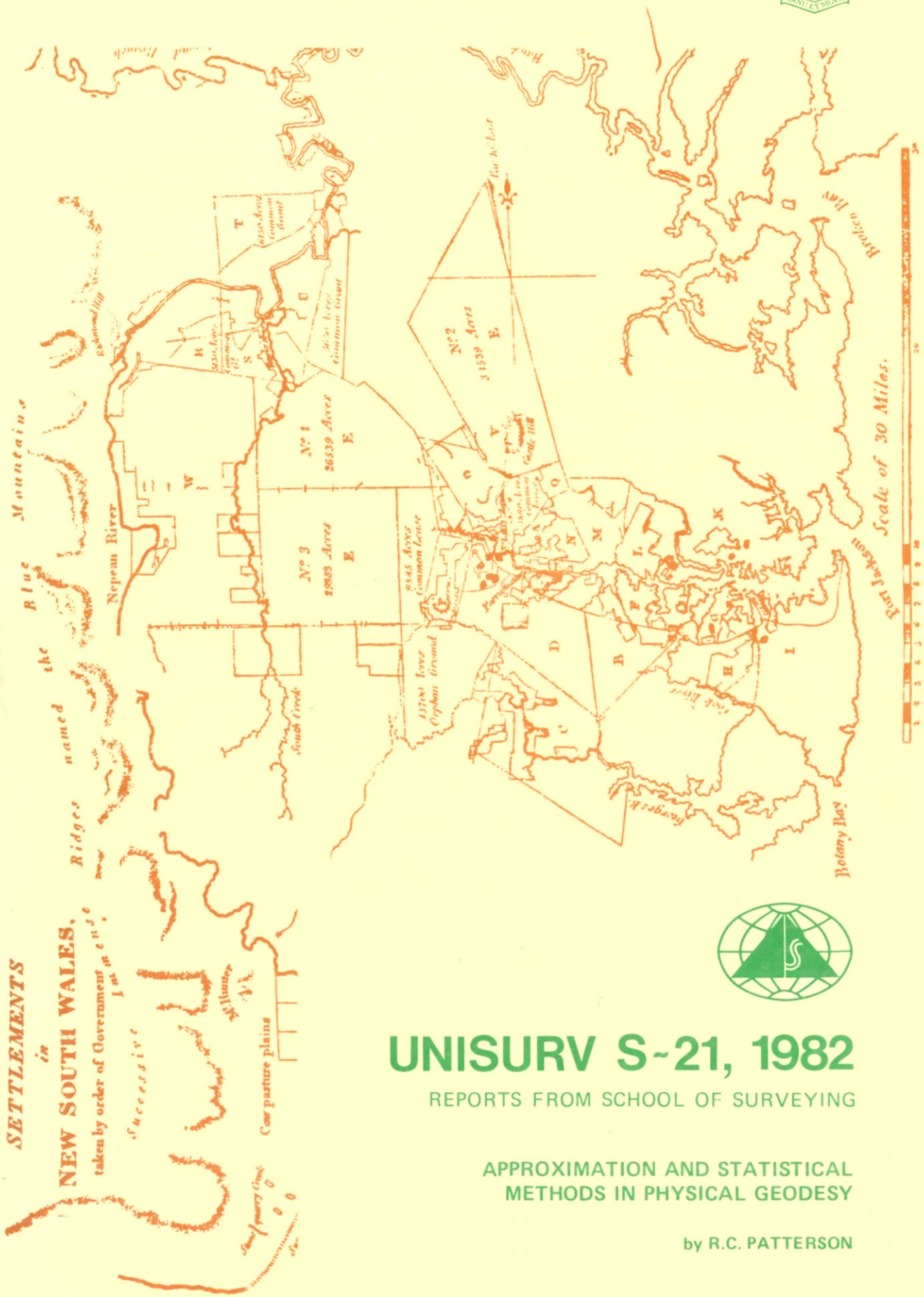
THE UNIVERSITY OF NEW SOUTH WALES
KENSINGTON, NSW AUSTRALIA 2033



UNISURV S-21, 1982

REPORTS FROM SCHOOL OF SURVEYING

APPROXIMATION AND STATISTICAL METHODS IN PHYSICAL GEODESY

by R.C. PATTERSON

UNISURV REPORT NO. 21, 1982

APPROXIMATION AND STATISTICAL METHODS IN PHYSICAL GEODESY

by

R.C. PATTERSON

Received September, 1982

SCHOOL OF SURVEYING
THE UNIVERSITY OF NEW SOUTH WALES
P.O. BOX 1
KENSINGTON, N.S.W. 2033
AUSTRALIA

National Library of Australia
Card No. and ISBN
0 85839029 9

SUMMARY

Various approximation methods of solving differential equations are reviewed and applied to the solution of the Geodetic Boundary value problem. Similarities between the methods when the underlying space is chosen to be a Reproducing Kernel Hilbert Space are discussed.

Statistical methods entailing the use of the projection theorem in the linear least squares prediction of quantities are then examined with a view to their application to Physical Geodesy.

The basic congruence theorem is then used to demonstrate how the statistical and approximation methods correspond in certain specified cases.

Finally the problems associated with the choice or computation of the Covariance function and thus the relevant Reproducing Kernel Hilbert Space and corresponding norm are reviewed and investigated. In particular, a derivation of Lauritzen's result on ergodicity which does not use topological group theory is presented, and a suggestion is made as to how the covariance function may be chosen in such a way as to prevent the norm, or variance, of the predicted quantity from becoming infinite.

TABLE OF CONTENTS

			Page
INTRODUCTION			1
CHAPTER 1	APPROXIMATION		11
	1.1	Introduction	11
	1.2	Methods of Undetermined Coefficients	12
	1.3	The Method of Weighted Residuals (M.W.R.)	14
	1.4	Choice of the Weighting Functions	15
		1.4.1 Galerkins Method	15
		1.4.2 Collocation	16
		1.4.3 Least Squares	18
		1.4.4 Other Methods	19
		1.4.5 Variations of the Collocation Method	21
		(i) Collocation with Derivatives	21
		(ii) Least Squares Collocation	23
		1.4.6 Value of a Linear Functional at the point t_p	29
	1.5	Choice of Functions ψ_i	31
	1.6	Reproducing Kernel Hilbert Spaces	39
		1.6.1 Definition	40
		1.6.2 Examples	40
		1.6.3 Properties of Reproducing Kernels	45
		1.6.4 Representers of Linear Functionals	69
		1.6.5 Projections and Subspaces	76
	1.7	Choice of H_n - A Summary	79
	1.8	Error Bounds	82
CHAPTER 2	STATISTICAL METHODS		93
	2.1	Introduction	93
	2.2	Conditional Expectation	95
	2.3	Second Order Random Functions	98
	2.4	Linear, Inner Product, and Hilbert Spaces	99
	2.5	Hilbert Space Spanned by a Random Function	102

Page

CHAPTER 2 (Continued)

2.6 The Projection Theorem and Linear Least Squares Prediction 105

2.6.1 The Projection Theorem 105
2.6.2 Application of the Projection Theorem to Random Variables 108
2.6.3 Further Properties of Projections 111

2.7 Relationship Between Approximation and Statistical Methods 116

2.7.1 Definitions 116
2.7.2 The Basic Congruence Theorem 117
2.7.3 Representation of Random Functions by Reproducing Kernel Hilbert Spaces 120

2.8 The Covariance Function 129

2.9 Estimating the Covariance Function: Ergodicity 137

2.10 Observation Noise and Trend Removal: Geodetic Least Squares Collocation 153

CONCLUSION 159

ACKNOWLEDGEMENTS 168

BIBLIOGRAPHY 169

INTRODUCTION

Since George Gabriel Stokes published a formula for determination of the Geoid-Spheroid separation in 1849 it has been no small problem for Geodesists to apply it, for there has never been a complete gravity coverage of the earth.

The formula is

$$N = \frac{R}{4\pi G} \iint_{\sigma} \Delta g \, S(\psi) \, d\sigma$$

where $S(\psi) = \frac{1}{\sin(\psi/2)} - 6 \sin\frac{\psi}{2} + 1 - 5\cos\psi - 3\cos\psi \, \ell n(\sin\frac{\psi}{2} + \sin^2\frac{\psi}{2})$

is known as Stokes' function,

R is the radius of the earth

G is the Mean Value of Gravity over the earth

ψ is the spherical distance between the point where N is being determined and the point where Δg is known

Δg is the gravity anomaly

$d\sigma$ is an element of surface area.

If gravity or gravity anomaly were known at all points on the earth with latitude ϕ' and longitude λ', the geoid spheroid separation at a point P with coordinates (ϕ, λ) could be expressed as

$$N(\phi,\lambda) = \frac{R}{4\pi G} \int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \Delta g(\phi', \lambda') S(\psi) \cos\phi' d\phi' d\lambda'$$

where $\psi = \cos^{-1}[\sin\phi\sin\phi' + \cos\phi\cos\phi'\cos(\lambda'-\lambda)]$

(HEISKANEN & MORITZ, 1967)

Now if $\Delta g(\phi', \lambda')$ had some functional form it would perhaps be possible to integrate the above expression analytically, or at least choose some numerical method of integration where it is possible to make the error as small as one wished.

But this is not the case, for the functional form of Δg will never be known as some convenient expression.

If $\Delta g(\phi', \lambda')$ were known at all points on the earth's surface it would not be at all difficult to evaluate the above integral numerically and suitably restrict any errors due to the approximate evaluation by summation simply by choosing ever decreasing element sizes until refinement produced negligible effect.

But this is not the case either for even at the present time, after many years of gravity measurements, and technological advancements making it easier to obtain such data in oceanic regions, there are still vast areas of the earth's surface where no measurements have been made.

The computation of N has been done by two methods (HEISKANEN & MORITZ, 1967)

1) Using templates subdividing the earth into concentric circles, a method now somewhat obsolete

2) By the use of grid lines. Here the earth is subdivided into "rectangular" blocks, the subdivision being effected by the grid lines of some coordinate system, usually latitude and longitude. The blocks are called "squares". The size of these squares varies, for the gravity anomalies close to the computation point affect N much more than those located far away. In fact beyond a certain spherical distance a square is a spherical cap diametrically opposite the computation point represented by a single anomaly value.

Instead of writing Stokes' formula in terms of Latitude and Longitude, it may be written in terms of the Spherical distance ψ, and Azimuth α from the computation point, ie

$$d\sigma = \sin\psi \, d\alpha d\psi. \quad \text{Thus}$$

$$N = \frac{R}{4\pi G} \int_0^{2\pi} \int_0^{\pi} \Delta g(\psi,\alpha) S(\psi) \sin\psi d\psi d\alpha$$

Now, whereas $S(\psi)$ is infinite at $\psi = 0$, the function $F(\psi) = S(\psi)\sin\psi$ is finite for $0 < \psi < \pi$.

Taking a value of Δg to represent each square the above integral can be replaced by a summation.

$$N = K \sum_{i=1}^{b} m_i^2 \sum_{j=1}^{n} F(\psi_{ij}) \, \Delta g_{ij}$$

where b = number of different classes of squares

ie $\frac{1}{2}^\circ \times \frac{1}{2}^\circ$, $1^\circ \times 1^\circ$, $5^\circ \times 5^\circ$ etc.

n = number of squares of a certain size

ψ_{ij} = spherical distance to centre of square i, j

Δg_{ij} = value of gravity anomaly chosen to represent square i, j ; in milligals

m_i^2 = area in $(\text{degrees})^2$ of a class i square

$$K = \frac{R}{4\pi G} \times \frac{\pi^2}{(180)^2} \doteq 1.58 \times 10^{-2}$$

This value of K is computed using R expressed in centimetres and G in milligals and gives N in centimetres.
In addition to these blocks or squares it is also advisable to choose a circular ring around the computation point as the innermost zone and evaluate it separately using individual anomalies.

Now if gravity data were available at all points on the earth's surface it is not difficult to see how one could evaluate N with a minimum of error. Since this is not the case two problems present themselves.

1) In squares where there is a less than adequate coverage of data, how is the existing data used to obtain a representative value for the value of Δg for the square?

2) What is done where there is no data at all within the square?

The earliest work on these problems appears to have been carried out by (DE GRAFF-HUNTER, 1935) and (HIRVONEN, 1956, 1962). In that era it was usual to "solve" problem (2) by using a zero gravity anomaly for all unsurveyed squares. This assumption, although not really valid for a particular square, was reasonable for a large number included in a summation for it was only the assumption that the unknown anomalies were symmetrically distributed about a zero mean. It was really just the same as asserting that the expected value of an anomaly is zero, which is true at least in a literal sense for the word "anomaly" means a departure from an expected value.
Also the field of anomalies should not contain spherical harmonics of orders zero and one, ie. the spheroid should be chosen to coincide with the centre of the earth and should be "best fitting" in that the algebraic sum of all anomalies on the earth should be zero.
Problem (1) was in general solved by using the mean anomaly and research mainly centred on attempts to determine the precision of the derived quantity N. For this the "error of representation" was used:

$$\sigma_S^2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \ldots$$

where $\overline{\Delta g_i}$ = mean anomaly for square i

Δg_{ij} = observed anomaly

n = number of stations in a square

t = number of squares considered.

The quantity G_s was also used:

$$G_s = \frac{1}{t} \sum_{i=1}^{t} (\overline{\Delta g_i} - \overline{\Delta g})^2$$

where $\overline{\Delta g}$ is the general mean = $\sum_{i=1}^{t} \frac{\overline{\Delta g_i}}{t}$,

but it does not appear that any analysis of variance model was applied to the data, nor was any sampling technique other than simple random sampling used.

Hirvonen used the following formula for computation of the mean error of N

$$m(N) = 1.035 \sqrt{\Sigma S^2 E^2 q^2}$$

S^2 being Stokes' function for the square

q^2 the area of the square

E^2 was the error of representation assumed constant for each square.

So at this stage the problem was still being dealt with at a somewhat rudimentary level.

However Hirvonen appears to have been aware that correlation functions or at least correlation could be used in some way to predict gravity anomalies in unsurveyed areas as were Moritz and Rapp. (MORITZ, 1962) gave a general least squares method for interpolation and prediction of gravity anomalies where an unknown anomaly was represented by a linear combination of surrounding anomalies and the difference between true and predicted anomalies minimised.

(RAPP, 1964, also applied this theory to gravity data in the United States.

The launching of Sputnik I in October 1957 drastically altered the Geodesist's viewpoint in that it was no longer earthbound. The perturbation of satellite orbits could now be used in the determination of the Geoid.

(KAULA, 1957, 1959, 1963, 1966, 1967) was prominent in the area of extension of gravity data to unsurveyed areas and the use of satellite derived harmonic models for geoid determinations in the decade following the launching. (UOTILA, 1962) also contributed to the area of Harmonic Analysis of available data.

The Kaula method of satellite enhancement involved the use of a set of satellite derived harmonic coefficients $\bar{C}^{S}_{nm}$, $\bar{S}^{S}_{nm}$ up to the order n = 8 (or less), plus a set of equal-area gravity anomaly area-means.

By means of a least squares adjustment using both sets of data it was possible to correct the $\bar{C}^{S}_{nm}$ and $\bar{S}^{S}_{nm}$ terms to $\bar{C}_{nm}$ and $\bar{S}_{nm}$ terms suitable for use in the formula.

$$\Delta g = \frac{GM}{R^2} \sum_{n=2}^{n(max)} (n-1) \sum_{m=0}^{n} P_{nm}(\sin\varphi)[\bar{C}_{nm}\cos m\lambda + \bar{S}_{nm}\sin m\lambda],$$

(Here, G is the Newtonian Gravitational constant).
The substitution of φ and λ for a point where Δg was required gave quantities usable in the summation formula for N. However with only about 75 $\bar{C}_{nm}$ and $\bar{S}_{nm}$ terms, a high accuracy could not be expected.

Nevertheless apart from satellite enhancement, Kaula's contribution included the introduction of many ideas from the field of [illegible]

(MATHER, 1967) used a least-squares fitting of two-dimensional trigonometrical Fourier Series in the extension of the gravity field in South Australia, and in addition to the statistical methods mentioned so far there were methods of prediction which attempted to use Geophysical phenomena or models such as isostasy to solve the problem of missing data.

So in the mid-sixties there was quite a range of theories and methods. Some unification was necessary.

The direction from which this came appeared basically to be an extension of (MORITZ, 1962) although it was more fundamental, and permitted the use of quantities other than gravity anomalies, not just in the prediction of Δg in unsurveyed areas, but in the direct prediction of N itself, as well as any potential-related quantity of the earth's gravity field. It was called Least-Squares Collocation.

