \; n_{k-j} \tag{5.19}$$

which is an one-sided moving average model.

Without any approximation it can be shown that a discrete parameter zero mean weakly stationary stochastic process can be expressed as the sum of a purely deterministic and a purely nondeterministic zero mean processes (Wold decomposition), the latter having an one-sided (infinite) moving average representation [Koopmans, 1974, p. 255]. This connection with the Cramer-Hida representation establishes the importance of the one-sided moving average and the autoregressive scheme (whose solution is an one-sided moving average) in modeling stationary stochastic processes.

Passing from the discrete to the continuous time case and restricting ourselves to the autoregressive model in view of its finite number of parameters, we have

$$\frac{d^n u(t)}{dt^n} + \sum_{j=0}^{n-1} b_j \frac{d^j u(t)}{dt^j} = g \cdot \frac{dW(t)}{dt} \qquad (5.20)$$

where $W(t)$ is the Wiener process. Generalizing to a vector process $u(t)$ and allowing the coefficient matrices to be functions of time we obtain the n^{th} order autoregressive model

$$\frac{d^n u(t)}{dt^n} + B_1(t) \frac{d^{n-1} u(t)}{dt^{n-1}} + \ldots + B_{n-1}(t) \frac{du(t)}{dt} + B_n(t)\, u(t) =$$

$$= G(t) \frac{d W(t)}{dt} \qquad (5.21)$$

The solution $u(t)$ of the above differential equation has multiplicity one and is a Gaussian Markov process [Ephremides and Thomas, 1973, p. 13]. We shall limit ourselves to the first order (n=1) autoregressive process and we shall investigate conditions for $u(t)$ to be stationary.

The use of the Wiener process instead of any other process with independent increments restricts us to Gaussian processes. The "Gaussianity" assumption is essential in making possible statistical inference on final Gaussian estimates, since means and covariances completely specify in that case the corresponding probability distributions.

Another important independent increment process is the Poisson (count) process and the generalized Poisson process [McGarty, 1974, pp. 83-86]. The formal derivative of the generalized Poisson process is also a white noise process, although non-Gaussian. For the study of Poisson driven Markov processes we refer to [Snyder, 1975, Section 4.2].

5.4 Construction of Exponentially Correlated Stationary Stochastic Processes from White Noise

The most simple continuous time autoregressive model is the first-order one-dimensional one with constant coefficients,

$$\frac{dx(t)}{dt} + \rho\, x(t) = \frac{dW(t)}{dt}, \qquad E\{\dot{W}(t)\,\dot{W}(s)\} = \sigma^2\,\delta(t-s) \tag{5.22}$$

where $\dot{W} = \frac{dW}{dt}$. This is a linear stochastic differential equation with state transition function

$$\Phi(t,s) = e^{-\rho(t-s)} \tag{5.23}$$

and solution

$$x(t) = e^{-\rho(t-t_o)} + \int_{t_o}^{t} e^{-\rho(t-\xi)}\, dW(\xi) \tag{5.24}$$

We can differentiate between two types of solution according to whether

$x(t_o)$ is a deterministic constant or a random variable:

A. For $x(t_o) = c_o = \text{const.}$, we have

$$m(t) = E\{x(t)\} = e^{-\rho(t-t_o)}\, c_o$$

$$R(t,s) = E\{x(t)\, x(s)\} = e^{-\rho(t-t_o)}\, c_o^2\, e^{-\rho(s-t_o)} +$$

$$+ \sigma^2 \int_{t_o}^{\min(t,s)} e^{-\rho(t-\xi)}\, e^{-\rho(s-\xi)}\, d\xi =$$

$$= \left[c_o^2 - \frac{\sigma^2}{2\rho}\right] e^{-\rho(t+s)}\, e^{2\rho t_o} + \frac{\sigma^2}{2\rho}\, e^{-\rho|t-s|} \tag{5.26}$$

The covariance function of $x(t)$ is

$$C(t,s) = R(t,s) - m(t)\, m(s) =$$

$$= -\frac{\sigma^2}{2\rho}\, e^{-\rho(t+s)}\, e^{2\rho t_o} + \frac{\sigma^2}{2\rho}\, e^{-\rho|t-s|} \tag{5.27}$$

and therefore $x(t)$ is nonstationary.

B. For $x(t_o) = x_o(\omega)$, with $E\{x_o\} = c_o$, $E\{x_o^2\} = \sigma_o^2$ and $E\{x_o\, \dot{W}(t)\} = 0$, we have

$$m(t) = e^{-\rho(t-t_o)}\, c_o \tag{5.28}$$

$$R(t,s) = \left[\sigma_o^2 - \frac{\sigma^2}{2\rho}\right] e^{-\rho(t+s)}\, e^{2\rho t_o} + \frac{\sigma^2}{2\rho}\, e^{-\rho|t-s|} \tag{5.29}$$

$$C(t,s) = \left[\sigma_o^2 - \frac{\sigma^2}{2\rho} - c_o^2\right] e^{-\rho(t+s)}\, e^{2\rho t_o} + \frac{\sigma^2}{2\rho}\, e^{-\rho|t-s|} \tag{5.30}$$

The solution $x(t)$ is in general a nonstationary process. It is possible to

obtain a stationary process by setting $c_o = 0$ (so that $E\{x(t)\} = 0 = \text{const.}$) and $\sigma_o^2 = \sigma^2 / 2\rho$ so that

$$C(t,s) = \frac{\sigma^2}{2\rho} e^{-\rho|t-s|} \tag{5.31}$$

We have succeeded in constructing a stationary stochastic process from white noise. With appropriate selection of the parameters σ^2, ρ we can obtain every such process from the class of exponentially correlated processes with covariance functions of the form

$$C(\tau) = C_o \exp[-|\tau| / T], \quad \tau = t-s, \quad T > 0 \tag{5.32}$$

where $C_o = C(0)$ is the variance of $x^t(\omega)$ for every t, and T is the "correlation time" related to the "sharpness" of $C(\tau)$. The smaller the correlation time T is, the more the covariance function is concentrated about the $C(\tau)$ axis.