Its invention is usually accredited to (KRARUP, 1969) although the contributions of Bjerhammar, Moritz, Rapp and Tscherning in the period at the close of the sixties should not be underestimated. (MORITZ, 1972) clarified the Hilbert space oriented theory which is somewhat difficult for the uninitiated to understand and extended the theory to the case where observational noise is admitted.

(LAURITZEN, 1973) set out to formalise the theory, and produced the famous result on the non-ergodicity of the empirical covariance function. He also placed great emphasis on the theory of Reproducing Kernel Hilbert spaces. Many aspects of the theory have been treated by a number of writers: (GROTEN, 1972, 1975), (DERMANIS, 1976, 1977), (KEARSLEY, 1977), (LACHAPELLE, 1975, 1977), (LOWREY & ARGENTIERO, 1977), (MEISSL, 1971, 1976), (MORITZ, 1973, 1977, 1978, 1980), (RAPP, 1972, 1973), (RUMMEL, 1976), (SCHWARZ, 1976, 1977), (TSCHERNING, 1970, 1971, 1972a, b, c, 1973, 1974, 1976, a, b, c, 1977, 1978, a, b, c, d, 1979, a, b, 1980) to name a few.

Two of the more interesting aspects dealt with are the result of (TSCHERNING, 1977) concerning the divergence of the norm when the empirical covariance function is used, and the (DERMANIS, 1976, 1977) articles which show a link between probabalistic and deterministic methods of prediction.

The purpose of this project is to review certain aspects of the theory in view of the reluctance of some Geodesists to accept it as a method as good as the solution of Stokes' equation by summation.

This suspicion towards the method possibly has something to do with its development as a statistical method. Accordingly in Chapter 1 the method has been developed as an approximate solution to Laplace's equation, the boundary condition being the spherical approximation of what is known as the "fundamental equation of physical geodesy" (HEISKANEN & MORITZ, 1967, pp 86, 98)

ie

$$\nabla^2 T = \frac{\partial^2 T}{\partial r^2} + \frac{2}{r}\frac{\partial T}{\partial r} + \frac{1}{r^2}\frac{\partial^2 T}{\partial \theta^2} + \frac{\cot\theta}{r^2}\frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 T}{\partial \lambda^2} = 0$$

where (r, θ, λ) are spherical polar coordinates, θ being the colatitude and λ the longitude.

This is Laplace's equation in terms of T, the disturbing potential, and is satisfied outside a sphere of radius R, ie. for $r > R$.

The boundary condition is

$$\Delta g = -\frac{\partial T}{\partial r} - \frac{2T}{r}$$

where Δg is the gravity anomaly and is satisfied on the boundary $r = R$.

So the same equation and boundary condition are used as in the derivation of Stokes' equation, the only difference being that a numerical solution is sought directly rather than taking the intermediate step of deriving Stokes' equation and then seeking a numerical solution.

Instead of using the above equations themselves, a more general operator notation has been used and a quite general method known as the Method of Weighted Residuals has been applied. There are many variations to this method and attention has been focussed specifically on the methods known as Least Squares, Galerkin's Method, and Collocation. It is found that if the test function is a linear combination of the representers of evaluation functionals in a Hilbert Space, the three methods coincide.

The kernel functions in a Reproducing Kernel Hilbert Space are an example of such representers and an extensive treatment of Reproducing Kernels is given, particularly to the class of Reproducing Kernels appropriate in the solution of Laplace's Equation.

Finally the concept of error bounds is dealt with and a suggestion for a method of deriving an optimal Kernel function is given. An example of a Kernel function for which both the norm of the function being predicted and the Kernel function itself are bounded is also given.

In Chapter 2 the statistical treatment is examined. The similarity of the approximation equation to that for conditional expectation is first noted, and the method is then derived as an example of Linear Least Squares Prediction. The Basic Congruence Theorem (PARZEN, 1959) is invoked and the statistical method is shown to give an identical result to that obtained by approximation when the covariance function is chosen to be the reproducing kernel. The result is not new, having been derived by (DERMANIS, 1976, 1977). However Dermanis used the Karhunen-Loeve expansion to do this, a method dependent on the index set in use. The result is here shown to be more general. However, in the Physical Geodesy case, Mercer's Theorem does hold and the Karhunen-Loeve expansion of a continuous

function on the sphere for both anisotropic, and isotropic covariance functions, is derived. On examination of the covariance function, these expansions are used to verify Lauritzen's result on the non-ergodicity of the empirical covariance function by the use of elementary probability theory rather than the rather abstract topological group theory, with which a large proportion of Geodesists are unfamiliar.

It is to be noted that throughout this project classical or Stokesian theory only has been used. However, for the newer theories such as Molodenski's, where integral equations instead of differential equations are involved, approximation methods such as collocation, Galerkin's method, and Least Squares are valid. Hence, much the same treatment could be given. But this was considered to be beyond the scope of the present work so it was decided to restrict the topic to the spherical approximation.

CHAPTER 1

APPROXIMATION

1.1 Introduction

In operator form a linear partial differential equation and its boundary conditions may be expressed as

The differential equation

$$Au = f \quad \text{on} \quad D$$

The boundary conditions

$$Bu = g \quad \text{on} \quad \partial D$$

A and B are linear operators, u, g and f are functions in some function space H, u being unknown. D is some domain in R^n and ∂D is the boundary of D.

In the case of the solution of Laplace's equation with the sperical approximation of the fundamental equation of physical geodesy as boundary condition,

D is the space outside a sphere of radius R,

∂D is the surface of this sphere

$f = 0$

$g = \Delta g\ (r, \theta, \lambda)$ the gravity anomaly function

A is the Laplacian

B is the operator $\{\frac{\partial}{\partial r} + \frac{2}{r}\}_{r=R}$

and u is the potential T.

Since f = 0 homogeneous equations alone need be of interest, ie

$$Au = 0 \quad \text{on} \quad D$$

$$Bu = g \quad \text{on} \quad \partial D.$$

An approximation of u is sought. It is of the form:

$$u \doteq u^* = c_1\psi_1 + c_2\psi_2 + \ldots + c_n\psi_n$$

where ψ_i , $i = 1,2, \ldots, n$ are linearly independent functions which span some n dimensional function space H_n.

These functions must all satisfy the equation

$$A\psi_i = 0 \qquad i = 1,2, \ldots, n \text{ on } D$$

(More specifically they must satisfy Laplace's equation).

They are thus chosen such that the above equation is satisfied.

This affects the choice of H_n.

Note also that $Au^* = 0$ on D

must therefore be satisfied for A is a linear operator and u* is a linear combination of the ψ_i's.

Of course there will be many such sets of functions $\{\psi_i\}$ which will satisfy the above,and many choices of H_n. This will be dealt with in sections 1.5 - 1.7. For the moment, however it will be assumed that these choices have already been made. The problem then is to determine values of $c_1, c_2, \ldots, c_n$ such that the linear combination u* satisfies the boundary conditions, ie such that both

$$Au^* = 0 \text{ on } D$$

$$\text{and } Bu^* = g \text{ on } \partial D.$$

This provides the approximation of u.

1.2 Methods of Undetermined Coefficients

These methods have been applied to the solution of differential equations since the early part of this century. I.G.Bubnov in 1913 and B.G. Galerkin in 1915 appear to have done the pioneer work leading to the Bubnov-Galerkin method. M. Picone in 1928 applied the method of least squares to this problem whilst J.C. Slater in 1934 and J.Barta

in 1937 applied the method of collocation in the solution of differential equations related to electronic energy bands in metals, and the torsion of a square prism, respectively.

Other methods include the subdomain method advanced in 1923 by C.B. Bilzeno and J.J.Koch, the method of moments developed by H. Yamada and H. Fyita in 1947-51 and the integral method of von Karman and Pohlhausen in 1921. In addition it is possible to combine the least squares and collocation method if there are more data points than there are constants to be determined. This method is called least squares collocation and was applied by S.L. Altman in 1958 to calculations in crystallography.

There are two ways that these methods may be applied: as interior methods and as boundary methods. In interior methods the trial functions ψ_1, ψ_2, ..., ψ_n are chosen to satisfy the boundary conditions and one of the abovementioned procedures is used to determine the constants c_1, c_2, ... c_n such that the differential equation is satisfied. In boundary methods the reverse is the case. Mixed methods also exist. Since in this project it is Laplace's equation outside a sphere which is of interest, only boundary methods will be dealt with.

S.H. Crandall in 1956 and L. Collatz in 1961 unified the various procedures, Crandall as the "Method of Weighted Residuals" and Collatz as "error distribution principles", although the similarity of the methods had been recognized by C.B. Biezeno and R. Courant as early as 1924. It may be noted that Crandall and Collatz merely use different terms for essentially the same unification. For the historical development of these methods see (FINLAYSON, 1972) where the above is dealt with in more detail, and references are supplied.

1.3 The Method of Weighted Residuals (M.W.R.)

Since the method applies where a functional form of the boundary condition is indeed known, and is applied mainly because the use of rigorous methods is too difficult, too cumbersome or even impossible the method will be described for this case first. Later it will be shown how this method may be adopted where no functional form is available and one is forced to use it because one has only values on the boundary at certain discrete points.

In general one has

$Au = 0$ in D as differential equation

$Bu = g$ on ∂D as the boundary condition

where ∂D is the boundary of D, and

A and B are linear operators.

$\psi_1, \psi_2, \ldots, \psi_n$ are chosen such that they are solutions of the differential equation.

An approximation of the form

$$u^* = c_1\psi_1 + c_2\psi_2 \ldots + c_n\psi_n$$

is sought such that in some sense it approximates the boundary condition.

Let the "residual" at some point t on ∂D be defined as $R(t)$,

where $R(t) = Bu^*(t) - g(t)$

or simply,

$$R = Bu^* - g$$

Let there be n "weighting functions" $W_i(t)$ also defined on ∂D.

This is represented as W_i in function form.

The the MWR criterion is that the inner product of the weighting function with the residual is zero.

ie $\langle W_i, R \rangle = 0 \qquad i = 1, 2, \ldots, n.$

Thus $\langle W_i, Bu^* - g \rangle = 0$

ie $\langle W_i, Bu^* \rangle = \langle W_i, g \rangle$

and hence $\langle W_i, c_1 B\psi_1 + c_2 B\psi_2 + \ldots + c_n B\psi_n \rangle = \langle W_i, g \rangle$

or $c_1 \langle W_i, B\psi_1 \rangle + c_2 \langle W_i, B\psi_2 \rangle + \ldots + c_n \langle W_i, B\psi_n \rangle = \langle W_i, g \rangle$

$i = 1, 2, \ldots, n.$

It can be seen that in all methods there will be n equations in the n unknowns $c_1, c_2, \ldots, c_n$. Each choice of the weighting function leads to a different method.

1.4 Choice of the Weighting Functions

1.4.1 Galerkin's Method (also called the Bubnov-Galerkin method)

Here the weighting functions $W_i(t)$ are chosen to be the same as the functions $\psi_i(t)$ which satisfy the original differential equation.

ie $\langle \psi_i, Bu^* - g \rangle = 0$

If H_n is a Hilbert Space then it is the same as requiring that the residuals are orthogonal to all $\underset{\sim}{v}$ in H_n or that the solution is the projection of g on H_n.

Hence

$$c_1 \langle \psi_i, B\psi_1 \rangle + c_2 \langle \psi_i, B\psi_2 \rangle + \ldots + c_n \langle \psi_i, B\psi_n \rangle = \langle \psi_i, g \rangle$$

$i = 1, 2, \ldots, n.$

In matrix form then

$$G\underset{\sim}{c} = \underset{\sim}{f}$$

where $g_{ij} = \langle \psi_i, B\psi_j \rangle$

$$\underset{\sim}{c}^T = [c_1 \; c_2 \; \ldots \; c_n]$$

and

$$\underset{\sim}{f}^T = [f_1 \; f_2 \; \ldots \; f_n] \quad \text{where } f_i = \langle \psi_i, g \rangle$$

1.4.2 Collocation

In this method the W_i's are chosen to be displaced dirac delta functions

$$W_i(t) = \delta(t - t_i) \qquad t, t_i \in \partial D \qquad i = 1, 2, \ldots n.$$

The dirac delta function for functions defined on the real line has the property

$$\int_{-\infty}^{\infty} \delta(t-a)\, dt = 1$$

ie. it vanishes everywhere except t = a where it has unit "area".

In R^3, $\delta(\underset{\sim}{r} - \underset{\sim}{a}) = \delta(x - a_1)\, \delta(y - a_2)\, \delta(z - a_3)$

and $$\int_{R^3} \delta(\underset{\sim}{r} - \underset{\sim}{a})\, dV = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x - a_1)\, \delta(y - a_2)\, \delta(z - a_3)\, dzdydx$$

$$= 1$$

Also, on the real line we have

$$\int_{-\infty}^{\infty} \delta(t - a)\, f(t)\, dt = f(a)$$

and in R^3 $$\int_{R^3} \delta(\underset{\sim}{r} - \underset{\sim}{a})\, f(\underset{\sim}{t})\, dV = f(\underset{\sim}{a})$$

In a general Hilbert space them, the dirac delta corresponds to the representer ℓ_{t_i} of an evaluation functional $\ell^*_{t_i}$

ie if f is in H_n, then $\ell^*_{t_i}$ is in H_n^*, the dual space of H_n

$$\ell^*_{t_i}(f) = f(t_i) = \langle \ell_{t_i}, f \rangle ,$$

where ℓ_{t_i} is the representer of $\ell^*_{t_i}$ in H_n.

Or, in dirac delta notation

$$\langle \delta(t - t_i),\ f(t) \rangle = f(t_i)$$

So with $W_i = \delta(t - t_i)$

$$\langle W_i, B\psi_j \rangle = \langle \delta(t - t_i),\ B\psi_j(t) \rangle$$

$$= B\psi_j \Big|_{t_i} = B\psi_j(t_i)$$

$$\text{and } \langle W_i, g \rangle = \langle \delta(t - t_i), g(t) \rangle$$

$$= g(t_i)$$

Again ,in matrix notation

$$G\,\underset{\sim}{c} = \underset{\sim}{f}$$

but this time

$$g_{ij} = B\psi_j(t_i)$$

$$\text{and } f_i = g(t_i)$$

$$\text{or } c_1 B\psi_1(t_i) + c_2 B\psi_2(t_i) + \ldots + c_n B_n(t_i) = g(t_i)$$

$$i = 1, 2, \ldots, n.$$

Thus what has been done is to choose as many locations t_i as there are coefficients c_i and make the function u^* equal to g at these locations. So at the n locations t_i the approximation is exact or "correct". Hence the name collocation. In using this method one would hope that the residual does not get very far away from zero between the points where it vanishes. The choice of collocation points would be a matter of some concern therefore, if the functional form of g were known. In the Geodetic problem however, there is no choice, since it is not known, although $g(t_i)$ is known at various points.

1.4.3 Least Squares

Geodesists are quite familiar with the Gauss-Legendre principle of least squares since it is used in the adjustment of survey networks, ie that the sum of the squares of the residuals should be a minimum for the best linear unbiassed estimator.

In the solution of differential equations the residual function is not a discrete point function, usually, but is known over a more or less continuous domain. So it is a question of the minimisation of an integral of squared residuals, rather than the sum of the squares of residuals, but then an integral is really a generalisation of a sum, so the principle is much the same. However by the use of Hilbert space theory there is a common link. What is minimised is the inner product of the residual with itself and thus the norm of the residual is also minimised.

$$\text{ie} \quad \min_{u^* \in H_n} \|R\| \quad \text{where} \quad \|R\| = (R, R)^{\frac{1}{2}}$$

so $c_1, c_2, \ldots, c_n$ are chosen to minimise

$$\begin{aligned} F &= \langle Bu^* - g,\ Bu^* - g \rangle \\ &= \langle Bu^*, Bu^* \rangle - 2\langle Bu^*, g \rangle + \langle g, g \rangle \\ &= \langle c_1 B\psi_1 + c_2 B\psi_2 + \ldots + c_n B\psi_n,\ c_1 B\psi_1 + \ldots + c_n B\psi_n \rangle \\ &\qquad - 2\langle c_1 B\psi_1 + c_2 B\psi_2 + \ldots + c_n B\psi_n, g \rangle + \langle g, g \rangle \\ &= \underset{\sim}{c}^T G \underset{\sim}{c} - 2\underset{\sim}{c}^T \underset{\sim}{f} + \langle g, g \rangle \end{aligned}$$

where $G = [g_{ij}]$, $g_{ij} = \langle B\psi_i, B\psi_j \rangle$

and $f_i = \langle B\psi_i, g \rangle$

To minimise F, $\frac{\partial F}{\partial \underset{\sim}{c}}$ is equated to zero giving

$$2G\underset{\sim}{c} - 2\underset{\sim}{f} = 0$$

$$\text{or} \quad G\underset{\sim}{c} = \underset{\sim}{f}$$

where G and $\underset{\sim}{f}$ are as stated above.