To extend these results to more dimensions consider the vector stochastic process $X(t)$ generated by a general first order autoregressive model of the form

$$\frac{d}{dt} X(t) = A(t) X(t) + \frac{d}{dt} W(t) \tag{5.33}$$

with $E\{\dot{W}(t) \dot{W}^T(s)\} = Q(t) \delta(t-s)$, $E\{X(t_o)\} = X_o$,

$E\{X(t_o) X^T(t_o)\} = R_o$, $E\{W(t) X^T(t_o)\} = 0$

The solution is

$$X(t) = \Phi(t,t_o) X(t_o) + \int_{t_o}^{t} \Phi(t,\xi) \, dW(\xi) \tag{5.34}$$

$$m(t) = E\{ X(t) \} = \Phi(t,t_o)\ X_o \tag{5.35}$$

and covariance function

$$C(t,s) = \Phi(t,t_o)[\ R_o - X_o\ X_o^T\ ]\ \Phi^T(s,t_o) + \int_{t_o}^{\min(t,s)} \Phi(t,\xi)\ Q(\xi)\ \Phi^T(s,\xi)\ d\xi \tag{5.36}$$

The necessary and sufficient conditions for stationarity are $m(t) = \text{const.}$ (implying $X_o = 0$), and $C(t,s) = C(t-s)$. It can be shown (see for example [Bucy and Joseph, 1968, p. 25]) that $K(t) = C(t,t)$ is the solution of the differential equation

$$\frac{d}{dt}\ K(t) = A(t)\ K(t) + K(t)\ A^T(t) + Q(t) \tag{5.37}$$

with initial condition $K(t_o) = C_o = R_o - X_o\ X_o^T$

In the stationary case $K(t) = K(0) = K(t_o) = R_o$ is constant and therefore

$$\frac{d}{dt}\ R_o = 0 = A(t)\ R_o + R_o\ A^T(t) + Q(t) \tag{5.38}$$

This is a necessary but not sufficient condition for stationarity. We can obtain sufficient conditions (without claims to necessity) by setting

$A(t) = A = \text{const.}$, $Q(t) = Q = \text{const.}$, $X_o = 0$, and

$$A\ R_o + R_o\ A^T + Q = 0 \tag{5.39}$$

This is equivalent to the condition $2\rho\ \sigma_o^2 = \sigma^2$ in the one-dimensional case. An additional condition corresponding to $\rho > 0$ is that the eigenvalues of A must have negative real parts (see [Arnold, 1974, p. 133]). If $X(t_o)$ is also multivariate Gaussian, $X(t)$ is a Gaussian process.

5.5 The Dynamic Model Compensation (DMC) Algorithm

Ingram and Tapley [1974] have deviced a method for the estimation of the state of a spacecraft in the presence of unmodeled accelerations which they call the "Dynamic Model Compensation" (DMC) algorithm. With the help of material from [Ingram, 1970, Tapley, 1973; and Ingram and Tapley, 1974] we shall give a somewhat generalized version of this technique. One part of our generalization is the consideration of an unspecified dynamical system with state governed by a linear differential equation

$$\frac{d}{dt} x(t) = A(t)\, x(t) + u(t) \tag{5.40}$$

in place of the specific equations of spacecraft motion. $u(t)$ is a vector stochastic process serving as a model for unknown (unmodeled) forcing terms, and generated by another first order differential equation

$$\frac{d}{dt} u(t) = B(\rho)\, u(t) + n(t) \tag{5.41}$$

where the coefficient matrix $B(\rho)$ depends on a vector ρ of unknown parameters. The unspecified nature of the dependence of B on ρ is the second aspect of our generalization. Ingram and Tapley consider a diagonal matrix B, and ρ has elements the negative inverses of the diagonal elements of B. $n(t)$ is a vector white noise process with $E\{n(t)\, n^T(s)\} = G(t)\, \delta(t-s)$, serving as a formal representation in place of the mathematically precise Ito stochastic differential equation

$$du(t) = B(\rho)\, u(t) + dW(t) \tag{5.42}$$

The lack of a coefficient matrix in fron of $n(t)$ poses no restriction since the corresponding effect can be trivially incorporated in the covariance matrix

G(t) by means of a simple propagation. The constant vector ρ can be also trivially modeled by the differential equation

$$\frac{d}{dt}\,\rho = 0 \tag{5.43}$$

Summarizing, and after some obvious change to a more compact notation, our state-model becomes

$$\dot{x}_t = A_t\, x_t + u_t \tag{5.44a}$$

$$\dot{u}_t = B(\rho)\, u_t + n_t \tag{5.44b}$$

$$\dot{\rho}_t = 0 \tag{5.44c}$$

Integrating with the help of initial values $\rho(t_1) = \rho_1$, $u(t_1) = u_1$, $x(t_1) = x_1$, we obtain

$$\rho_t = \rho_1 \tag{5.45a}$$

$$u_t = \Phi_u(t,t_1)_{(\rho_1)}\, u_1 + \int_{t_1}^{t} \Phi_u(t,\xi)_{(\rho_1)}\, n_\xi\, d\xi =$$

$$= \Phi_u(t,t_1)_{(\rho_1)}\, u_1 + l_1(t,t_1,\rho_1) \tag{5.45b}$$

$$x_t = \Phi_x(t,t_1)\, x_1 + \int_{t_1}^{t} \Phi_x(t,\xi)\, u_\xi\, d\xi =$$

$$= \Phi_x(t,t_1)\, x_1 + \Phi_{xu}(t,t_1)_{(\rho_1)}\, u_1 + \int_{t_1}^{t} \Phi_x(t,\xi)\, l_1(\xi)\, d\xi \tag{5.45c}$$

where

$$\Phi_{xu}(t,t_1) = \int_{t_1}^{t} \Phi_x(t,\xi) \quad \Phi_u(\xi,t_1) \quad d\xi \tag{5.46}$$