This is equivalent to choosing $\frac{\partial Bu^*}{\partial c_i}$ as the weighting function W_i in the equation

$$\langle W_j, R \rangle = 0$$

ie

$$\langle \frac{\partial Bu^*}{\partial c_i}, R \rangle = 0 \qquad i = 1, 2, \ldots, n$$

$$\langle \frac{\partial Bu^*}{\partial c_i}, Bu^* - g \rangle = 0$$

$$\therefore \quad \langle \frac{\partial Bu^*}{\partial c_i}, Bu^* \rangle = \langle \frac{\partial Bu^*}{\partial c_i}, g \rangle$$

$$\therefore \quad \langle B\psi_i, c_1 B\psi_1 + c_2 B\psi_2 + \ldots + c_n B\psi_n \rangle = \langle B\psi_i, g \rangle$$

$$i = 1, 2, \ldots, n$$

since B is a linear operator,

$$\text{or} \quad c_1 \langle B\psi_i, B\psi_1 \rangle + c_2 \langle B\psi_i, B\psi_2 \rangle + \ldots + c_n \langle B\psi_i, B\psi_n \rangle = \langle B\psi_i, g \rangle,$$

$$i = 1, 2, \ldots, n$$

as before

1.4.4 Other Methods

In the subdomain method, the domain ∂D is divided into n subdomains, where n is the number of coefficients to be determined, then the coefficients are determined in such a way as to make the average value of the residual in each subdomain zero.

This is equivalent to using an indicator function as the weighting function in M.W.R.

ie. $$W_i(t) = I_{\partial D_i}(t) = \begin{cases} 1, & \text{for } t \text{ in } \partial D_i \\ 0, & \text{for } t \text{ not in } \partial D_i \end{cases}$$

where ∂D_i is a subdomain of ∂D. $i = 1, 2, \ldots, n$.

This method could be used in Physical Geodesy if one were to divide the earth into "squares" and allow the average value of the measured gravity anomalies in a square represent the product

$$\langle I_{\partial_i D}(t), g(t) \rangle$$

But the problem with this method is that there are squares where no anomaly has been determined, and so the method has the same difficulties as the use of Stokes' Integral. There is no need to make the "squares" of equal size, however.

In the Method of Moments, successively higher moments of the residual are equated to zero. For ordinary differential equations then, the weighting functions $W_i(t)$ would be 1, t, t^2, ..., t^{n-1}. It would appear that in partial differential equations, product moments would make the solution somewhat cumbersome. The integral method is just a first approximation of the Method of Moments.

In Orthogonal Collocation the functions ψ_1, ψ_2, ... ψ_n are chosen to be orthogonal polynomials and the collocation points are chosen as the roots of the polynomials. However since it is not possible to choose the collocation points in the Geodetic Boundary Value Problem it appears that the method is not of great interest here.

If the weighting functions are a set of complete functions but are not the trial functions the method is merely a general MWR.

One way of doing this is to choose some differential operator D of the same order as A and form the weighting functions $W_i(t) = D\psi_i(t)$. This method was proposed by Krawchuk in 1932.

It appears so far that most of the methods mentioned will not be of great use in solving the Geodetic Boundary Value Problem (GVBP) since they require some knowledge of the functional form of g(t) rather than its evaluation at various points. An exception to this is the Collocation method of which there are some variations.

1.4.5 Variations of the Collocation Method

(i) Collocation with Derivatives

(COLLATZ, 1966) describes a procedure which he calls "Collocation with Derivatives" in which g is not known at all the collocation points but derivatives of g are. Rather that restricting the disucssion to derivatives however, the general case where various linear functionals of g are known at the n points, some of which could be evaluation functionals (but not necessarily all of them) will be dealt with.

The equations are essentially the same except that the operator B is now partitioned

ie $$B = \begin{pmatrix} B_1 \\ \hline B_2 \\ \hline \vdots \\ \hline B_m \end{pmatrix}$$

where there are m linear operators of different type, ie.

$$B_k = D_k B \quad \text{where } D_k \text{ is a linear operator}$$

g is also partitioned into m functions, each related to g by a linear operator equation of the form $g_k = D_k g$

ie. $Au = 0$ for t in D

$$\begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_m \end{pmatrix} u = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{pmatrix}$$ for t in ∂D

which is the same as writing

$$\left.\begin{aligned} Au &= 0 \\ B_1u &= g_1 \\ B_2u &= g_2 \\ \cdots &\quad \cdots \\ B_mu &= g_m \end{aligned}\right\}$$ for t in D, for t in ∂D

Note $m \leq n$.

The solution is given by

$$G\underset{\sim}{c} = \underset{\sim}{f}$$

But this time $g_{ij} = B_k\psi_j(t_i)$

with $f_i = g_k(t_i)$

This is to say that it is now important to use the operator B_k and function g_k relevant to the evaluation point t_i.

Now, g_{ij} may also be expressed as follows:

$$g_{ij} = \langle \ell_i, B_k\psi_j \rangle .$$

We may regard B_k as the composition of two operators D_k and B such that $B_k = D_kB$.

Then $g_{ij} = \langle \ell_i, D_kB_j \rangle$

or $g_{ij} = \langle D_k^*\ell_i, B\psi_j \rangle$

where D_k^* is the adjoint of D_k.

But $D_k^* \ell_i$ is itself the representer of a functional since it gives a real number when its inner product is taken with $B\psi_j$. Let this functional be called $\bar{\ell}_i^*$ with representer $\bar{\ell}_i$

Then $g_{ij} = \langle \bar{\ell}_i, B\psi_j \rangle$

Note also that $f_i = g_k(t_i)$

But if $D_k B = B_k$

and $B_k u = g_k$

then $D_k B u = g_k = D_k g$

ie $f_i = \langle \ell_i, D_k g \rangle$

$= \langle D_k^* \ell_i, g \rangle$

$= \langle \bar{\ell}_i, g \rangle$

So the general form of a collocation solution is

$$c_1 \langle \bar{\ell}_i, B\psi_1 \rangle + c_2 \langle \bar{\ell}_i, B\psi_2 \rangle + \ldots + c_n \langle \bar{\ell}_i, B\psi_n \rangle = \langle \bar{\ell}_i, g \rangle$$

$$i = 1, 2, \ldots, n.$$

(ii) Least Squares-Collocation

In the pure collocation and collocation with derivatives methods, the value of g(t) or some functional of g(t) was known at n points, which was the same number of points that there were constants c_i. In least squares g(t) was known for all $t \in \partial D$. But what is to be done if g(t),(or some functional of g(t)),is known at $M > n$ points?

It is simpler to deal with pure collocation first, the extension to collocation with derivatives then follows immediately.

The residual function was

$$R(t) = Bu^*(t) - g(t).$$

Using evaluation functional notation ℓ_i^* rather than dirac deltas, we would have for the evaluation of R(t) at the M points t_i

$$W_i = \ell_i \qquad i = 1, 2, \ldots, M$$

ie

$$R(t_i) = \langle \ell_i, Bu^* - g \rangle \qquad i = 1, 2, \ldots, M.$$

If each of these residuals is set equal to zero, M equations of the form

$$\langle \ell_i, Bu^* \rangle = \langle \ell_i, g \rangle \quad \text{are obtained.}$$

But since

$$u^* = c_1\psi_1 + c_2\psi_2 + \ldots + c_n\psi_n$$

and M > n there is no solution for $\underset{\sim}{c}$ as the equations are inconsistent. So the best that can be done is to minimise the (weighted) sum of the squares of the residuals instead of equating each residual to zero. We have then

$$\underset{\sim}{R} = H\underset{\sim}{c} - \underset{\sim}{f}$$

where

$$\underset{\sim}{R}^T = [R(t_1), R(t_2) \ \ldots \ R(t_M)]$$

$$\underset{\sim}{c}^T = [c_1 c_2 \ \ldots \ c_n]$$

$$\underset{\sim}{f}^T = [\langle \ell_1, g \rangle \ \langle \ell_2, g \rangle \ \ldots \ \langle \ell_M, g \rangle]$$

$$H = [h_{ij}] \text{ where } h_{ij} = \langle \ell_i, B\psi_j \rangle$$

The minimisation of $\underset{\sim}{R}^T W \underset{\sim}{R}$ with respect to $\underset{\sim}{c}$ gives

$$\hat{\underset{\sim}{c}} = (H^T W H)^{-1} H^T W \underset{\sim}{f}$$

where W is a suitable M x M weighting matrix.

After minimisation the residual is

$$\hat{\underset{\sim}{R}} = \{H(H^T W H)^{-1} H^T W - I\}\underset{\sim}{f}$$

which in general is not equal to $\underset{\sim}{0}$, so the method is not a true collocation method in the sense of the residuals vanishing at all of the collocation points.

When M = n the method reduces to collocation and as M becomes infinite the method tends towards least squares. In the case of least squares-collocation with derivatives it is only necessary to substitute $\bar{\ell}_i^*$, a general linear functional, for ℓ_i^* the evaluation functional ie substitute $\bar{\ell}_i$ for ℓ_i in each of the above equations where ℓ_i appears.

The weight matrix W allows one to give relative importance to each of the collocation points. For example, n of the points could be given weight 1 and M-n of the points weight zero. In this case the method reverts to pure collocation on the n points with weight 1. In the case where $g(t_i)$ is determined by measurement the weights can be made equal to the reciprocal of the variances of each measurement. If the measurements are independent the matrix W will be diagonal but if they are correlated W will be the inverse of the covariance matrix Σ of the observations. If it is assumed that the observations are error free an identity matrix I_{MxM} should be used, the error being assumed to be solely in the approximation method used rather than distributed between the observations and the approximation.

The similarity of this method to the least squares method is seen by examining the equation for the pure least squares method.

Here

$$c_1 \langle B\psi_i, B\psi_1 \rangle + c_2 \langle B\psi_i, B\psi_2 \rangle + \ldots + c_n \langle B\psi_i, B\psi_M \rangle = \langle B\psi_i, g \rangle \qquad i = 1, 2, \ldots, n.$$

If the inner product $\langle \, , \rangle$ which is usually some sort of integral is replaced by $(\, ,)$ where

$$(B\underset{\sim}{\psi}_i, B\underset{\sim}{\psi}_j) = \underset{\sim}{b}_i^T W \underset{\sim}{b}_j$$

where $\underset{\sim}{b}_k^T = [B\psi_k(t_1) \quad B\psi_k(t_2) \quad \ldots \quad B\psi_k(t_M)]$

$= [\langle \ell_1, B\psi_k \rangle \quad \langle \ell_2, B\psi_k \rangle \quad \ldots \quad \langle \ell_M, B\psi_k \rangle]$

ie $\underset{\sim}{\psi}_k^T = [\psi_k(t_1) \quad \psi_k(t_2) \quad \ldots \quad \psi_k(t_M)]$

and W is the MxM weight matrix mentioned above, then $(B\underset{\sim}{\psi}_i, B\underset{\sim}{\psi}_j)$ is obviously the i, j th element of the matrix H^TWH.

In addition

$(B\underset{\sim}{\psi}_i, \underset{\sim}{g}) = \underset{\sim}{b}_i^T \; W\underset{\sim}{f}$, the ith element of the vector $H^TW\underset{\sim}{f}$

So the equations

$$c_1(B\underset{\sim}{\psi}_i, B\underset{\sim}{\psi}_1) + c_2(B\underset{\sim}{\psi}_i, B\underset{\sim}{\psi}_2) + \ldots + c_n(B\underset{\sim}{\psi}_i B\underset{\sim}{\psi}_n) = (B\psi_i, g)$$

$$i = 1, 2, \ldots, n$$

with the inner product (,) as defined are equivalent to the matrix equation

$$\hat{\underset{\sim}{c}} = (H^TWH)^{-1} H^TWf$$

It would appear that if it was required that the inner product (,) converge to $\langle , \rangle$ for an infinite number of points where $\langle , \rangle$ is of a form similar to

$$\int_\Omega w(t) \; \phi_i(t) \quad \phi_j(t) \; dt, \qquad t \in \Omega$$

then the weight matrix would be diagonal and of the form $w_{ii} = w(t_i)$ where w(t) is the weight function of the inner product $\langle , \rangle$. Notwithstanding this, the weight matrix also gives the opportunity to introduce the factor of observation noise. Perhaps a way of including both elements is to use a matrix of the form $W = (W_1^{-1} + W_2^{-1})^{-1}$ where W_1 represents observation noise and W_2 represents the weighting function of the inner product $\langle , \rangle$.

Finally, it may be noted that this method resembles the least squares method more than the collocation method. In fact it is not really a collocation method at all since the residual is not made zero at the collocation points. "Discrete least-squares" would probably be a better name for it, but it is best not to introduce new names for a method which already has an established name. It should also not be confused with what (MORITZ 1972) refers to as least squares collocation. To distinguish between the two methods in this project the method of Moritz will be called "Geodetic Least Squares Collocation", the present method remaining as is.

1.4.6 Summary of MWR methods

In all cases the matrix equation is

$$G\underset{\sim}{c} = \underset{\sim}{f} ,$$

where $\underset{\sim}{c}^T = [c_1 \; c_2 \; \ldots \; c_n]$

and g_{ij}, the i, j th element of G and f_i the i th element of $\underset{\sim}{f}$ are given by the choice of the method

(i) Galerkin's method

Weighting function: $W_i = \psi_i$

The G matrix: $g_{ij} = \langle \psi_i, B\psi_j \rangle$

The $\underset{\sim}{f}$ vector: $f_i = \langle \psi_i, g \rangle$

(ii) Collocation

Weighting function: $W_i = \ell_i$, the representer of the evaluation functional at t_i, ℓ_i^*

or $W_i = \delta(t - t_i)$ the displaced dirac delta.

The G matrix: $g_{ij} = \langle \ell_i, B\psi_j \rangle = B\psi_j(t_i)$

The $\underset{\sim}{f}$ vector: $f_i = \langle \ell_i, g \rangle = g(t_i)$

(iii) Least Squares

Weighting function: $W_i = \frac{\partial Bu}{\partial c_i}^*$

The G matrix: $g_{ij} = \langle B\psi_i, B\psi_j \rangle$

The f vector: $f_i = \langle B\psi_i, g \rangle$

(iv) Collocation with Derivatives

Weighting function: $W_i = \bar{\ell}_i = D_k^* \ell_i$

where D_k^* is some linear operator, ℓ_i an evaluation functional

The G matrix: $g_{ij} = \langle \bar{\ell}_i, B\psi_j \rangle$

The $\underset{\sim}{f}$ vector: $f_i = \langle \bar{\ell}_i, g \rangle$

(v) Least Squares Collocation

Weighting function: $W_i = \frac{\partial Bu}{\partial c_i}^*$

The G matrix: $g_{ij} = (B\underset{\sim}{\psi}_i, B\underset{\sim}{\psi}_j)$

The $\underset{\sim}{f}$ vector: $f_i = (B\psi_i, g)$

where the inner product (,) is as defined above. The inner product $\langle\ ,\ \rangle$ is in all other cases the inner product of the solution space of $Au = 0$, eg. a Hilbert Space of functions harmonic outside a sphere of radius R and regular at infinity.

On obtaining $\underset{\sim}{c}$, these coefficients can be substituted into

$$u^* = c_1\psi_1 + c_2\psi_2 + \ldots + c_n\psi_n$$

to give the solution to the differential equation which approximately satisfies the boundary conditions.

However, it is not so much the functional form of the approximate solution which is of interest in this project. What is of interest is the approximate value of u(t), or of some linear functional $\bar{\bar{\ell}}_p^*$ of u(t) at some point t_p, $t_p \in \partial D$, for example the geoid-spheroid separation N, the value of the gravity anomaly Δg or a deflection of the vertical in the meridian ξ at some point t_p on the surface of the earth where no quantity has actually been observed.

1.4.7 Value of a Linear Functional at the point t_p

What is sought is the quantity

$$d_p = \langle \bar{\bar{\ell}}_p, u \rangle .$$

An approximation may be obtained using u^*

ie. $$\hat{d}_p = \langle \bar{\bar{\ell}}_p, u^* \rangle$$

$$= \langle \bar{\bar{\ell}}_p, \sum_{i=1}^{n} c_i \psi_i \rangle$$

$$= \sum_{i=1}^{n} c_i \langle \bar{\bar{\ell}}_p, \psi_i \rangle$$

$$= \underset{\sim}{p}^T \underset{\sim}{c}$$

where $$p^T = [\langle \bar{\bar{\ell}}_p, \psi_1 \rangle \quad \langle \bar{\bar{\ell}}_p, \psi_2 \rangle \quad \dots \quad \langle \bar{\bar{\ell}}_p, \psi_n \rangle]$$

So

$$\hat{d}_p = \underset{\sim}{p}^T G^{-1} \underset{\sim}{f},$$ G and $\underset{\sim}{f}$ having been obtained by one of the various M.W.R.