Introducing

$$\theta_x(t,t_1,x_1,u_1,\rho_1) = \Phi_x(t,t_1)\ x_1 + \Phi_{xu}(t,t_1,\rho_1)\ u_1 \tag{5.47a}$$

$$\theta_u(t,t_1,u_1,\rho_1) = \Phi_u(t,t_1)_{(\rho_1)}\ u_1 \tag{5.47b}$$

$$\theta_\rho(\rho_1) = \rho_1 \tag{5.47c}$$

and

$$\eta_{\rho 1} = 0 \tag{5.48a}$$

$$\eta_{u1}(t,t_1,\rho_1) = 1_1(t,t_1,\rho_1) \tag{5.48b}$$

$$\eta_{x1}(t,t_1,\rho_1) = \int_{t_1}^{t} \Phi_x(t,\xi)\ 1_1(\xi,t_1,\rho_1)\ d\xi \tag{5.48c}$$

we arrive at

$$X_t = \theta(t,t_1,X_1) + \eta_1(t,t_1,X_1) \tag{5.49}$$

where

$$X_t^T = [\ x_t^T\ u_t^T\ \rho_t^T\ ],\quad \theta_t = [\ \theta_x^T\ \theta_u^T\ \theta_\rho^T\ ],\quad \eta_1^T = [\ \eta_{x1}^T\ \eta_{u1}^T\ 0\ ]$$

For $t = t_{i+1}$ we have

$$X_{i+1} = \theta(t_{i+1},t_1,X_1) + \eta_1(t_{i+1},t_1,X_1) \tag{5.50}$$

This is a discrete state model and can be used together with a set of

observations (assuming linear dependence on the state for simplicity)

$$y_i = H_i X_i + v_i, \quad E\{v_i\} = 0, \quad E\{v_i v_j^T\} = R_i \delta_{ij} \tag{5.51}$$

to solve the prediction-smoothing problem. However, there are some difficulties associated with this model. One of them is the nonlinearity of θ with respect to ρ_i, which we can overcome by linearization. The second is the dependence of the state noise η_i on the state X_i (actually only on ρ_i), and the final one is the non-whiteness of the sequence $\{\eta_i\}$. If the values $\{\rho_i\}$ are considered to be constants, it can be easily seen that

$$E\{\eta_i\} = 0 \qquad \text{and} \qquad E\{\eta_i \eta_j^T\} = 0 \quad \text{for } i \neq j \tag{5.52}$$

However, ρ_i is part of the state X_i and as such it is a random vector. In view of $\rho_{i+1} = \rho_i$, we have

$$E\{\rho_i\} = E\{\rho_0\} \qquad \text{and} \qquad E\{\rho_i \rho_j^T\} = E\{\rho_0 \rho_0^T\} \neq 0 \tag{5.53}$$

Because of the complex dependence of η_i on ρ_i it is difficult to evaluate $E\{\eta_i \eta_j^T\}$, but the two vectors are in general correlated since ρ_i and ρ_j are. To overcome these difficulties the sequence $\{\eta_i\}$ is approximated by a new sequence $\{\hat{\eta}_i\}$, which does not depend on the state (ρ_i) and is also white ($E\{\hat{\eta}_i \hat{\eta}_j^T\} = 0$ for $i \neq j$).

The approximation, which is essential to the resulting algorithm, is to replace $l_i(t_{i+1}, t_i, \rho_i)$ by a random vector $\hat{l}_i$ such that $E\{\hat{l}_i\} = 0$ and $E\{\hat{l}_i \hat{l}_j^T\} = E\{l_i l_j^T\} \delta_{ij}$. This we can achieve with the help of a white sequence n_i, such that $E\{n_i\} = 0$ and $E\{n_i n_j^T\} = I \delta_{ij}$, by setting

$$\hat{l}_i = K_i^{\frac{1}{2}} n_i \tag{5.54}$$

where $K_i = E\{l_i l_j^T\}$, and the expectation is taken by treating ρ_i as a nonrandom constant,

$$K_i = \int_{t_i}^{t_{i+1}} \Phi_u(t_{i+1}, \xi)\ G(\xi)\ \Phi_u^T(t_{i+1}, \xi)\ d\xi = K_i(t_{i+1}, t_i, \rho_i) \qquad (5.55)$$

K_i depends on ρ_i since Φ_u depends on ρ_i. It follows easily that

$$E\{\hat{l}_i \hat{l}_j^T\} = K_i^{\frac{1}{2}}\, E\{n_i n_j^T\}\, (K_i^{\frac{1}{2}})^T = K_i^{\frac{1}{2}} (K_i^{\frac{1}{2}})^T \cdot \delta_{ij} = K_i\, \delta_{ij} \qquad (5.56)$$

By setting

$$\hat{\eta}_{ui} = \hat{l}_i, \qquad \hat{\eta}_{xi} = \left[\int_{t_i}^{t_{i+1}} \Phi_x(t_{i+1}, \xi)\ d\xi \right] \hat{l}_i = S_i\, \hat{l}_i$$

and $\hat{\eta}_i^T = [\ \hat{\eta}_{xi}^T\ \ \hat{\eta}_{ui}^T\ \ 0\ ]$, we have

$$E\{\hat{\eta}_i\} = 0 \quad \text{and} \quad E\{\hat{\eta}_i \hat{\eta}_j^T\} = Q_i\, \delta_{ij} \qquad (5.57)$$

where

$$Q_i = \begin{bmatrix} S_i K_i S_i^T & S_i K_i & 0 \\ K_i S_i^T & K_i & 0 \\ 0 & 0 & 0 \end{bmatrix} = Q_i(\rho_i) \qquad (5.58)$$

We have now arrived at the state model

$$X_{i+1} = \theta(t_{i+1}, t_i, X_i) + \hat{\eta}_i \qquad (5.59)$$

such that the smoothing-prediction problem can be solved with the help of standard techniques (extended Kalman-Bucy filter, see [Jazwinski, 1978, p. 278; and Tapley, 1973, p. 411]), except for the dependence of the state noise covariance matrix $Q_i(\rho_i)$ on part of the state. Ingram and Tapley simply write $Q_i(\rho)$ without specifying which value of ρ is used (T_x, T_y, T_z in their notation). It is natural to assume that ρ_i is replaced

by its current estimate ρ_i based upon all past observations y_r $(r=1,2,..,i)$.