Note that where ℓ_p^* is the evaluation functional of Bu it is also the evaluation functional of u if B^{-1} exists.

ie if $\ell_p^*(Bu) = Bu(t_p)$

then $\ell_p^*(u) = u(t_p)$

For suppose B has an inverse operator B^{-1}

$$\langle \ell_p, u \rangle = \langle \ell_p, B^{-1}Bu \rangle = \langle B^{*-1}\ell_p, Bu \rangle = B^{-1}\ell_p^*(Bu)$$

$$= B^{-1}Bu(t_p) = u(t_p)$$

If $\bar{\bar{\ell}}_p = D_k^*\ell_p = \bar{\ell}_p$, one of the functionals in the original formulation,

then $p^T = [\langle \bar{\ell}_p, \psi_1 \rangle \quad \langle \bar{\ell}_p, \psi_2 \rangle \quad \ldots \quad \langle \bar{\ell}_p, \psi_n \rangle]$

In particular if t_p is one of the collocation points t_i in the pure collocation method, and

$$\langle \bar{\ell}_i, u \rangle = Bu(t_i)$$

then $\bar{\ell}_i = B^*\ell_i$ and

$$\underset{\sim}{p}^T = [\langle B^*\ell_i, \psi_1 \rangle \ \langle B^*\ell_i, \psi_2 \rangle \ \ldots \ \langle B^*\ell_i, \psi_n \rangle]$$

$$= [\langle \ell_i, B\psi_1 \rangle \ \langle \ell_i, B\psi_2 \rangle \ \ldots \ \langle \ell_i, B\psi_n \rangle]$$

But this is the i th row of the matrix G in the pure collocation method. When multiplied by G^{-1} it will provide a row matrix with all elements zero except the i th one which is unity.

Calling this row matrix $\underset{\sim}{r}_i^T$, then

$$\hat{d}_i = \underset{\sim}{p}^T G^{-1}\underset{\sim}{f} = \underset{\sim}{r}_i^T\underset{\sim}{f} = f_i = g(t_i)$$

In other words the original value is returned.

In the case of collocation with derivatives, much the same thing happens. At the point t_i the value observed was, say, $D_k g(t_i)$.

Now $D_k g(t_i) = D_k Bu(t_i) = \ell_i^*\{D_k Bu\}$

$$= \langle \ell_i, D_k Bu \rangle = \langle D_k^*\ell_i, Bu \rangle = \langle \bar{\ell}_i, Bu \rangle$$

$$= \langle B^*\bar{\ell}_i, u \rangle$$

So the representer of the functional required is $B^*\ell_i$.

Thus

$$\underset{\sim}{P}^T = [\langle \bar{\ell}_i, B\psi_1 \rangle \ \langle \bar{\ell}_i, B\psi_2 \rangle \ \ldots \ \langle \bar{\ell}_i, B\psi_n \rangle], \text{ which}$$

is the i th row of the G matrix in the collocation with derivatives or "mixed" collocation method and hence $\underset{\sim}{P}^T G^{-1} = \underset{\sim}{r}_i^T$.

Thus $\hat{d}_i = \underset{\sim}{r}_i^T \underset{\sim}{f} = \underset{\sim}{f}_i = \langle \bar{\ell}_i, g \rangle = \langle \bar{\ell}_i, Bu \rangle = D_k g(t_i)$, and the original functional of g which was observed is returned.

1.5 Choice of Functions ψ_i

In the case of least squares collocation, however, it is not expected that the observed value $g(t_i)$ will be returned.

But it may be possible by some choice of the functions ψ_i to obtain a situation where

$$\underset{\sim}{P}^T = [(B\underset{\sim}{\psi}_i, B\underset{\sim}{\psi}_1) \ (B\underset{\sim}{\psi}_i, B\underset{\sim}{\psi}_2) \ \ldots \ (B\underset{\sim}{\psi}_i, B\underset{\sim}{\psi}_n)]$$

Then, in the manner described above, the i th element of the vector $\underset{\sim}{f}$ would be returned.

Now $f_i = (B\underset{\sim}{\psi}_i, \underset{\sim}{g})$

But if $\underset{\sim}{P}^T$ was indeed as described above then the evaluation functional under this inner product would in fact be $B\psi_i$, and thus f_i would equal $g(t_i)$. Although at first it appears unlikely that this speculation would amount to anything, it illustrates the similarity between the methods and may give some clue as to the way the functions $\psi_1, \psi_2, \ldots \psi_n$ could be chosen.

Similar remarks would again apply to the pure least squares method the difference being that the inner product $\langle , \rangle$ is used instead of $(,)$. ie. here one would require that

$$B\psi_j(t_i) = \langle B\psi_i, B\psi_j \rangle$$

for a collocation type situation.

More interesting in this regard is Galerkin's method. A choice of functions $\psi_1, \psi_2, \ldots \psi_n$ would have to be made such that

$$\underset{\sim}{p}^T = [\langle \psi_i, B\psi_1 \rangle \ \langle \psi_i, B\psi_2 \rangle \ \ldots \ \langle \psi_i, B\psi_n \rangle]$$

ie such that $B\psi_j(t_i) = \langle \psi_i, B\psi_j \rangle$.

Then $\hat{d}_i = \langle \psi_i, g \rangle$

But note that $B\psi_j(t_i) = \langle \ell_i, B\psi_i \rangle$

This means that the functions that would be chosen would be the representers of the evaluation functionals ℓ_i^*. Now, if in much the same way that collocation was generalised to collocation with derivatives or mixed collocation Galerkin's method were generalised to a mixed method, then instead of only the representers of evaluation functions being chosen as the functions $\psi_1, \psi_2, \ldots \psi_n$, they would be chosen as the representers of general functionals ie $\bar{\ell}_1, \bar{\ell}_2, \ldots, \bar{\ell}_n$.

Returning to the collocation method with derivatives and substituting $\bar{\ell}_i$ for ψ_i, $i = 1, 2, \ldots, n$:

$$g_{ij} = \langle \bar{\ell}_i, B\bar{\ell}_j \rangle$$
$$f_i = \langle \bar{\ell}_i, g \rangle$$

Doing the same thing for Galerkin's method

$$g_{ij} = \langle \bar{\ell}_i, B\bar{\ell}_j \rangle$$
$$f_i = \langle \bar{\ell}_i, g \rangle$$

But, for least squares

$$g_{ij} = \langle B\bar{\ell}_i, B\bar{\ell}_j \rangle$$
$$f_i = \langle B\bar{\ell}_i, g \rangle.$$

So it appears at first that the result obtained by least squares may differ from that obtained by the other methods even if the representers of linear functionals are chosen as the functions ψ_j, $j = 1, 2, \ldots n$.

However it is worthwhile to examine the methods further and perhaps work out conditions where they are the same.

Letting $\ell_i = D_i^* \ell_i \quad i = 1, 2, \ldots, n$ as before, for least squares

$$g_{ij} = \langle B\bar{\ell}_i, B\bar{\ell}_j \rangle = \langle BD_i^*\ell_i, BD_j^*\ell_j \rangle$$

$$f_i = \langle B\bar{\ell}_i, g \rangle = \langle BD_i\ell_i, g \rangle$$

Then for $i = 1, 2, \ldots, n$ it is known that

$$c_1 \langle BD_i^*\ell_i, BD_1^*\ell_1 \rangle + c_2 \langle BD_i^*\ell_i, BD_2^*\ell_2 \rangle + \ldots c_n \langle BD_i^*\ell_i, BD_n^*\ell_n \rangle = \langle BD_i^*\ell_i, g \rangle$$

$$\therefore \sum_{j=1}^{n} c_j \langle \ell_i, D_iB^*BD_j^*\ell_j \rangle = \langle \ell_i, D_iB^*g \rangle = D_iB^*g(t_i)$$

$$\therefore \sum_{j=1}^{n} c_j D_iB^*BD_j^*\ell_i^*\{\ell_j\} = D_iB^* \sum_{j=1}^{n} c_j BD_j^*\ell_i^*\{\ell_j\} = D_iB^*g(t_i)$$

In most cases encountered in Geodesy D_i and B are differential or integral operators which possess inverses. So assuming then that the operator D_i^*B has an inverse it follows that

$$\sum_{j=1}^{n} c_j BD_j^* \ell_i^*\{\ell_j\} = g(t_i)$$

Hence

$$\sum_{j=1}^{n} c_j D_i BD_j^*\ell_i^*\{\ell_j\} = D_i g(t_i) = \langle \ell_i, D_i g \rangle$$

$$\therefore \sum_{j=1}^{n} c_j \langle \ell_i, D_iBD_j^* \ell_j \rangle = \sum_{j=1}^{n} c_j \langle D_i^*\ell_i, BD_j^*\ell_j \rangle = \langle D_i^*\ell_i, g \rangle$$

or

$$\sum_{j=1}^{n} c_j \langle \bar{\ell}_i, B\bar{\ell}_j \rangle = \langle \bar{\ell}_i, g \rangle, \quad i = 1, 2, \ldots n$$

which is the set of equations for Galerkin's method as before.

ie., $$c_1 \langle \bar{\ell}_i, B\bar{\ell}_1 \rangle + c_2 \langle \bar{\ell}_i, B\bar{\ell}_2 \rangle + \ldots + c_n \langle \bar{\ell}_i, B\bar{\ell}_n \rangle = \langle \bar{\ell}_i, g \rangle$$

$i = 1, 2, \ldots n.$

That is to say that if D_i, $i = 1, 2, \ldots, n$ and B both possess inverses, then the values of c_i, $i = 1,2, \ldots, n$ obtained by least squares are the same as those obtained by collocation or Galerkin's method provided that $\bar{\ell}_i = \psi_i$, $i = 1, 2, \ldots, n$.

Recalling the formula for the value of some linear functional of u(t) at the point t_p

$$\hat{d}_p = \underset{\sim}{P}^T \underset{\sim}{c} ,$$

it is seen that since $\underset{\sim}{c}$ is the same for all three methods under the conditions stated, the least squares method returns the "observed" value of Bu(t) at $t = t_p$ and so becomes a collocation method without the use of any "unusual" choice of $\underset{\sim}{P}$.

As well, it should be noted that although $\underset{\sim}{c}$ is the same for all three methods, nowhere has it been stated that

$$\langle \bar{\ell}_i, B\bar{\ell}_j \rangle = \langle B\bar{\ell}_i, B\bar{\ell}_j \rangle , \quad \text{or that}$$
$$\langle \bar{\ell}_i, g \rangle = \langle B\bar{\ell}_i, g \rangle$$

In fact it is highly likely that the matrix G in Galerkin's method or for collocation is not symmetric whereas in least squares it must be symmetric. Computational advantages must surely follow from this.

Now all that is necessary to do is to solve the system

$$\underset{\sim}{c} = G^{-1}\underset{\sim}{f}$$

with $G = [g_{ij}]$ where $g_{ij} = \langle B\bar{\ell}_i, B\bar{\ell}_j \rangle$

$$f_i = \langle B\bar{\ell}_i, g \rangle$$

and a collocation as well as a least squares solution is obtained.

In fact since $B\bar{\ell}_i$ is itself a linear functional one may without loss of generality write

$$g_{ij} = \langle \bar{\ell}_i, \bar{\ell}_j \rangle$$
$$f_i = \langle \bar{\ell}_i, g \rangle$$

where the $\bar{\ell}_i$'s are linear functionals.

It is now necessary to find some means of getting numerical values for the inner products.

Before examining this matter in some detail, perhaps it is worthwhile to revisit the method known as least-squares collocation. Here, dealing only with pure collocation,

$$g_{ij} = (B\ell_i, B\ell_j)$$
$$f_i = (B\ell_i, g) \qquad \text{if } \psi_i = \ell_i \qquad i = 1, 2, \ldots, n$$

Now the inner products are different in this case, but by a similar reasoning as that used in least squares, the system may be reduced to one where

$$g_{ij} = (\ell_i, B\ell_j)$$
$$f_i = (\ell_i, g)$$

Expressing this in matrix form,

$$\underset{\sim}{c} = (L^T W H)^{-1} L^T W \underset{\sim}{f}$$

where $L = [\ell_{ij}]$,

$$\ell_{ij} = \langle \ell_i, \ell_j \rangle$$
$$i = 1, 2, \ldots, M$$
$$j = 1, 2, \ldots, n$$

$$H = [h_{ij}]_{Mxn}$$

$$h_{ij} = \langle \ell_i, B\ell_j \rangle$$
$$i = 1, 2, \ldots, M$$
$$j = 1. 2, \ldots, n$$
$$M > n.$$

$$f_i = \langle \ell_i, g \rangle$$

Now where $\underset{\sim}{L}_i$ is the i th column of L it is easily seen that

$$\underset{\sim}{L}_i^T W H (L^T W H)^{-1} L^T W \underset{\sim}{f} = \underset{\sim}{L}_i^T W \underset{\sim}{f} = (\ell_i, g)$$

However, it is the quantity $\langle \ell_i, g \rangle$ which is of interest, ie where

$$\underset{\sim}{P}_i^T = [\langle \ell_i, B\ell_1 \rangle \ \langle \ell_i, B\ell_2 \rangle \ \ldots \langle \ell_i, B\ell_n \rangle],$$

it would be required for a collocation solution that:

$$\underset{\sim}{P}_i^T \underset{\sim}{c} = \underset{\sim}{P}_i^T (L^T W H)^{-1} L^T W \underset{\sim}{f} = f_i = \langle \ell_i, g \rangle .$$

This will happen if $\underset{\sim}{L}_i^T W = \underset{\sim}{r}_i^T$,

where $\underset{\sim}{r}_i$ is a vector with a unity as the i th element and zero elsewhere, for then

$$\underset{\sim}{L}_i^T WH = \underset{\sim}{r}_i^T H = \underset{\sim}{P}_i^T \quad \text{and} \quad \underset{\sim}{L}_i^T W \underset{\sim}{f} = \underset{\sim}{r}_i^T \underset{\sim}{f} = f_i$$

Now if $\underset{\sim}{L}_i^T W = \underset{\sim}{r}_i^T$

$$\underset{\sim}{L}_i^T = \underset{\sim}{r}_i^T W^{-1}$$

ie. $\langle \ell_i, \ell_j \rangle = S_{ij}$

where S_{ij} is the i, j th element of the inverse of the weight matrix.

And indeed $L^T W = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ & & & & & \\ 0 & 0 & 0 & 0 & .1. & 0 \end{pmatrix}_{n \times M}$

$$= [\, I_{n \times n} \,\vdots\, 0_{n \times M-n}]$$

So $L^T WH = G = [g_{ij}]$ $\quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots n.$

where $g_{ij} = \ell_i, B\ell_j$

And $L^T Wf = [\, I_{n \times n} \,\vdots\, 0_{n \times M-n}] \cdot \begin{pmatrix} \langle \ell_1, g \rangle \\ \langle \ell_2, g \rangle \\ \vdots \\ \langle \ell_M, g \rangle \end{pmatrix} = \begin{pmatrix} \langle \ell_1, g \rangle \\ \langle \ell_2, g \rangle \\ \vdots \\ \langle \ell_n, g \rangle \end{pmatrix} = \begin{pmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{pmatrix}$

So the equation for $\underset{\sim}{c}$ becomes

$$c_1 \langle \ell_i, B\ell_1 \rangle + c_2 \langle \ell_i, B\ell_2 \rangle + \ldots + c_n \langle \ell_i, B\ell_n \rangle = \langle \ell_i, g \rangle$$

$$i = 1, 2, \ldots, n.$$

So if the inner product is chosen as above the collocation or Galerkin equations are obtained. And since these equations give the same result for $\underset{\sim}{c}$ as the least squares equations we may say that the system

$$\underset{\sim}{c} = G^{-1}\underset{\sim}{f}$$

where

$$G = [g_{ij}] \, , \; \langle B\ell_i, B\ell_j \rangle = g_{ij}$$

$$f_i = \langle B\ell_i, g \rangle$$

gives an identical result to least squares collocation provided that $\psi_i = \ell_i$ and $\langle \ell_i, \ell_j \rangle = S_{ij}$,

$$i, j = 1, 2, \ldots, n.$$

Since inner products are to be determined as described above it is instructive to evaluate the inner product

$$\langle B\ell_i, B\ell_j \rangle .$$

Now $\langle B\ell_i, B\ell_j \rangle = \langle \ell_i, B^*B \, \ell_j \rangle$

$$= \ell_i^* (B^*B\ell_j)$$

$$= B^*B \, \ell_i^*(\ell_j)$$

$$= B^*B \langle \ell_i, \ell_j \rangle = B^*B(S_{ij})$$

If there were some function S(t,u) such that

$$S_{ij} = S(t,u) \Big|_{\substack{t = t_i \\ u = t_j}} = S(t_i, t_j)$$

with t, u in ∂D then the notation above would mean

$$B^*B \langle \ell_i, \ell_j \rangle = B^* \Big| BS(t,u) \Big|_{t=t_i} \Big| \Big|_{u=t_j}$$

ie. first hold u fixed, perform the operation B on it and evaluate at $t = t_i$. Then perform the operation B^* on the result and evaluate it at $u = t_j$.