The central point in the derivation of the algorithm is the approximation of l_i by $\tilde{l}_i$. Ingram [1970] offers no explanation or motivation about this approximation. Further reference is given to [Jordan, 1966] (to which we had no access with the remark [Ingram, 1970, p. 22] that the original white noise process $n(t)$ is considered constant over the interval $[t_i, t_{i+1}]$. This, at least mathematically, is absurd. The very essence of the white noise process is the everywhere discontinuity of its sample functions, and $n(\xi) - n(\zeta)$, $\xi, \zeta \in [t_i, t_{i+1}]$, is unbounded no matter how small $|t_{i+1}-t_i|$ is.

It must be pointed out that the same white sequence $\hat{\eta}_i$ may be obtained by a much simpler argument. The random vector $\rho(t)$ is approximated over the interval $[t_i, t_{i+1}]$ by a constant nonrandom vector equal to its current estimate $\hat{\rho}_i$ based upon past observations up to the epoch t_i. In this case we replace $l_i(t_{i+1}, t_i, \rho_i)$ by $\hat{l}_i(t_{i+1}, t_i, \hat{\rho}_i)$, so that $E\{\hat{l}_i \hat{l}_i^T\} = K_i(\rho_i)$, and proceed to obtain the white sequence $\hat{\eta}_i$ with $E\{\hat{\eta}_i \hat{\eta}_i^T\} = Q_i(\hat{\rho}_i)$.

Assuming for a moment that despite the approximations involved, the final solution (after smoothing where state noise covariance matrices are also updated) is sufficiently close to the "true" solution of the original model

$$\dot{x}_t = A_t x_t + u_t \tag{5.44a}$$

$$\dot{u}_t = B(\rho) u_t + n_t \tag{5.44b}$$

$$\dot{\rho}_t = 0 \tag{5.44c}$$

we may identify the problem with the one described by

$$\dot{x}_t = A_t x_t + u_t \tag{5.60a}$$

$$\dot{u}_t = B(\rho^*) u_t + n_t \tag{5.60b}$$

where ρ^* is the final optimal estimate of ρ based upon all the observations.

In this case we can directly model u_t as a stochastic process with known mean and covariance function, specified by the solution of eq. (5.60b) and the initial value $u_0 = u(t_0)$.

For the specific choice of

$$B(\rho) = \begin{bmatrix} -\rho_x^{-1} & 0 & 0 \\ 0 & -\rho_y^{-1} & 0 \\ 0 & 0 & -\rho_z^{-1} \end{bmatrix} \qquad E\{n_t n_s^T\} = \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{bmatrix} \delta(t-s)$$

each component of u_t is an independent stochastic process with mean

$$E\{u_x(t)\} = e^{-(t-t_0)\rho_x^{-1}} \quad E\{u_x(t_0)\} \tag{5.61}$$

and covariance

$$C(t,s) = \left[E\{u_x^2(t_0)\} - \tfrac{1}{2}\sigma_x^2 \rho_x\right] \quad e^{-(t+s)\rho_x^{-1}} \quad e^{2t_0\rho_x^{-1}} \quad +$$

$$+ \tfrac{1}{2}\sigma_x^2 \rho_x \quad e^{-|t-s|\rho_x^{-1}} \quad +$$

$$- e^{-(t+s)\rho_x^{-1}} \quad e^{2t_0\rho_x^{-1}} \quad \left(E\{u_x(t_0)\}\right)^2 \tag{5.62}$$

with similar expressions for $u_y(t)$ and $u_z(t)$.

Ingram [1970, p. 22] claims stationarity of $u_x(t)$, recognizing the necessity of ρ_x being a constant, but not the special relation between ρ_x and the initial value $u_x(t_0)$ statistics also necessary for stationarity (see Section 5.3). The initial values are modeled as random variables with $E\{u(t_0)\} = u_0$, $E\{u(t_0)\,u^T(t_0)\} - u_0 u_0^T = \sigma_0^2 I$ [Ingram, 1970, pp.76-77]. If $u_0 \neq 0$, then $E\{u_t\} \neq$ const., and u_t is certainly nonstationary. If $u_0 = 0$, then, to secure stationarity one must set

$$\rho_x = \frac{2\sigma_{0x}^2}{\sigma_x^2} \qquad \text{where} \quad \sigma_{0x}^2 = E\{u_x^2(t_0)\} \tag{5.63}$$

in which case ρ_x should be a fixed constant and not a parameter under estimation. Ingram and Tapley [1974, p. 195] mention that the initial conditions are unknown. If this implies that u_0 is a deterministic constant to be estimated simultaneously, we again arrive to a nonstationary solution in general (see Section 5.3).

In view of the reported success of the DMC algorithm, its adaptivity must be looked for in the approximations themselves rather, than in the original rigorous state equations. If ρ has been fixed, the state noise matrix $Q(t_{k+1}, t_k, \rho)$ should have been a priori defined. On the other hand, the state noise matrix $Q(t_{k+1}, t_k, \hat{\rho}_k)$ is allowed to vary in the filtering algorithm, according to the current estimate $\hat{\rho}_k$ of the parameter ρ. Although $\rho_k = \rho_{k-1}$ in the discrete state model, the same equality does not hold for the corresponding estimates

$$\hat{\rho}_k \neq \hat{\rho}_{k-1} \tag{5.64}$$

where $\hat{\rho}_i$ is the estimate of ρ based upon past observations y_j up to the present epoch ($j=1, 2, \ldots, i$), and the one additional observation y_k produces in general a change on the new estimate $\hat{\rho}_k$ of ρ.

The concept of an exponentially correlated forcing term can still be maintained within each interval $[t_i, t_{i+1}]$ where ρ_i is treated as a constant and corresponds to a "local covariance function" formulation designed to fit the local behavior of the unknown disturbing function. On the whole time interval $[t_1, t_n]$, however, the variation of the estimates $\hat{\rho}_i$ of ρ used, gives rise to a nonstationary stochastic process $u(t)$.