S(t,u) would of course have to be an element of the Hilbert space in question whenever either t or u were held fixed. There would also have to be some resemblance to a linear functional in this case. Such functions do exist and will be the topic of the next section.

However, before this is dealt with, a final word on the matrix W. It was stated before that this matrix could perhaps be used to introduce observational noise into the problem. But now it may well be imagined that something more fundamental to the problem than merely observational error is involved. It is conceivable that if the matrix W is the inverse of the covariance matrix of quantities of the form $B^{-1}g$ evaluated or observed at points t_i $i = 1, 2, \ldots, n$ then due to the strong resemblance of the above expression to the matrix expression

$$BW^{-1}B^T,$$

the matrix G would be the covariance matrix of the quantity g evaluated at the points t_i. The covariance matrix would then have much more to do with the signal rather than noise.

But rather than introducing probabalistic aspects to the problem at this stage the problem will be allowed to remain deterministic.

To gain an appreciation of the significance of the function S(t,u) it is necessary to introduce the concept of a Reproducing Kernel.

1.6 Reproducing Kernel Hilbert Spaces

The standard text on kernel functions is (MESCHKOWSKI, 1962) which contains most of the proofs. Unfortunately there does not appear to be an English translation of this work. There are also a number of good references on this topic. (ARONSZAJN, 1950) is perhaps the definitive work, and for quite good sections one may consult (LAURITZEN, 1973), (DAVIS, 1963), (MEISSL, 1975/76), (MORITZ, 1978) or (TSCHERNING, 1978a), to name but a few. Consequently, rather than reproduce proofs from these articles it is intended to list those theorems which may be considered important to this project and attempt to illustrate their meanings by use of simple examples.

Let M be some point set in R^n and let H be a Hilbert space of functions from M to the real numbers R, with inner product denoted by $\langle\ ,\ \rangle$ and with norm $\|f\| = (\langle f,f\rangle)^{\frac{1}{2}}$, for all f in H.

For example M could be the integers, the interval $[a,b] \in R$, a subset of the integers, or D or ∂D from previous sections.

Now let K be a function of two variables

$$K : M \times M \to R$$

ie. $z = K(s,t) \quad z \in R,\ s,t \in M.$

Let s_i now be a fixed value.

Let K^{s_i} denote the function

$$K^{s_i} : M \to R$$

ie. $z = K(s_i,t) \quad z \in R,\ t \in M,\ s_i \in M$

1.6.1 Definition

The function K is said to be a reproducing kernel for the space H if the following properties hold

i) For all s_i in M, K^{s_i} is an element of H

ii) For any function $f \in H$ the reproducing property holds

$$f(s_i) = \langle f, K^{s_i} \rangle \quad \text{for all } s_i \in M$$

or written another way

$$f(s_i) = \langle f(f), K(s_i, t) \rangle$$

for all $s_i \in M$.

1.6.2 Examples

i) Euclidean Vector Space of four dimensions E^4

$M = \{1, 2, 3, 4\}$

$H = \{h: M \to R \mid M = \{1, 2, 3, 4\}\}$

$h = [h(1)\ h(2)\ h(3)\ h(4)]$

eg $h_1 = [15 \quad 18.2 \quad 6\pi \quad 93.8]$

These may be written as vectors and the inner product is

$$\langle h_1, h_2 \rangle = \underset{\sim}{h}_1^T \underset{\sim}{h}_2.$$

The reproducing kernel is $K(i,j) = \delta_{ij}$

$i,j \in M$, the kroneker delta.

$$K(i,j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

This may be written in a matrix form as

$$K(i,j) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Now for property (i) to be satisfied,

$K^i(j)$ should be an element of H

H

eg

$$K^1(j) = K(i,j) = \delta_{ij} = \begin{matrix} 1 & j = 1 \\ 0 & j \neq 1 \end{matrix}$$

So $K^1(1) = 1 \quad K^1(2) = 0 \quad K^1(3) = 0 \quad K^1(4) = 0$

$K^1(j)$ could be written as a vector $[1 \quad 0 \quad 0 \quad 0]^T$.

Similarly $K^2(j)$ $K^3(j)$ and $K^4(j)$ could be written as vectors $[0 \quad 1 \quad 0 \quad 0]^T$, $[0 \quad 0 \quad 1 \quad 0]^T$, $[0 \quad 0 \quad 0 \quad 1]^T$ respectively.

So for all i, $K^i(j) \varepsilon H$.

To check the reproducing property it is required that

$\langle K^i(j), \ h(j) \rangle = h(i)$, or written in vector form with the inner product as described above

$$\langle K^i(j), \ h(j) \rangle = \underset{\sim}{K}^{iT} \underset{\sim}{h} = [K^i(1) \ K^i(2) \ K^i(3) \ K^i(4)] \begin{bmatrix} h(1) \\ h(2) \\ h(3) \\ h(4) \end{bmatrix}$$

So for example $\langle K^3(j), h(j) \rangle = [0 \quad 0 \quad 1 \quad 0] \begin{bmatrix} h(1) \\ h(2) \\ h(3) \\ h(4) \end{bmatrix}$

$= h(3)$ as required.

ii) R^3 with inner product $\underset{\sim}{u}^T P \underset{\sim}{v}$

where $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

$$\text{Here } K(i,j) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \qquad M = \{1, 2, 3\}$$

$$\underset{\sim}{h}^T = [h(1)\ h(2)\ h(3)]$$

$$\text{ie } K(i,j) = \begin{cases} 0 & j \neq i \\ 1 & j = i = 1 \\ 1/2 & j = i = 2 \\ 1/3 & j = i = 3 \end{cases}$$

So for property (i)

$$\underset{\sim}{K}^1 = [1 \quad 0 \quad 0]^T \qquad \underset{\sim}{K}^2 = [0 \quad 1/2 \quad 0]^T$$

and $\underset{\sim}{K}^3 = [0 \quad 0 \quad 1/3]^T$ which are all vectors in R^3.

And for the reproducing property, for example

$$\langle K^2, h \rangle = [0 \quad 1/2 \quad 0] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} h(1) \\ h(2) \\ h(3) \end{bmatrix} = h(2)$$

or in general

$$\langle K^i, h \rangle = h(i) \text{ as required.}$$

iii) $M = N$ the natural numbers $\{1, 2, 3, \ldots\}$

$$H = \{ h : N \to R \mid \sum_{n=1}^{\infty} | h(n) |^2 < \infty \}$$

These functions are the sequences chosen such that they are bounded under the norm defined by the inner product

$$\langle h_1, h_2 \rangle = \sum_{n=1}^{\infty} h_1(n)\ h_2(n)$$

$$\text{e.g. } h_1(n) = \frac{1}{2^n} = \frac{1}{2}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \ldots$$

$$h_2(n) = \frac{1}{n!} = 1, \frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \ldots$$

These functions could be written as (infinite) vectors

$$h_1 = [\, \frac{1}{2} \quad \frac{1}{4} \quad \frac{1}{8} \quad \frac{1}{16} \quad \frac{1}{32} \quad \cdots \,]^T$$

$$h_2 = [\, 1 \quad \frac{1}{2} \quad \frac{1}{6} \quad \frac{1}{24} \quad \frac{1}{120} \quad \cdots]^T$$

with inner product $\langle h_1, h_2 \rangle = \underset{\sim}{h_1}^T \underset{\sim}{h_2}$

$$= \frac{1}{2} + \frac{1}{8} = \frac{1}{48} + \frac{1}{384} + \frac{1}{3840} + \cdots = \sum_{n=1}^{\infty} \frac{1}{2^n n!}$$

Now $K(m,n) = \delta_{mn} = \sum_{i=1}^{\infty} e_i(m)\, e_i(n)$

and $e_i(m) = \delta_{im}$, $e_i(n) = \delta_{in}$ $\qquad m, n \in N.$

K(m,n) could be represented as an infinite order square matrix with unity on the diagonal and zero everywhere else.

Now $K^p(m) = \sum_{i=1}^{\infty} e_i(p)\, e_i(m)$ for a fixed p

ie $K^p(m) = 0 \qquad p \neq m$

$\qquad\qquad\quad 1 \qquad p = m$

which could be written as an infinite vector with a "1" in the p th position and zero everywhere else. So it is an element of H, and property (i) is satisfied.

Now the inner product is

$$\langle K^p, h \rangle = \sum_{n=1}^{\infty} K^p(n)\, h(n)$$

$$= h(p)$$

or in vector notation

$$\langle K^p, h \rangle = K^{pT}h = [0 \;\; 0 \;\; 0 \;\; 0 \;\; \ldots 1 \;\; 0 \ldots] \begin{bmatrix} h(1) \\ h(2) \\ \vdots \\ h(p) \\ \vdots \end{bmatrix}$$

$$= h(p)$$

So the reproducing property is satisfied. The Hilbert space in this example is often called the ℓ^2 space.

iv) Let H consist of functions of the form

$$H(x) = \int_0^x h(t)\, dt$$

where $h(t) \in L^2\,[0, 1]$ ie. $M = [0, 1]$

The inner product is defined by

$$\langle H, G \rangle = \int_0^1 H^1(x)\, G^1(x)\, dx$$

$$= \int_0^1 h(x)\, g(x)\, dx \;, \quad H^1(x) \text{ being the derivative of } H(x), \text{ etc.}$$

Now this is the inner product of $L^2\,[0, 1]$ but H is not $L^2\,[0, 1]$. $L^2\,[0, 1]$ is not really a function space in terms of the definition for here it would be necessary to identify all functions $f(x)$ and $g(x)$ for which

$$\int_0^1 [f(x) - g(x)]^2 dx = 0.$$ ie. it is a set of equivalence classes which, in fact, does not possess a reproducing kernel for reasons which will be apparent later.

However, it is possible to construct one for H.

For $K(x,y)$ to be a reproducing kernel then, holding x fixed $K^x(t)$ must be an element of H.

The inner product of $H(t)$ with $K^x(t)$ is $H(x)$.

$$\text{ie. } H(x) = \langle H(t), K^x(t) \rangle_t$$

$$= \int_0^1 H^1(t)\, K^1(x,t)\, dt$$

But also, by definition

$$H(x) = \int_0^x H^1(t)\, dt$$

$$\text{ie } . \int_0^x H^1(t)\, dt = \int_0^1 H^1(t)\, K^1(x,t)\, dt.$$

$$\int_0^x H^1(t) \circ 1\, dt + \int_x^1 H^1(t) \circ 0\, dt$$

ie. it would be required that

$$K^1(t,x) = \begin{cases} 1 & 0 < t < x \\ 0 & x < t < 1 \end{cases}$$

Now

$$K^1(t,x) = \frac{\partial K(t,x)}{\partial t} \quad \text{and letting}$$

$$K(t,x) = \begin{cases} t & 0 < t < x \\ x & x < t < 1 \end{cases},$$

the results above hold.

1.6.3 Properties of Reproducing Kernels

i) A reproducing kernel is symmetric

$$K(s,t) = \langle K^s(v), K^t(v) \rangle = \langle K^t(v), K^s(v) \rangle$$

$$= K(t,s)$$

1.6.3 ii) A reproducing kernel is positive

ie. $\langle K^s, K^s \rangle = K(s,s) \geq 0$

and equals zero if $K^s(t) = 0$, almost everywhere

so $\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j K(t_j, t_i) \geq 0$

In examples (i) and (ii) then, the matrices would be (at least) positive semi-definite.

Theorem 1.6.1 A necessary and sufficient condition for a Hilbert space H to have a reproducing kernel is that all the evaluation functionals $\ell^*_{t_i}$, $t_i \in M$ defined by

$$\ell^*_{t_i}(h) = h(t_i) \quad \text{for all } h \in H$$

are bounded. Their representers will thus be in H, so they will be members of the dual space H^*.

To illustrate, it was previously mentioned that the space $L^2[0, 1]$ did not have a reproducing kernel. If instead of examining $L^2[0, 1]$, the case of $L^2[-\pi, \pi]$ is taken, the same statement naturally applies.

However, one may ask about the dirac delta function. Surely this is a function from $M \times M$ to R where the reproducing property holds. In fact, for all $h \in H$,

$$\int_{-\pi}^{\pi} h(x)\, \delta(x-t)\, dx = h(t)$$

Unfortunately however, although the reproducing property holds, the first property, which demands that for fixed t (say), $\delta(x-t)$ is an element of H, does not.

For the dirac delta function has the following expansion [see (DAVIS, 1963) or (STAKGOLD, 1967)]:

$$\delta(x-t) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \cos k\,(x-t)$$

Now $$\| K^t(x) \|^2 = \langle K^t(x),\ K^t(x) \rangle = K(t,t)$$

and letting $x = t$

$$\delta(t,t) = \delta(0) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \cos k\,(0) = \infty$$

So for $t = t_i$, $\delta(x-t_i)$ is not square integrable. Also the evaluation functional is unbounded, so $L^2[-\pi,\pi]$ does not have a reproducing kernel. By much the same reasoning $L^2[a,b]$ will not have a reproducing kernel, nor will $L^2[-\infty,\infty]$.

In fact in the case of $L^2[-\infty,\infty]$ the delta function can be shown not to be square integrable more directly. For the case $t = 0$ the δ function may be approximated by a sequence of ordinary functions

$$\delta_n(x) = \sqrt{\frac{n}{\pi}}\, e^{-nx^2} \qquad n = 1, 2, 3 \ldots$$

with $\delta(x)$, being thought of as the limit as n goes to infinity of $\delta_n(x)$

Note that

$$\int_{-\infty}^{\infty} \sqrt{\frac{n}{\pi}}\, e^{-nx^2}\, dx = 1 \qquad n = 1, 2, \ldots$$

$$\text{so } \lim_{n \to \infty} \int_{-\infty}^{\infty} \delta_n(x)\, dx = 1 = \int_{-\infty}^{\infty} \delta(x)\, dx\ .$$

$$\text{However, } \int_{-\infty}^{\infty} \delta_n{}^2(x)\, dx = \frac{n}{\pi} \int_{-\infty}^{\infty} e^{-2nx^2}\, dx$$

$$= \frac{n}{\pi} \sqrt{\frac{\pi}{2n}} = \sqrt{\frac{n}{2\pi}}$$

$$\text{ie} \quad \lim_{n \to \infty} \int_{-\infty}^{\infty} \delta_n(x)\, dx = \infty$$

So the δ function is not square integrable, as before, and thus not an element of H.

Consequently it would be considered inprudent to use $L^2[a, b]$ or $L^2[\ \Omega\ ]$ where Ω is some set in R^n as the space of functions from which the approximate solution of the differential equation will be sought in this work. This is because of the rather attractive properties which Hilbert Spaces with reproducing kernels possess.

Theorem 1.6.2 If H has a reproducing kernel, then the kernel is unique.

Theorem 1.6.3 Let M be an arbitrary set and let $K : M \times M \to R$ be positive and symmetric. Then there exists a Hilbert Space, unique up to isomorphism which has K as its reproducing kernel.

Now since the kernel for such a Hilbert Space is unique by theorem 2, this theorem implies a one to one correspondence between reproducing kernels and Hilbert Spaces for a given set M, at least up to isomorphism.