5.6 Adaptive Estimation in General

The general adaptive estimation problem may be defined as follows: The evolution in time of the state x_t of a dynamical system is described by a differential equation of the form

$$x_t = f(x_t, t, \epsilon_t) \tag{5.65}$$

where ϵ_t represents a function of unknown disturbances. It is possible to model ϵ_t, not as a single stochastic process u_t, but as a whole class of stochastic processes generated with the help of some other known process n_t and a set of unspecified parameters p,

$$\dot{u}_t = g(u_t, t, n_t, p) \tag{5.66}$$

Every set of parameters p specifies a different stochastic process u_t, and an optimal set p* is sought, such that the resulting stochastic process u_t^* is an optimal model (in a sense that remains to be specified) of the unknown disturbances ϵ_t, for the purpose of estimating the state x_t (and possibly another set of parameters a) from a finite set of observations

$$y_i = h_i(x_{t_i}, t_i, a) + v_i \tag{5.67}$$

where v_i is a sequence of zero mean random vectors with known covariances.

It is essential to realize that the differential equation (5.66) does not model any physically realizable dynamical system, but merely serves as an artificial means for constructing a class of stochastic processes $u_t(p)$. The nature of this class must be a priori specified (by means of selecting a certai process n_t and a certain function $g(u_t, t, n_t, p)$), and its appropriateness must be justified by considerations related to the nature of the unknown disturbances ϵ_t to be modeled.

Since p is a vector of constants, we have $\dot{p} = 0$, and through a technique called "vector augmentation" we may rewrite our model as

$$\dot{X}_t = \begin{bmatrix} \dot{x}_t \\ \dot{u}_t \\ \dot{p}_t \end{bmatrix} = \begin{bmatrix} f(x_t, t, u_t) \\ g(u_t, t, n_t, p) \\ 0 \end{bmatrix} = F(X_t, t, n_t) \tag{5.68}$$

The solution to the above stochastic differential equation is by no means

trivial because of the nonlinear dependence of $F(X_t, t, n_t)$ on X_t and n_t. The simplest linear model correspond to

$$\underline{f}(x_t, t, n_t) = A_t\ x_t + u_t \tag{5.69a}$$

$$g(u_t, t, n_t, p) = B_t\ u_t + D_t\ p + n_t . \tag{5.69b}$$

In this case the solution u_t is of the form

$$u_t = \Phi_u(t, t_o)\ u_o + \int_{t_o}^{t} \Phi_u(t, \xi)\ n_\xi\ d\xi + \left[\int_{t_o}^{t} \Phi_u(t, \xi)\ D_\xi\ d\xi \right] p =$$

$$= u_{1t} + \tilde{D}_t\ p = u_{1t} + u_{2t} \tag{5.70}$$

u_t is now the sum of a known stochastic process u_{1t} and a parametrized deterministic function $u_{2t}(p)$ which does not contribute to the covariance of u_t. It is possible to include the known mean of u_{1t} in u_{2t}, so that u_t is modeled as a stochastic process with known covariance and parametrized mean. This is a well known technique (see for example [Parzen, 1961]), but deviates from our main objective which is the adaptivity (parametrization) of the covariance function also.

The next simpler model is similar to the one in the DMC algorithm, obtained by setting

$$g(u_t, t, n_t, p) = B(p)\ u_t + n_t \tag{5.71}$$

the only nonlinearity being with respect to p.

Returning to the general nonlinear stochastic differential equation (5.68), let us assume that a solution has been somehow obtained. Such a solution is a stochastic process X_t with mean and covariance depending on those of n_t and the initial values X_o (x_o, u_o, p_o) modeled as random variables. Assuming for simplicity linear observations and no parameters,

we have

$$y_i = H_i x_i + v_i , \qquad i=1, 2, \ldots, n \tag{5.72}$$

Let X denote the vector of all states x_i,

$$X^T = [\, x_1^T \;\; x_2^T \;\; \ldots \;\; x_n^T \,]$$

U that of u_i, and $P = p_0$, since all states p_i are the same. From the known mean and covariance of the solution process X_t it is possible to obtain the mean and covariance of the vector

$$Z^T = [\, X^T \;\; U^T \;\; P^T \,],$$

$$E\{Z\} = \begin{bmatrix} \bar{X} \\ \bar{U} \\ \bar{P} \end{bmatrix} = \begin{bmatrix} X - \delta X \\ U - \delta U \\ P - \delta P \end{bmatrix}, \quad E\{ZZ^T\} - Z Z^T = \begin{bmatrix} S_{XX} & S_{XU} & S_{XP} \\ S_{UX} & S_{UU} & S_{UP} \\ S_{PX} & S_{PU} & S_{PP} \end{bmatrix}$$

The observation equations can be written in the following matrix form:

$$Y = H X + V \quad \text{or} \quad L = Y - H \bar{X} = [\, H \;\; I \,] \begin{bmatrix} \delta X \\ V \end{bmatrix}$$

Solving this least squares adjustment (condition equations) problem we obtain an estimate $\delta\hat{X}$ and $\hat{X} = \bar{X} + \delta\hat{X}$. The effect of the observations on the nonobservable parts of the state U and P can be evaluated through least squares prediction,

$$\delta\hat{P} = S_{PX} S_{XX}^{-1} \delta\hat{X}, \qquad \hat{P} = \bar{P} + \delta\hat{P}$$

and similarly for U. $\hat{X}$, the final estimate of the state at observation epochs does not depend at all on the final estimate $\hat{p}$ of p, but only on the mean and covariance function of the solution process X_t and consequently on the mean and covariance of the initial value p_0. In fact there is no adaptivity of the

mean and covariance function of the state noise process u_t in the vector augmentation technique. The solution differs from one with fixed parameters p, only in the fact that the uncertainty in an original estimate p_o of p has been taken into account.

Adaptivity can be introduced by resolving the problem, fixing p to the value of the obtained estimate $\hat{p}$, evaluating mean and covariance function of u_t, and using only the first part of the state equations

$$\dot{x}_t = f(x_t, t, u_t)$$

The optimality of the estimate $\hat{p}$ is hard to define. This estimate depends on the observations (through $\hat{X}$ being a function of Y), and to this extend is adapted to the observational data, but it also largely depends on the mean and covariance of the initial estimate p_o. Because of this, the vector augmentation technique fails to provide a solution to the following more idealized problem:
Given the mean and covariance of the state noise u_t within a set of unknown parameters p, find among all permissible values of p the one which is optimal in a sence to be defined.

The solution x_t of the state differential equation has mean and covariance function depending on the unspecified parameters p. The vector X corresponding to states at observational epochs has mean $\bar{X}(p)$ and covariance matrix S(p), both depending on p. For linear observations the model for the adjustment is

$$L(p) = Y - H\,\bar{X}(p) = H\,\delta X + V, \qquad E\{\delta X\,\delta X^T\} = S(p)$$

and the final estimate $\hat{X}(p) = \bar{X}(p) + \delta\hat{X}(p)$ depends also on p.
A simple way to define optimality of $\hat{X}(p)$ is the minimization of the variance of any scalar linear function of X

$$q = b^T X$$

i.e., to find the optimal vector p which minimizes

$$\sigma_q^2(p) = b^T [E\{X(p) X^T(p)\} - X(p) X^T(p)] b =$$

$$= b^T E\{\delta\hat{X}(p) \delta\hat{X}^T(p)\} b .$$

Such a criterion however, can be justified only when the variance to be minimized is the "true" variance based upon "true" a priori statistics of both the observational errors and the signal δX. The result is the minimization of the nominal rather than the true variances of estimates and the appropriateness of such an optimality criterion is questionable.

A more reasonable criterion of optimality is to be found through intuitive reasoning, rather than in any rigorous definition. In essence, the role of estimation is to separate the useful signal related to the states of the dynamical system from the unwanted observational noise. If such a separation has not taken place because of the use of an incorrect model, the effect is to be seen on the estimates of the observational residuals v_i. If the residuals are too small, too much signal has been taken out. If they are too large, or they just show some systematic pattern, then part of the signal has not been detected. In general, the inconsistency of residuals with their a priori statistics strongly indicates the use of an inappropriate model. This effect manifests itself during computations associated with filtering in what is commonly called "filter divergence" [Jazwinski, 1970, p. 302].

In a global (nonsequential) solution, the parameters p determining the statistics of the unknown disturbing function u_t must be varied by "trial and error" until consistency of residuals with their a priori statistics is reached. This presupposes that first, the parameters p give rise to a class of statistics for the process u_t which is wide enough, so as to contain an element close to the "true" statistics. The second assumption is that the a priori statistics of the observational errors are accurately known through calibration and standard statistical techniques.

Related to global adaptivity are techniques, such as the DMC algorithm, of local or real time adaptivity. These techniques aim at modifying the state noise covariance matrix during each step of the filtering process, in order to overcome the filter divergence problem. Such techniques are discussed in [Jazwinski, 1970, Chapter 8, especially Section 11; and Gelb, 1974, Sections 8.1 and 9.1].

Jazwinski [1969] considers a discrete linear dynamical system model

$$x_{k+1} = \Phi_{k+1}\, x_k + G_k\, w_k \qquad (5.73)$$

with observations

$$y_k = H_k\, x_k + v_k \qquad (5.74)$$

where the state noise w_k is a zero mean white Gaussian sequence, accounting for errors made in modeling the dynamics of the system. The state noise covariance matrices Q_k are considered to be the same over every N observations

$$Q_{k+i} = Q_{k,N}, \quad i=1, 2, \ldots, N$$

and an algorithm is devised for determining $Q_{k,N}$ so that the produced residuals are consistent with their a priori statistics.

Another approach very much similar to the DMC algorithm, is given in [Jazwinski, 1974]. The adopted system model is

$$\dot{x}_t = f(x_t,t) + G\, u_t \qquad (5.75)$$

and u_t is modeled over each interval $[t_i, t_{i+1}]$ between observations as

$$u_t = u_i + \beta_i\,(t-t_i) \qquad (5.76)$$

where β_i is a sequence of random variables with fixed a priori uncertainty.

The discrete state equations of the form

$$x_{i+1} = F(x_i, u_t) \qquad t_i \le t \le t_{i+1} \tag{5.77}$$

are augmented by

$$u_{i+1} = u_i + \beta_i \ (t_{i+1} - t_i) \tag{5.78a}$$

$$\beta_{i+1} = \beta_i \qquad (\beta_t = 0 \quad \text{for} \quad t_i \le t \le t_{i+1}) \tag{5.78b}$$

and the problem is solved with the help of the extended Kalman filter algorithm.

6. SUMMARY AND RECOMMENDATIONS

The objective of this work has been twofold: First, to clarify the mathematical and probabilistic background of standard linear estimation techniques used in geodesy and to reveal their interrelationship. Secondly, to address what we considered to be the two most important estimation problems in geodesy: the norm choice problem in gravimetric collocation, and the adaptive determination of the stochastic models for not directly observable physical processes.

Linear least squares adjustment and linear least squares prediction have been shown to reduce to linear best approximation problems. In the former, the observations are best approximated from elements of the "model space," while in the latter, the unknown parameters are best approximated from elements of the "data space." Probabilistic concepts in least squares adjustment have been shown to refer only to the definition of the metric for the approximation, while in least squares prediction the structure of the approximation space itself (space of second-order random variables) in addition relies on probabilistic ideas. However, for the case of Hilbert space valued random variables, least squares prediction has been related to deterministic (exact) collocation. Estimate optimality criteria (minimum error bounds) have been identified for the single parameter prediction as opposed to the global minimum norm solution for the unknown function.

Kalman-Bucy filtering techniques have been shown to reduce in the case of discrete observations to least squares adjustment (for filtering-smoothing) and least squares prediction (for prediction).