Any positive symmetric function is thus the reproducing kernel for some Hilbert Space. Given such a function, the Hilbert Space can be constructed as follows:

Since K(s,t) is a reproducing kernel then for any fixed $t_i \in M$, $K^{t_i}(s)$ is an element of some linear space H_0. Now if $a_i \in R$ the linear combinations $\sum_{i=1}^{n} a_i K^{t_i}$ will also be elements of H_0 for $n \in N$.

$$\text{ie.} \quad H_0 = \{ \sum_{i=1}^{n} a_i K^{t_i} \mid a_i \in R, \quad t_i \in M, \quad n \in N \}$$

The inner product may be defined as

$$\left\langle \sum_{i=1}^{n} a_i K^{t_i}, \sum_{j=1}^{n} b_j K^{t_i} \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j K(t_i, t_j)$$

$$n \in N \qquad a_i, b_j \in R \qquad t_i, t_j \in M$$

Let the norm be defined as the square root of the inner product of an element with itself and H_0 is a pre-Hilbert Space. Adding limits of sequences of functions in H_0 makes it a Hilbert Space H.
Let us examine such a sequence of functions $h_1, h_2, h_3, \ldots$ which converges to some element $h \in H$ in the strong sense (convergence in norm)

$$\text{ie.} \quad \lim_{n \to \infty} \| h - h_n \| = 0$$

Since H is a reproducing kernel space,

$$\begin{aligned} | h(s) - h_n(s) |^2 &= | \langle h - h_n, K^s \rangle |^2 \\ &\leq \| h - h_n \| \cdot \| K^s \| \\ &= \| h - h_n \| \cdot \{K(s,s)\}^{\frac{1}{2}} \\ &\leq \| h - h_n \| \cdot c \end{aligned}$$

for any s in M^1 a subset of M and where $| \{K(s,s)\}^{\frac{1}{2}} | \leq c$.
This gives the following result:

Theorem 1.6.4 Let H have a reproducing kernel K(s,t) and let a sequence of functions $h_i(s)$ $s \in M$ converge in the strong sense.

Then there will be pointwise convergence in M. In addition, there will be uniform convergence in any subset M^1 of M where K is bounded on the diagonal of $M^1 \times M^1$.

$$\text{ie} \qquad \lim_{n \to \infty} h_n = h.$$

If, furthermore, the Hilbert Space is separable, one is permitted to write an equality sign between an element and its expansion with respect to an orthonormal basis.

In fact the reproducing kernel and the orthonormal basis are intimately related in a separable reproducing kernel Hilbert Space, as stated in the following theorem:

Theorem 1.6.5 A reproducing kernel Hilbert Space H of functions on M is separable if, and only if, there exists an orthonormal basis e_n, $n \in N$ such that for all $s,t \in M$,

$$K(s,t) = \sum_{n=1}^{\infty} e_n(s)\, e_n(t)$$

The functions e_n will then constitute a complete orthonormal system in H.

Now if e_n, $n \in N$ and f_n, $n \in N$ are both orthonormal systems in H, then due to the uniqueness of the reproducing kernel, and the above theorem,

$$\sum_{n=1}^{\infty} e_n(s)\, e_n(t) = \sum_{n=1}^{\infty} f_n(s)\, f_n(t) = K(s,t)$$

for all s,t in M.

Since it is known that in a separable Hilbert Space with a reproducing kernel there exists a complete orthonormal basis it would be interesting to see if the functions $K(t_i,s)$ $i = 1, 2, \ldots$ would form a complete set in H.

Here is a particular case where this happens:

Let H be a separable Hilbert Space of continuous functions $H : M \to R$ which has a reproducing kernel. Suppose that within M there exists an ε-net of points t_i, $i = 1, 2, \ldots, \infty$. That is to say, suppose the $\{t_i\}$ is everywhere dense in M ie. for any point t and any $\varepsilon > 0$ there exists a point t_i so that the distance defined by the norm is less than ε.

An orthonormal basis may be constructed from the function $K^{t_i}(t)$ $i = 1, 2, \ldots, \infty$ by the Gram-Schmidt method. Let $\phi_1, \phi_2 \ldots$ be the orthonormal basis

$$\phi_1 = a_{11}K^{t_1}(t)$$

$$\phi_2 = a_{21}K^{t_1}(t) + a_{22}K^{t_2}(t)$$

$$\vdots \quad \vdots \quad \vdots$$

$$\phi_n = a_{n1}K^{t_1}(t) + a_{n2}K^{t_2}(t) + \ldots + a_{nn}K^{t_n}(t)$$

$$\phi_{n+1} = a_{n+1,1}K^{t_1}(t) + a_{n+1,2}K^{t_2}(t) + \ldots a_{n+1,n+1}K^{t_{n+1}}(t)$$

It is easy to show that $a_{ii} \neq 0$ for all i in N.

Now let f within H, be some function orthogonal to all ϕ_i, where i is in N.

So $\langle \phi_1, f \rangle = \langle a_{11}K^{t_1}(t), f(t) \rangle = a_{11}f(t_1) = 0$

Since $a_{11} \neq 0$, $f(t_1) = 0$.

Now assume that $f(t_i) = 0$, $i = 1, 2, \ldots, n$.

$$\langle \phi_{n+1}, f \rangle = \langle \sum_{i=1}^{n} a_{n+1,i}K^{t_i}(t) + a_{n+1,n+1}K^{t_{n+1}}(t), f(t) \rangle$$

$$= \sum_{i=1}^{n} a_{n+1,i}f(t_i) + a_{n+1,n+1}f(t_{n+1})$$

$$= a_{n+1,n+1}f(t_{n+1}) = 0$$

And since $a_{ii} \neq 0$ for each i in N, $f(t_{n+1}) = 0$

Thus, by induction $f(t_i) = 0$ for all i in N.

Now since t_i, $i = 1, 2, \ldots, \infty$ is an ε-net for any point t in M a point t_i may be found such that

$$\| t - t_i \| < \varepsilon$$

And by continuity of f a δ may be found such that

$$\| t - x \| < \varepsilon \Rightarrow \| f(t) - f(x) \| < \delta$$

In particular

$$\| t - t_i \| < \varepsilon \Rightarrow \| f(t) - f(t_i) \| < \delta$$

But $f(t_i) = 0$ for all i in N.

Hence $\| f(t) \| < \varepsilon$

Since this is true for any ε, $f(t) \equiv 0$ or $f = 0$.

Thus by (DAVIS, 1975 Theorem 8.9.1) ϕ_i, $i = 1, 2, \ldots, \infty$ is a complete basis for H.

So any f in H can be expressed as a linear combination of ϕ_i's. But these in turn are just linear combinations of the functions K^{t_i} for i in N. So any f in H may be expressed as

$$f = \sum_{i=1}^{\infty} a_i K^{t_i} \quad \text{where } a_i \text{ is a real number}$$

and thus K^{t_i}, i in N forms a complete set in H. This is a most useful result for in approximation it is often necessary to find the projection of f on some subspace of H. K^{t_i}, $i = 1, 2, \ldots, n$ will form a basis for an n dimensional subspace and the projection of f on this subspace will converge pointwise to f as $n \to \infty$.

Before examining this aspect, however, there are some more examples of reproducing kernel Hilbert Spaces to be examined.

1.6.3 iii) Isotropic Kernels on the Unit sphere σ

Let $M = \sigma$, the unit sphere centred at the origin, and H be square integrable functions defined on σ with inner product

$$\langle g, h \rangle = \int_{\sigma} \int g(t)\, h(t)\, dt \qquad t \in \sigma.$$

The space H is thus $L^2[\sigma]$, and the functions f, g can be expanded into a series of Spherical Harmonics.

eg. $$f(\theta,\lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [a_{nm}R_{nm}(\theta,\lambda) + b_{nm}S_{nm}(\theta,\lambda)]$$

where $\pi > \theta > 0$ and $2\pi > \lambda > 0$, θ and λ being the co-latitude and longitude respectively.

$$R_{nm}(\theta,\lambda) = P_{nm}(\cos\theta) \cos m\lambda$$

$$S_{nm}(\theta,\lambda) = P_{nm}(\cos\theta) \sin m\lambda$$

where $P_{nm}(\cos\theta)$ is a solution of Legendre's differential equation, known as a Legendre Function

ie $$\sin\theta\, g''(\theta) + \cos\theta\, g'(\theta) + [n(n+1)\sin\theta - \frac{m^2}{\sin\theta}]g(\theta) = 0$$

has solution

$$g(\theta) = P_{nm}(\cos\theta)$$

Putting $t = \cos\theta$,

$$P_{nm}(t) = \frac{1}{2^n n!}(1-t)^{m/2} \frac{d^{n+m}}{dt^{n+m}}(t^2 - 1)^n$$

When m = 0 the Legendre functions become the well-known Legendre Polynomials $P_n(t)$.

It may be shown that

$$\int_\sigma \int R_{nm}(\theta,\lambda)\, R_{sr}(\theta,\lambda) d\sigma = 0$$

$$\int_\sigma \int S_{nm}(\theta,\lambda)\, S_{sr}(\theta,\lambda) d\sigma = 0 \qquad \text{for } n \neq s \text{ or } r \neq m$$

and

$$\int_\sigma \int R_{nm}(\theta,\lambda)\, S_{sr}(\theta,\lambda)\, d\sigma = 0 \quad \text{always.}$$

Also

$$\int_\sigma \int [R_{no}(\theta,\lambda)]^2\, d\sigma = \frac{4\pi}{2n+1}, \quad \text{and}$$

$$\int_\sigma \int [R_{nm}(\theta,\lambda)]^2 \, d\sigma = \int_\sigma \int [S_{nm}(\theta,\lambda)]^2 \, d\sigma$$

$$= \frac{2\pi}{2n+1} \quad \frac{(n+m)!}{(n-m)!}$$

A complete orthonormal set of functions is obtained by dividing each of the spherical harmonics by its respective norm.

ie. $$\bar{R}_{no}(\theta,\lambda) = \sqrt{\frac{2n+1}{4\pi}} R_{no}(\theta,\lambda) = \sqrt{\frac{2n+1}{4\pi}} P_n(\cos\theta)$$

$$\bar{R}_{nm}(\theta,\lambda) = \sqrt{\frac{2(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} R_{nm}(\theta,\lambda) \quad , m \neq 0$$

$$\bar{S}_{nm}(\theta,\lambda) = \sqrt{\frac{2(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} S_{nm}(\theta,\lambda) \quad , m \neq 0$$

(Note that these definitions differ from those in (HEISKANEN & MORITZ, 1967) where a factor of $\frac{1}{4\pi}$ is used outside the integral sign).

So now $$\int_\sigma \int \bar{R}_{nm}^2 \, d\sigma = \int_\sigma \int \bar{S}_{nm}^2 \, d\sigma = 1.$$

The function $f(\theta,\lambda)$ can be expanded in terms of this orthonormal set as

$$f(\theta,\lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\bar{a}_{nm}\bar{R}_{nm}(\theta,\lambda) + \bar{b}_{nm}\bar{S}_{nm}(\theta,\lambda)]$$

where

$$\bar{a}_{nm} = \int_\sigma \int f(\theta,\lambda) \, \bar{R}_{nm}(\theta,\lambda) \, d\sigma$$

and

$$\bar{b}_{nm} = \int_\sigma \int f(\theta,\lambda) \, \bar{S}_{nm}(\theta,\lambda) \, d\sigma$$

(Here we let $\bar{S}_{no} = 0$)

Since the spherical harmonics are a countable set they can be identified with the natural numbers and so be expressed as functions with a single index.

eg.
$$\phi_1 = \overline{R}_{00}(\theta,\lambda) \qquad \phi_5 = \overline{R}_{20}(\theta,\lambda)$$
$$\phi_2 = \overline{R}_{10}(\theta,\lambda) \qquad \phi_6 = \overline{R}_{21}(\theta,\lambda)$$
$$\phi_3 = \overline{R}_{11}(\theta,\lambda) \qquad \phi_7 = \overline{S}_{21}(\theta,\lambda)$$
$$\phi_4 = \overline{S}_{11}(\theta,\lambda) \qquad \text{etc. (MORITZ, 1966)}$$

Note that $S_{no} = 0$, for all $n \in N$ since $\sin 0\lambda = 0$.

The orthogonal expansion for $f(\theta,\lambda)$ is now

$$f(\theta,\lambda) = \sum_{i=1}^{\infty} \overline{c}_i \, \phi_i(\theta,\lambda)$$

and

$$\langle \phi_i, \phi_\ell \rangle = \int_\sigma \int \phi_i(\theta,\lambda) \, \phi_\ell(\theta,\lambda) d\sigma = \delta_{i\ell}$$

Now, as for $L^2[-\pi\ \pi]$, the space $L^2[\sigma]$ does not possess a reproducing kernel as the evaluation functionals will not have bounded representers within the space. In other words, the series

$$\sum_{i=1}^{\infty} \phi_i(t) \, \phi_i(s), \quad t, s \in \sigma \text{ diverges.}$$

Now
$$\delta(s,t) = \sum_{i=1}^{\infty} \phi_i(t)\phi_i(s) = \sum_{n=0}^{\infty} \overline{R}_{no}(\theta_s,\lambda_s) \, \overline{R}_{no}(\theta_t,\lambda_t)$$
$$+ \sum_{n=1}^{\infty} \sum_{m=1}^{n} [\overline{R}_{nm}(\theta_s,\lambda_s) \, \overline{R}_{nm}(\theta_t,\lambda_t)$$
$$+ \overline{S}_{nm}(\theta_s,\lambda_s) \, \overline{S}_{nm}(\theta_t,\lambda_t)]$$

$$= \frac{1}{4\pi} \sum_{n=0}^{\infty} (2n+1)\, P_n(\cos\theta_s)\, P_n(\cos\theta_t)$$

$$+ \sum_{n=1}^{\infty} \frac{2(2n+1)}{4\pi} \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} [P_{nm}(\cos\theta_s)\, P_{nm}(\cos\theta_t)(\cos m\lambda_t \cos m\lambda_s - \sin m\lambda_t \sin m\lambda_s)]$$

$$= \frac{1}{4\pi} \{1 + \sum_{n=1}^{\infty} (2n+1)[P_n(\cos\theta_s)P_n(\cos\theta_t) + 2 \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_{nm}(\cos\theta_s)P_{nm}(\cos\theta_t)\cos m(\lambda_s$$

$$\therefore \quad \delta(s,t) = \frac{1}{4\pi} + \frac{1}{4\pi} \sum_{n=1}^{\infty} (2n+1)\, P_n(\cos\psi)$$

by the addition rule for spherical harmonics (HEISKANEN & MORITZ, 1967) and where $\cos\psi = \cos\theta_s\cos\theta_t + \sin\theta_t\sin\theta_s\cos(\lambda_t - \lambda_s)$.
Note the similarity between this expression and the expansion of the delta function $\delta(s-t)$ in $L^2[-\pi,\pi]$. Now since $P_0(\cos\psi) = 1$,

$$\delta(s,t) = \frac{1}{4\pi} \sum_{n=0}^{\infty} (2n+1)\, P_n(\cos\psi)$$

Putting $\psi = 0$, $\delta(s,s) = \frac{1}{4\pi} \sum_{n=0}^{\infty} (2n+1) = \infty$

ie. $\| \delta^s \| = \infty$ and the evaluation functionals are unbounded.

Now since $L^2[\sigma]$ does not have a reproducing kernel it would be of benefit to find a space H which is not unlike it but which does have one.

The orthogonal expansion of $L^2[\sigma]$ was

$$f(t) = \sum_{n=1}^{\infty} \bar{c}_n \phi_n(t) \qquad t \in \sigma$$

By Parseval's equality,

$$\| f \|^2_{L^2[\sigma]} = \sum_{n=1}^{\infty} \bar{c}_n^2$$

and indeed, $\langle f,g \rangle = \sum_{n=1}^{\infty} \bar{c}_n \bar{g}_n$.

Let the elements remain the same as those in $L^2[\sigma]$ but construct a new Hilbert Space H with norm defined by

$$\| f \|_H^2 = \frac{1}{4\pi} \sum_{j=1}^{\infty} \frac{2n+1}{k^2_j} \bar{c}_j^2$$

and inner product, the j being related to n by Moritz's numbering:

$$\langle f,g \rangle_H = \frac{1}{4\pi} \sum_{j=1}^{\infty} \frac{2n+1}{k^2_j} \bar{c}_j \bar{g}_j$$

$\phi_j(t)$ $j = 1, 2, \ldots \infty$ are no longer orthonormal since the inner product has been changed. In fact

$$\| \phi_j \|_H = \langle \phi_j, \phi_j \rangle_H^{\frac{1}{2}} = \frac{\sqrt{2n+1}}{k_j \sqrt{4\pi}}$$

An orthonormal basis ψ_j, is formed by dividing ϕ_j by its norm.

ie. $$\psi_j = \frac{k_j \sqrt{4\pi}}{\sqrt{2n+1}} \phi_j$$

Then

$$K(s,t) = \sum_{j=1}^{\infty} \psi_j(s) \psi_j(t)$$

$$= 4\pi \sum_{j=1}^{\infty} \frac{k^2_j}{2n+1} \phi_j(s) \phi_j(t)$$

$$= \sum_{j=0}^{\infty} K^2_j P_j(\cos\psi), \text{ under certain conditions}$$

For $\psi = 0$ $K(s,s) = \sum_{n=1}^{\infty} K^2_j$

So for H to possess a reproducing kernel the constants K_j must be chosen such that

$$\sum_{j=1}^{\infty} K^2_j < \infty .$$

A kernel of this type is known as an isotropic kernel, for it only depends on the spherical distance ψ between the points s and t $\in$ c, and not on the positions of the points (θ_s,λ_s), (θ_t,λ_t).