The use of stochastic models has been investigated for the two most important physical processes related to geodetic work: the gravity field and the rotation of the earth. Motivated by Lauritzen's proof of nonergodicity

for the gravity field of the earth, a deterministic criterion for optimality of related predictions has been introduced leading to a proposed solution of the optimal norm choice problem in (exact) collocation. In the case of the rotation of the earth, the relation of stochastic models for polar motion and diurnal rotation to those for the excitation function has been shown, through the use of the linearized Liouville equation for both cases: with and without damping.

The possibility of constructing simple stochastic models from white noise has been explored, and the conditions for stationarity of the output of first-order autoregressive models, excited by white noise, have been established.

The Dynamic Model Compensation algorithm has been generalized, and its adaptive structure and inherent approximations have been clarified. Other possibilities for adaptively estimating both state and unknown effects on dynamical systems and the nature of the optimality of such estimates have been explored.

Of course a great deal has yet to be done both with respect to gravimetric problems and to the adaptive estimation in geometric-dynamic geodesy.

The computational feasibility of obtaining an optimal model covariance or norm, using minimum error-bound criteria, remains to be demonstrated. A way must be found to include the effect of observational errors.

Adaptive estimation techniques, because of their vary nature (adaptivity to observational evidence), can only be compared and justified in connection with real problems. The success of the Dynamic Model Compensation algorithm in estimating unmodeled accelerations on satellites encourages the use of this and similar adaptive algorithms, especially for VLBI observations when estimating the unknown excitation function giving rise to earth rotation.

REFERENCES

Arnold, L. (1974). Stochastic Differential Equations: Theory and Applications, New York: Wiley-Interscience.

Aronszajn, N. (1950). "The Theory of Reproducing Kernels," Trans. Amer. Math. Soc., Vol. 68, pp. 337-404.

Bharucha-Reid, A.T. (1972). Random Integral Equations, New York: Academic Press.

Bucy, R.S. and P.D. Joseph. (1968). Filtering for Stochastic Processes with Application to Guidance, New York: Wiley-Interscience.

Burril, C.W. (1972). Measure, Integration and Probability, New York: McGraw-Hill.

Cholshevnikov, K.V. (1965). "On the Values of the Coefficients in the Expansion of the Potential" (in Russian), Proc. Univ. Leningrad, Series on Math., Mech., & Astronomy, No. 13, pp. 155-158.

Cholshevnikov, K.V. (1968). "On the Value of the Tesseral Harmonic Coefficients" (in Russian), Proc. Univ. Leningrad, Series on Math., Mech., & Astronomy, 1, pp. 149-153.

Cramer, H. (1964). "Stochastic Processes as Curves in Hilbert Space," Theory of Probability and Its Applications, Vol. 9, No. 2, pp. 169-179.

Cramer, H. (1965). "A Contribution to the Multiplicity Theory of Stochastic Processes," Proc. 5th Berkeley Symposium on Statistics and Applied Probability, Vol. 2, pp. 215-221.

Cramer, H. (1971). Structural and Statistical Problems for a Class of Stochastic Processes, Princeton: Princeton Univ. Press.

Cramer, H. and M.R. Leadbetter. (1967). Stationary and Related Stochastic Processes, New York: John Wiley.

Dahlen, F.A. (1971). "The Excitation of the Chandler Wobble by Earth-Quakes," Geoph. J. Roy. Astr. Society, Vol. 25, pp. 157-206.

Davis, P.J. (1963). Interpolation and Approximation, New York: Dover.

Deutsch, R. (1965). Estimation Theory, Englewood Cliffs, N.J.: Prentice-Hall.

Doob, J.L. (1953). Stochastic Processes, New York: John Wiley.

Dunford, N. and J.T. Schwartz. (1963). Linear Operators, Part II, New York: Wiley-Interscience.

Eeg, J. and T. Krarup. (1975). "Integrated Geodesy," from Mathematical Methods in Physical Geodesy Lecture Notes, Ramsau, 1973 published in: Brosowski, B. and E. Martensen, eds., Methoden und Verfahren der mathematischen Physik/Band 13, Bibliographisches Institut, Manheim.

Ephremides, A. and J.B. Thomas (eds.). (1973). Random Processes: Multiplicity Theory and Canonical Decompositions, Stroudsburg, Pa.: Dowden Hutchinson and Ross.

Fine, T.L. (1973). Theories of Probability, New York: Academic Press.

Finetti, B. de. (1974). Theory of Probability, Vol. 1, New York: John Wiley.

Gelb. A. (1974). Applied Optimal Estimation, Cambridge, Mass.: The MIT Press.

Gikhman, I.I. and A.V. Skorokhod. (1969). Introduction to the Theory of Random Processes, Philadelphia: W.B. Saunders Co.

Goldstein, H. (1950). Classical Mechanics, Reading, Mass.: Addison-Wesley Publ. Co.

Grenander, U. (1963). Probabilities on Algebraic Structures, New York: John Wiley.

Heiskanen, W. and H. Moritz. (1967). Physical Geodesy, San Francisco: W.H. Freeman and Co.

Ingram, D.S. (1970). Orbit Determination in the Presence of Unmodeled Accelerations, PhD Dissertation, The University of Texas, Austin.

Ingram, D.S. and B.D. Tapley. (1974). "Lunar Orbit Determination in the Presence of Unmodeled Accelerations," Celestial Mechanics, Vol. 9, pp. 191-211.

Jazwinski, A.H. (1969). "Adaptive Filtering," Automatica, Vol. 5, pp. 475-485.

Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory, New York: Academic Press.

Jazwinski, A.H. (1974). "Adaptive Sequential Estimation with Applications," Automatica, Vol. 10, pp. 203-207.

Jordan, J.F. (1966). Optimal Stochastic Control Theory Applied to Interplanetary Guidance, PhD Dissertation, The University of Texas, Austin.

Kalman, R.E. (1960). "A New Approach to Linear Filtering and Prediction Problems," J. Basic Eng., ASME, 82, pp. 35-45.

Kalman, R.E. and R.S. Bucy. (1961). "New Results in Linear Filtering and Prediction Theory," J. Basic Eng., ASME, Ser. D, 83, pp. 95-108.

Kolmogorov, A.N. (1950). Foundations of the Theory of Probability, New York: Chelsea Publ. Co.

Kolmogorov, A.N. and S.V. Fomin. (1960). Measure Lebesgue Integrals and Hilbert Space, New York: Academic Press.

Kolmogorov, A.N. and S.V. Fomin. (1970). Introductory Real Analysis, New York: Prentice-Hall. New York: Dover Reprint, 1975.

Koopmans, L.H. (1974). The Spectral Analysis of Time Series, New York: Academic Press.

Krakiwsky, E.J. (1975). "A Synthesis of Recent Advances in the Method of Least Squares," Lecture Notes, No. 42, Dept. of Surv. Eng., Univ. of New Brunswick, Fredericton, N.B., Canada.

Krarup, T. (1969). "A Contribution to the Mathematical Foundation of Physical Geodesy," Danish Geodetic Inst., Meddelelse No. 44, Copenhagen.

Lauritzen, S.L. (1973). "The Probabilistic Background of Some Statistical Methods in Physical Geodesy," Danish Geodetic Inst., Meddelelse No. 48, Copenhagen.

Liebelt, P.B. (1967). An Introduction to Optimal Estimation, Reading, Mass.: Addison-Wesley Publ. Co.

MacMillan, W.D. (1960). Dynamics of Rigid Bodies, New York: Dover.

McGarty, T.P. (1974). Stochastic Systems and State Estimation, New York: John Wiley.

Mansinha, I. and D.E. Smylie. (1967). "Effect of Earthquakes in the Chandler Wobble and the Secular Polar Shift," J. of Geophysical Research, Vol. 72, No. 18, pp. 4731-4743.

Moritz, H. (1972). "Advanced Least Squares Methods," Reports of the Dept. of Geodetic Science, No. 175, The Ohio State Univ., Columbus.

Mueller, I.I. (1969). Spherical and Practical Astronomy As Applied to Geodesy, New York: Frederick Ungar Publ. Co.

Müller, C. (1966). "Spherical Harmonics," Lecture Notes in Mathematics No. 17, Berlin: Springer-Verlag.

Munk, W.H. and G.J.F. MacDonald. (1960). The Rotation of the Earth, London: Cambridge University Press.

Obukhov, A.M. (1947). "Statistically Homogenous Random Fields on a Sphere" (in Russian), Uspekhi Mat. Nauk, Vol. 2, No. 2, pp. 196-198.

Papoulis, A. (1965). Probability, Random Variables and Stochastic Processes, New York: McGraw-Hill.

Parthasarathy, K.R. (1967). Probability Measures on Metric Spaces, New York: Academic Press.

Parzen, E. (1959). "Statistical Inference on Time Series by Hilbert Space Methods, I," Technical Rept. No. 23, Dept. of Statistics, Stanford Univ.

Parzen, E. (1960). "An Approach to Time Series Analysis," The Annals of Math. Statistics, Vol. 32, No. 4, pp. 951-989.

Parzen, E. (1962). Stochastic Processes, San Francisco: Holden-Day.

Payne, M.H. (1973). "Estimation of Bounds for the Geopotential Ceofficients," Rept. 73-28, Analytical Mechanics Associates, Inc., Jericho, N.Y. 11753.

Pitt, H.R. (1963). Integration Measure and Probability, New York: Hafner Publ. Co.

Prokhorov, Yu. V. (1956). "Convergence of Random Processes and Limit Theorems in Probability Theory," Theory of Probability and Its Applications, Vol. 1, No. 2, pp. 157-214.

Rochester, M.G. (1970). "Polar Wobble and Drift: A Brief History." In: L. Mansinha, D.E. Smylie and A.E. Beck (eds.), Earthquake Displacement Fields and the Rotation of the Earth, pp. 3-13, New York: Springer-Verlag/Dordrecht, Holland: D. Reidel Publ. Co.

Rochester, M.G. (1973). "The Earth's Rotation," EOS, Trans. Amer. Geoph. Union, Vol. 54, No. 8, pp. 769-781.

Royden, H.L. (1968). Real Analysis, London: The Macmillan Co.

Rozanov, J.A. (1968). Infinite Dimensional Gaussian Distributions, Proc. of the Steklov Inst. of Math., No. 108. English translation by Amer. Math. Soc., Providence, R.I., 1971.

Rudnick, P. (1956). "The spectrum of the Variation in Latitude," Trans. Amer. Geoph. Union, Vol. 37, No. 2, pp. 137-142.

Sage, A.P. and J.L. Melsa. (1971). Estimation Theory with Applications to Communication Theory and Control, New York: McGraw-Hill.

Shapiro, H.S. (1971). "Topics in Approximation Theory," Lecture Notes in Mathematics, No. 187, Berlin: Springer-Verlag.

Smylie, D.E., G.K.C. Clarke and T.J. Ulrych. (1973). "Analysis of Irregularities in the Earth's Rotation," Methods of Computational Physics, Vol. 13, pp. 391-430.

Snyder, D.L. (1975). Random Point Processes, New York: John Wiley.

Tapley, B.D. (1973). "Statistical Orbit Determination Theory," Recent Advances in Dynamical Astronomy, B.D. Tapley and V. Szebehely (eds.), pp. 396-425, Dordrecht, Holland: D. Reidel Publ. Co.

Taylor, A.E. (1958). Introduction to Functional Analysis, New York: John Wiley.

Tscherning, C.C. (1973). "Application of Collocation. Determination of a Local Approximation to the Anomalous Potential of the Earth Using 'Exact' Astrogravimetric Collocation," from Mathematical Methods in Physical Geodesy Lecture Notes, Ramsau, 1973 published in: Brosowski, B. and E. Martensen, eds., Methoden und Verfahren der mathematischen Physik/Band 14, Bibliographisches Institut, Manheim.

Uotila, U.A. (1967). Introduction to Adjustment Computations with Matrices, Lecture Notes, Dept. of Geodetic Science, The Ohio State Univ., Columbus.

Wiener, N. (1949). Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications, New York: John Wiley.