As will be seen later if the constants K^2_j are chosen to be the "degree-variances", the kernel is the empirical covariance function of a homogeneous and isotropic random field on the sphere.

Note also that the orthonormal expansion in H is

$$f(t) = \sum_{n=1}^{\infty} \hat{c}_j \psi_j \quad \text{ie. the coefficients } \hat{c}_j \text{ become}$$

$$\hat{c}_j = \sqrt{\frac{2n+1}{4\pi}} \frac{\bar{c}_j}{k_j} .$$

The new orthonormal set in terms of R_{nm} and S_{nm} becomes

$$\hat{R}_{no} = k_{no} P_n(\cos\theta)$$

$$\hat{R}_{nm} = k_{nm} \sqrt{\frac{2(n-m)!}{(n+m)!}} R_{nm}$$

$$\hat{S}_{nm} = h_{nm} \sqrt{\frac{2(n-m)!}{(n+m)!}} S_{nm}$$

$$\hat{a}_{nm} = \langle f, \hat{R}_{nm} \rangle_H = \frac{\bar{a}_{nm}}{k_{nm}} \sqrt{\frac{2n+1}{4\pi}}$$

$$\hat{b}_{nm} = \langle f, \hat{S}_{nm} \rangle_H = \frac{\bar{b}_{nm}}{h_{nm}} \sqrt{\frac{2n+1}{4\pi}}$$

The orthogonal expansion is

$$f = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\hat{a}_{nm}\hat{R}_{nm} + \hat{b}_{nm}\hat{S}_{nm}] = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\bar{a}_{nm}\bar{R}_{nm} + \bar{b}_{nm}\bar{S}_{nm}]$$

as is easily verified by substitution.

Now $$K(s,t) = 4\pi \sum_{j=1}^{\infty} \frac{k_j^2}{2n+1} \phi_j(s)\, \phi_j(t) = \sum_{j=0}^{\infty} K_j^2\, P_j(\cos\psi)$$

Since k_j represents h_{nm} or k_{nm} depending on the value of j it is not the same as K_j. Indeed it is not always possible to express K(s,t) in this form.

So now the conditions for which

$$K(s,t) = \sum_{n=0}^{\infty} K_j^2\, P_n(\cos\psi)$$

holds, will be determined.

Now $$K(s,t) = 4\pi \sum_{j=1}^{\infty} \frac{k_j^2}{2n+1} \phi_j(s)\, \phi_j(t)$$

$$= 4\pi \sum_{n=0}^{\infty} \frac{k_{no}^2}{2n+1} \bar{R}_{no}(\theta_s,\lambda_s)\, \bar{R}_{no}(\theta_t,\lambda_t)$$

$$+ 4\pi \sum_{n=1}^{\infty} \sum_{m=1}^{n} \frac{k_{nm}^2}{2n+1} \bar{R}_{nm}(\theta_s,\lambda_s)\, \bar{R}_{nm}(\theta_t,\lambda_t) + \frac{h_{nm}^2}{2n+1} \bar{S}_{nm}(\theta_s,\lambda_s)\, \bar{S}_{nm}(\theta_t,\lambda_t).$$

$$= \sum_{n=0}^{\infty} k_{no}^2\, P_n(\cos\theta_s)\, P_n(\cos\theta_t)$$

$$+ \sum_{n=1}^{\infty} 2\sum_{m=1}^{n} P_{nm}(\cos\theta_s)\, P_{nm}(\cos\theta_t)\, [k_{nm}^2 \cos m\lambda_t \cos m\lambda_s - h_{nm}^2 \sin m\lambda_t \sin m\lambda_s] \frac{(n-m)}{(n+m)}$$

$$= \sum_{n=0}^{\infty} k_{no}^2\, P_n(\cos\theta_s)\, P_n(\cos\theta_t)$$

$$+ \sum_{n=1}^{\infty} 2\sum_{m=1}^{n} k_{nm}^2 P_{nm}(\cos\theta_s)\, P_{nm}(\cos\theta_t)[\cos m\lambda_t \cos m\lambda_s - \sin m\lambda_t \sin m\lambda_s] \frac{(n-m)!}{(n+m)!}$$

$$+ \sum_{n=1}^{\infty} 2\sum_{m=1}^{n} -(h_{nm}^2 - k_{nm}^2)\, P_{nm}(\cos\theta_s)\, P_{nm}(\cos\theta_t) \sin m\lambda_t \sin m\lambda_s \frac{(n-m)!}{(n+m)!}$$

$$= \sum_{n=0}^{\infty} k^2_{no}\, P_n(\cos\theta_s)\, P_n(\cos\theta_t)$$

$$+ \sum_{n=1}^{\infty} 2k^2_{no} \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_{nm}(\cos\theta_s)\, P_{nm}(\cos\theta_t)\, \cos m(\lambda_s-\lambda_t)$$

$$+ \sum_{n=1}^{\infty} 2\sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_{nm}(\cos\theta_s)\, P_{nm}(\cos\theta_t)\, [(k^2_{nm}-h^2_{nm})\, \sin m\lambda_t \sin m\lambda_s$$

$$+ (k^2_{nm}-k^2_{no})(\cos m\lambda_t \cos m\lambda_s - \sin m\lambda_t \sin m\lambda_s)]$$

$$K(s,t) = \sum_{n=0}^{\infty} k^2_{no}\, P_n(\cos\psi)$$

$$+ \sum_{n=1}^{\infty} 2\sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_{nm}(\cos\theta_s) P_{nm}\cos\theta_t)[(k^2_{nm}-k^2_{no})\cos m\lambda_s \cos m\lambda_t$$

$$- (h^2_{nm}-k^2_{no})\sin m\lambda_s \sin m\lambda_t]$$

The first term is the same as the isotropic kernel expression when $k_{no} = K_n$.

The second term is an expression for the anisotropic part of the kernel. It becomes zero when

$$k_{nm} = h_{nm} = k_{no} = K_n \qquad m = 1, 2, \ldots, n, \text{ for all } n.$$

This is the condition for a kernel to be <u>isotropic</u>, when the norm is defined as previously stated.

In other words the norm should be chosen so that

$$\| f \|^2_H = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{2n+1}{K^2_n} \sum_{m=0}^{n} (\bar{a}^2_{nm} + \bar{b}^2_{nm}) \; ;$$

for an isotropic kernel.

Letting

$$\hat{C}^2_n = \frac{1}{4\pi} \sum_{m=0}^{n} (\bar{a}^2_{nm} + \bar{b}^2_{nn})$$

$$\| f \|^2_H = \sum_{n=0}^{\infty} \frac{\hat{C}^2_n}{K^2_n} (2n+1)$$

$$= 2\sum_{n=0}^{\infty} \frac{n\hat{C}^2_n}{K^2_n} + \sum_{n=0}^{\infty} \frac{\hat{C}^2_n}{K^2_n}$$

it is seen that for f to be bounded ΣK^2_n must converge at a slower rate than $\Sigma n\hat{C}^2_n$. But for H to possess a reproducing kernel ΣK^2_n must not diverge. So the values of K_n which may be chosen are limited by these considerations. Note that if K^2_n is chosen to equal $\hat{C}^2$ or even $(2n+1)\ \hat{C}^2_n$, although $\Sigma K^2_n < \infty$ in both cases, f is unbounded. So in this case, K(s,t) would be the reproducing kernel of some Hilbert Space V in which f is not included. ΣK^2_n must converge slowly enough for f to be included if K(s,t) is to be the reproducing kernel of H.

1.6.3 iv) Let M be the space outside a sphere with radius R in R^3, centre at the origin.

Let H be the set of functions harmonic in M and regular at infinity.

For harmonicity it is required that

$$\nabla^2 h = 0 \qquad \text{for all } h \in H$$

or in terms of θ, λ, r, where r is the distance to a point from the origin,

$$r^2 \frac{\partial^2 h}{\partial r^2} + 2r \frac{\partial h}{\partial r} + \frac{\partial^2 h}{\partial \theta} + \cot\theta \frac{\partial h}{\partial \theta} + \frac{1}{\sin^2\theta} \frac{\partial^2 h}{\partial \lambda^2} = 0$$

Regularity at infinity means that

$$\lim_{r \to \infty} rh \quad \text{exists}$$

and $r^2 \frac{\partial h}{\partial r}$, $r^2 \frac{\partial h}{\partial \theta}$ and $r^2 \frac{\partial h}{\partial \lambda}$ are all bounded.

In fact it is shown (MACMILLAN, 1930) that if h is the potential of the sphere and the mass is finite,

$$M = \lim_{r \to \infty} rh \quad \text{is the mass of the sphere.}$$

Note also that since h is harmonic in M, it is also analytic, that is, it is continuous and has continuous derivatives of any order (HEISKANEN & MORITZ, 1967).

Let the norm be defined as

$$\| h \| = \int_M (\nabla h)^2 \, dM \quad \text{with inner product}$$

$$\langle h, g \rangle = \int (\nabla h \cdot \nabla g) \, dM$$

where $\nabla h = \underset{\sim}{i} \frac{\partial h}{\partial x_1} + \underset{\sim}{j} \frac{\partial h}{\partial x_2} + \underset{\sim}{k} \frac{\partial h}{\partial x_3}$

$$\text{ie } (\nabla h)^2 = \left(\frac{\partial h}{\partial x_1}\right)^2 + \left(\frac{\partial h}{\partial x_2}\right)^2 + \left(\frac{\partial h}{\partial x_3}\right)^2$$

Now using Green's Theorem and the fact that $\nabla^2 h = 0$,

$$\int_M (\nabla h)^2 \, dM = -\int_{\sigma_R} \int h \left.\frac{\partial h}{\partial r}\right|_{r=R} d\sigma$$

Where σ_R is a sphere of radius R.

And $$\int_{\sigma_R} \int \overline{R}^2_{nm} = \int_{\sigma_R} \int \overline{S}^2_{nm} = R^2$$

and $$\int_\sigma \int \overline{R}_{nm} \overline{R}_{pq} d\sigma = 0 = \int_\sigma \int \overline{S}_{nm} \overline{S}_{pq} d\sigma = \int_\sigma \int \overline{R}_{nm} \overline{S}_{nm} d\sigma \quad \text{as before.}$$

A solution of Laplace's equation in the region M is given by

$$f(r,\theta,\lambda) = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sum_{m=0}^{n} [\bar{a}_{nm}\bar{R}_{nm}(\theta,\lambda) + \bar{b}_{nm}\bar{S}_{nm}(\theta,\lambda)]$$

So an orthogonal set of solutions is given by

$\psi_j = \left(\frac{R}{r}\right)^{n+1}\phi_j$, $j = 1, 2, \ldots \infty$ where ϕ_j is given by the notation suggested by (MORITZ, 1966).

ie $\left(\frac{R}{r}\right)^{n+1} \bar{R}_{nm}$, $m, n = 0, 1, 2, \ldots \infty$

$\left(\frac{R}{r}\right)^{n+1} \bar{S}_{nm}$, $m, n = 1, 2, \ldots \infty$

$$\int_{\sigma_R}\int - \psi_n \frac{\partial \psi_n}{\partial r}\Big|_{r=R} d\sigma$$

$$= \int_{\sigma_R}\int \left(\frac{R}{r}\right)^{n+1} \bar{R}_{nm} \cdot \frac{R^{n+1}}{r^{n+2}} (n+1) \Big|\bar{R}_{nm}\Big|_{r=R} d\sigma$$

$$= \int_{\sigma_R}\int \frac{R^{n+1}}{r^{n+1}} \bar{S}_{nm} \frac{R^{n+1}}{r^{n+2}} (n+1) \Big|\bar{S}_{nm}\Big|_{r=R} d\sigma$$

$$= \int_{\sigma_R}\int \frac{R^{2n+1}}{R^{2n+2}} (n+1)\, \bar{S}^2{}_{nm}\, d\sigma = \int_{\sigma_R}\int \frac{1}{R}(n+1)\bar{R}^2{}_{nm}\, d\sigma$$

$$= \frac{R^2(n+1)}{R} = R(n+1)$$

So to form an orthonormal set, one must divide each ψ_j by $\sqrt{R(n+1)}$

So $\psi_j = \frac{R^{n+\frac{1}{2}}}{(n+1)^{\frac{1}{2}} r^{n+1}} \phi_j$

To form the reproducing kernel, Theorem 1.6.5 is applied.

ie

$$K(s,t) = \sum_{j=1}^{\infty} \frac{R^{n+1\frac{1}{2}}}{(n+1)^{\frac{1}{2}} r_s^{n+1}} \phi_j(s) \frac{R^{n+1\frac{1}{2}}}{(n+1)^{\frac{1}{2}} r_t^{n+1}} \phi_j(t)$$

$$s,t \in M.$$

$$= \sum_{j=1}^{\infty} \frac{R^{2n+1}}{(n+1) r_s^{n+1} r_t^{n+1}} \phi_j(s) \phi_j(t)$$

$$= \frac{1}{4\pi} \sum_{n=1}^{\infty} \frac{2n+1}{R(n+1)} \left(\frac{R^2}{r_s r_t}\right)^{n+1} P_n(\cos\psi)$$

where ψ is the angle between s and t $\in$ M subtended at the origin, or centre of the sphere.

Note that when $r_s = r_t = R$ and $\psi = 0$

$$K(s,s) = \frac{1}{4\pi R} \sum_{n=1}^{\infty} \frac{2n+1}{n+1} = \infty$$

However when this is so, s and t are no longer in M, and as long as r_s, $r_t > R$ the series converges, so for all s, t $\in$ M, K(s,t) is a reproducjng kernel but not on boundary, ie. the surface of the sphere σ_R.

A closed expression (which has been modified to the present notation) for this kernel is given by (LAURITZEN, 1973):

$$K(s,t) = \frac{1}{4\pi} \left\{ \frac{2}{L} + \frac{1}{R} \ell n \left(\frac{r_s r_t (1-\cos\psi)}{R.L + R^2 - r_s r_t \cos\psi} \right) \right\}$$

$$\text{where } L = \left(\left(\frac{r_s r_t}{R}\right)^2 - 2 r_s r_t \cos\psi + R^2 \right)^{\frac{1}{2}}$$

This kernel is often referred to as the Dirichlet Kernel for this particular space of functions.

This is not the only reproducing kernel for the index set M.

For example (KRARUP, 1969) shows that

$$K(s,t) = \frac{2R}{L} \quad \text{with R and L}$$

representing the same quantities as above and s, t ε M as defined, is the reproducing kernel for the Hilbert Space whose inner product is defined as

$$\langle f,g \rangle = \frac{1}{4\pi} \int_M \frac{1}{r} \nabla f \cdot \nabla g \, dM$$

Another example of a reproducing kernel Hilbert Space on the same set of functions is the Poisson Kernel.

1.6.3 v) Let M and the functions $f : M \to R$ be as defined in 1.6.3 (iv). This time, however the Hilbert Space F will differ from H in that its inner product is defined as

$$\langle f,g \rangle = \lim_{r \to R} \frac{1}{R^2} \int_{\sigma_r} \int f(r_s,\theta_s,\lambda_s)\, g(r_t,\theta_t,\lambda_t)\, d\sigma_r$$

$$\text{where} \quad r = (r_s r_t)^{\frac{1}{2}} > R.$$

An orthonormal set of regular harmonic functions in M is

$$\left(\frac{R}{r_t}\right)^{n+1} \overline{R}_{nm}(\theta_t,\lambda_t) \qquad n, m = 0, 1, 2, \ldots, \infty$$

$$\left(\frac{R}{r_t}\right)^{n+1} \overline{S}_{nm}(\theta_t,\lambda_t) \qquad n, m = 1, 2, \ldots, \infty$$

Note that

$$\int_{\sigma_r} \int [\overline{R}_{nm}(\theta_t,\lambda_t)]^2 d\sigma_r = \int_{\sigma_r} \int [\overline{S}_{nm}(\theta_t,\lambda_t)]^2 d\sigma_r = r^2.$$

So $$\int_{\sigma_r}\int \left(\frac{R}{r_t}\right)^{2n+2}[\bar{R}_{nm}(\theta_t,\lambda_t)]^2 d\sigma_r = \int_{\sigma_r}\int \left(\frac{R}{r_t}\right)^{2n+2}[\bar{S}_{nm}(\theta_t,\lambda_t)]^2 d\sigma_r$$

$$= \left(\frac{R^2}{r_t^2}\right)^{n+1} r^2$$

$$= \left(\frac{R^2}{r^2}\right)^{n+1} r^2 \qquad \text{since } (r_t r_t)^{\frac{1}{2}} = r$$

ie $s = t$

And $$\lim_{r\to R} \frac{1}{R^2}\left(\frac{R^2}{r^2}\right)^{n+1} r^2 = 1$$

Thus the reproducing kernel for F may be found by theorem 1.6.5

$$K(s,t) = \sum_{j=1}^{\infty}\left(\frac{R^2}{r_s r_t}\right)^{n+1}\phi_j(s)\,\phi_j(t)$$

where $\phi_j(s)$, $\phi_j(t)$ are the same as in (MORITZ, 1966)

ie $$K(s,t) = \frac{1}{4\pi}\sum_{n=0}^{\infty}\left(\frac{R^2}{r_s r_t}\right)^{n+1}(2n+1)\,P_n(\cos\psi)$$

And, $$K(s,t) = \frac{1}{4\pi}\left\{\frac{r_s r_t - R^4}{R\,L^3}\right\}$$

with L the same as for the Dirichlet Kernel, is a closed expression for K(s,t).

Again it is noticed that when $r_s = r_t = R, \psi = 0, K(s,t)$ becomes infinite. But, once again, this does not matter because M was the set outside but not including a sphere of radius R in R^3.

Now it is seen that this kernel as well as the Dirichlet Kernel depend on ψ rather than the points (θ_s,λ_s) and (θ_t,λ_t) so in a sense they are isotropic kernels for functions on any sphere of radius $R^1 > R$, when $r_s = r_t = R^1$.

By similar reasoning to that in 1.6.3 (iii) it can be shown that a very general class of kernel is given by

$$K(s,t) = \sum_{n=0}^{\infty} k^2_{no} \left(\frac{R^2}{r_s r_t}\right)^{n+1} P_n(\cos\psi)$$

$$+ \sum_{n=1}^{\infty} 2\left(\frac{R^2}{r_s r_t}\right)^{n+1} \Sigma \frac{(n-m)!}{(n+m)!} P_{nm}(\cos\theta_s) P_{nm}(\cos\theta_t) \times$$

$$[(k^2_{nm} - k^2_{no}) \cos m\lambda_t \cos m\lambda_s - (h^2_{nm} - k^2_{no}) \sin m\lambda_t \sin m\lambda_s)]$$

where the first term is the isotropic part of the kernel and the second term is the anisotropic term which becomes zero when $k_{nm} = h_{nm} = k_{no}$. There is no reason why a Hilbert Space with such an anisotropic reproducing kernel should not be used for functions harmonic and regular outside a sphere, but the class of isotropic kernels certainly seems more "natural" because of its similarity to such kernels as the Dirichlet and Poisson kernels.

In fact, in the above equation when $k_{nm} = h_{nm} = k_{no}$ for all m, the general form of the isotropic kernel for harmonic functions outside a sphere of radius R is

$$K(s,t) = \sum_{n=0}^{\infty} K^2_n \left(\frac{R^2}{r_s r_t}\right)^{n+1} P_n(\cos\psi)$$

Obviously, when

$$K^2_n = \frac{2n+1}{4\pi} ,$$

the Poisson kernel is obtained and when

$$K^2_n = \frac{1}{4\pi} \left\{\frac{2n+1}{R(n+1)}\right\} ,$$

K(s,t) is the Dirichlet kernel.

To obtain the kernel K(s,t) given by (KRARUP, 1969), one makes the substitution

$$K^2_n = 2 \qquad \text{for all } n$$

Then

$$K(s,t) = 2 \sum_{n=0}^{\infty} \left(\frac{R^2}{r_s r_t}\right)^{n+1} P_n(\cos\psi)$$

$$= \frac{2R}{L} \ .$$

If ΣK^2_n converges, then when $r_s = r_t = R$ and $\psi = 0$, K(s,t) is not infinite on the sphere. But one needs to be extremely careful to take notice of what is being modelled in this case. For example ΣK^2_n could converge at such a rate that not only ΣK^2_n converges, but so does $\Sigma K^2_n r^{n+1}$, where $0 < r < 1$.

In this case it would be possible for $r_s r_t < R$ and K(s,t) would still be a reproducing kernel. But $r_s r_t < R$ implies that one or both of r_s, r_t is less than R. So K(s,t) would be the reproducing kernel for H' the space of functions regular and harmonic outside a sphere R' where $R' < R$.

Physical considerations may preclude this from happening, and so the kernel would not really be appropriate.

For K(s,t) to be a reproducing kernel for functions which are harmonic and regular outside a sphere of radius R then, the minimum condition is that

$$\sum_{n=0}^{\infty} K^2_n \left(\frac{R^2}{r_s r_t}\right)^{n+1} < \infty \ .$$

This applies whether the kernel is or is not isotropic.

In addition if the kernel is anisotropic it is required that

$$\sum_{n=1}^{\infty} 2\left(\frac{R^2}{r_s r_t}\right)^{n+1} \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} (k^2_{nm} + h^2_{nm}) < \infty$$

which will happen if k_{nm}, h_{nm} are of the same order as k_{no} (ie K_n) for each n.

The norm of f in a space with an isotropic kernel is given by

$$\| f \|^2 = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{2n+1}{K^2_n} \sum_{m=0}^{n} (\overline{a}^2_{nm} + \overline{b}^2_{nm})$$

For the Poisson kernel, then with $K^2_n = \frac{2n+1}{4\pi}$

$$\| f \|^2 = \sum_{n=0}^{\infty} \sum_{m=0}^{n} (\overline{a}^2_{nm} + \overline{b}^2_{nm})$$

as would be expected, from Parseval's equality.

For the Dirichlet kernel $K^2_n = \frac{2n+1}{4\pi R(n+1)}$

$$\text{So} \quad \| f \|^2 = R \sum_{n=0}^{\infty} (n+1) \sum_{m=0}^{n} (\overline{a}^2_{nm} + \overline{b}^2_{nm})$$

And in the case of the kernel in (KRARUP, 1969)

$$\| f \|^2 = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{2n+1}{2} \sum_{m=0}^{n} (\overline{a}^2_{nm} + \overline{b}^2_{nm})$$

It can be seen that there is considerable variation between the norm of f depending on what reproducing kernel is chosen. In other words the choice of reproducing kernel is really a "norm choice" problem.

1.6.4 Representers of Linear Functionals

Let $H : M \to R$ be a reproducing kernel Hilbert Space of functions. Then, because of the reproducing property

$$\langle K^{s_i}, f \rangle = f(s_i) ,$$

the functions $K^{s_i}(t)$ are obviously the representers of the evaluation functionals.

But these are not the only functionals which exist in the dual space H^* of H.

If B is some linear operator

$$B : H \to H ,$$

then the evaluation of Bf at some point $t_i \in M$ is a linear functional also. Let such a linear functional be called $\bar{\ell}_i^*$ with representer $\bar{\ell}_i$. Then

$$\langle \bar{\ell}_i, f \rangle = Bf(t_i)$$

But since $K^{t_i}(s)$ is the evaluation functional for any $f \in H$, then

$$\begin{aligned} Bf(t_i) &= \langle K^{t_i}(s), Bf(s) \rangle \\ &= \langle B^* K^{t_i}(s), f(s) \rangle \end{aligned}$$

ie $B^*K^{t_i}$ is the representer of the linear functional $\bar{\ell}_i^*$, or

$$\bar{\ell}_i = B^* K^{t_i}$$

where B* is the adjoint of the operator B.

Note that $Bf(t_i) = B \sum_{n=1}^{\infty} c_n \psi_n(s) \Big|_{s=t_i}$

$$= \sum_{n=1}^{\infty} c_n B\psi_n(s) \Big|_{s=t_i}$$

where $\{\psi_n(s)\}$ is an orthonormal basis.

Also $\langle B^*K(t_i,s), f(s) \rangle = Bf(t_i)$

where $K(t,s) = \sum_{n=1}^{\infty} \psi_n(t)\, \psi_n(s)$

$$\text{So } \langle B^*K(t,s), f(s) \rangle = \langle \sum_{n=1}^{\infty} B^*\psi_n(t)\, \psi_n(s), f(s) \rangle$$

$$= \sum_{n=1}^{\infty} \langle \psi_n(s), f(s) \rangle\, B^*\psi_n(t)$$

$$= \sum_{n=1}^{\infty} c_n\, B^*\psi_n(t)$$

$$\text{or} \quad \langle B^* K^{t_i}(s),\ f(s) \rangle = \sum_{n=1}^{\infty} c_n B^* \psi_n(t_i)$$

$$\text{where} \quad B^* \psi_n(t_i) = B^* \psi_n(t)\Big|_{t=t_i}$$

So $B^* K^{t_i}(s)$ is not difficult to determine. One holds s fixed, performs the operation B on K(t,s) as a function of t then lets $t = t_i$ and allows s to become the variable once again.

This is illustrated with the following example:

Let K(s,t) be the Poisson Kernel of example 1.6.3(iii)

Let

$$\Delta g = \frac{\partial f}{\partial r} - \frac{2f}{R} \quad \text{for} \quad f \in H.$$

It is intended to evaluate Δg at the point t_i on the boundary of M ie where $r_{t_i} = R$.

The reproducing kernel can be written in the form

$$K(s,t) = \sum_{j=1}^{\infty} \left(\frac{R^2}{r_s r_t}\right)^{n+1} \phi_j(s)\ \phi_j(t)$$

where $\phi_j(s)$, $\phi_j(t)$ are such that $\left(\frac{R}{r_t}\right)^{n+1} \phi_j(t)$ form an orthonormal set.

Now $$f(r_s,\theta_s,\lambda_s) = \sum_{n=0}^{\infty} \left(\frac{R}{r_s}\right)^{n+1} \sum_{m=0}^{n} [\bar{a}_{nm}\bar{R}_{nm}(\theta_s,\lambda_s) + \bar{b}_{nm}\bar{S}_{nm}(\theta_s,\lambda_s)]$$

which may be written using the notation of (MORITZ, 1966) as

$$f(s) = \sum_{j=1}^{\infty} c_j \left(\frac{R}{r_s}\right)^{n+1} \phi_j(s) \quad \text{where } c_j = \sum_{n=0}^{m} [\bar{a}_{nm} + \bar{b}_{nm}]$$

Now where B is $\frac{\partial}{\partial r_s} - \frac{2}{R}$,

$$Bf(s) = \sum_{j=1}^{\infty} c_j \left\{\frac{n+1}{r_s} - \frac{2}{R}\right\} \left(\frac{R}{r_s}\right)^{n+1} \phi_j(s)$$

Now

$$\langle K^{t_i}(s), Bf(s) \rangle = \sum_{j=1}^{\infty} c_j \left(\frac{R}{r_{t_i}}\right)^{n+1} \phi_j(t_i) \left\langle \left(\frac{n+1}{r_s} - \frac{2}{R}\right)\left(\frac{R}{r_s}\right)^{n+1} \phi_j(s), \left(\frac{R}{r_s}\right)^{n+1} \phi_j(s\right.$$

Now $\left\langle -\frac{2}{R}\left(\frac{R}{r_s}\right)^{n+1} \phi_j(s), \left(\frac{R}{r_s}\right)^{n+1} \phi_j(s) \right\rangle = -\frac{2}{R}$

and $\left\langle \frac{n+1}{r_s}\left(\frac{R}{r_s}\right)^{n+1} \phi_j(s), \left(\frac{R}{r_s}\right)^{n+1} \phi_j(s) \right\rangle = \frac{n+1}{R}$ by

definition of the inner product in example 1.6.3(v).

So $\langle K^{t_i}(s), Bf(s) \rangle = \sum_{j=1}^{\infty} c_j \left(\frac{n+1}{R} - \frac{2}{R}\right) \left(\frac{R}{r_{t_i}}\right)^{n+1} \phi_j(t_i)$

But this must equal $Bf(t_i)$

$$Bf(t_i) = \sum_{j=1}^{\infty} c_j \left(\frac{n+1}{r_{t_i}} - \frac{2}{R}\right)\left(\frac{R}{r_{t_i}}\right)^{n+1} \phi_j(t_i)$$

Thus $r_{t_i} = R$ which is not surprising since t_i was chosen to be on the boundary of M.

So $Bf(t_i) = \sum_{j=1}^{\infty} c_n \left(\frac{n-1}{R}\right) \phi_j(t_i)$,

$$= \sum_{n=0}^{\infty} \frac{n-1}{R} \sum_{m=0}^{n} [\bar{a}_{nm}\bar{R}_{nm} + \bar{b}_{nm}\bar{S}_{nm}]$$

Now

$$B^*K^{t_i}(s) = B_t^* K(s,t)\Big|_{t_i} = \sum_{j=1}^{\infty} \left(\frac{R}{r_s}\right)^{n+1} \phi_j(s) \left(\frac{n+1}{r_{t_i}} - \frac{2}{R}\right)\left(\frac{R}{r_{t_i}}\right)^{n+1} \phi_j(t_i)$$

and $\langle B^*K^{t_i}(s), f(s) \rangle = \sum_{n=1}^{\infty} \left(\frac{n+1}{r_{t_i}} - \frac{2}{R}\right)\left(\frac{R}{r_{t_i}}\right)^{n+1} \phi_j(t_i) \left\langle \left(\frac{R}{r_s}\right)^{n+1} \phi_n(s), c_j \left(\frac{R}{r_s}\right)^{n+1} \phi_j(s\right.$

$$= \sum_{n=1}^{\infty} c_j \left(\frac{n+1}{r_{t_i}} - \frac{2}{R}\right)\left(\frac{R}{r_{t_i}}\right)^{n+1} \phi_j(t_i)$$

$$= Bf(t_i)$$

And since t_i is on the surface of the sphere of radius R

$$\langle B^* K^{t_i}(s),\ f(s) \rangle = \sum_{j=1}^{\infty} c_j \left(\frac{n-1}{R}\right) \phi_j(t_i) \text{ as before.}$$

$$= \sum_{n=0}^{\infty} \frac{n-1}{R} \sum_{m=0}^{n} [\bar{a}_{nm}\bar{R}_{nm} + \bar{b}_{nm}\bar{S}_{nm}]$$

Now that a way of representing linear functionals has been found a method of determining inner products between the representers of these linear functionals is required.

With evaluation functionals there is no problem. For, by the definition of a reproducing kernel,

$$\langle K^{s_i}(t),\ K^{s_i}(t) \rangle = \langle K(s_i,\ t),\ K^{s_j}(t) \rangle$$

$$= K^{s_j}(s_i) = K(s_i, s_j).$$

Now if B_1 and B_2 are two linear operators and

$$\bar{\ell}_1^*(f) = \ell_1^*(Bf) = B_1 f(t_1) = \langle B_1^* K^{t_2}, f \rangle$$

$$\bar{\ell}_2^*(f) = B_2 f(t) = \langle B_2^* K^{t_1},\ f \rangle ,$$

then $$\langle \bar{\ell}_1, \bar{\ell}_2 \rangle = \langle B_1^* K^{t_1},\ B_2^* K^{t_2} \rangle$$

$$= \langle B_2 B_1^* K^{t_1},\ K^{t_2} \rangle = C(t_1,\ t_2), \quad \text{say}$$

which may be written $B_2 B_1^* K(t_1, t_2)$ ie operate in the manner previously described for B_1^* on $K(s,t)$ with s as the variable and evaluate at $t = t_1$, giving $C^{t_1}(t)$. Then operate using B_2 with $C^{t_1}(t)$ as a function of t, then evaluate at t_2 to give $C(t_1, t_2)$.

It is possible therefore to find the norm of an evaluation or other functional.

For example using the closed expression for the Poisson Kernel,

$$K(s,t) = \frac{R^2(r_s^2 r_t^2 - R^4)}{\{r_s^2 r_t^2 - 2R^2 r_s r_s \cos\psi + R^4\}^{3/2}}$$

Then $$\| K^{t_i} \|^2 = \langle K^{t_i}(s), K^{t_i}(s) \rangle = K(t_i,t_i)$$

$$= \frac{R^2(r_{t_i}^2 + R^2)}{(r_{t_i}^2 - R^2)^2}$$

Note that as $r_{t_i} \to R$ then $\| K^{t_i}(s) \| \to \infty$.

With $$B = -\frac{\partial}{\partial r} - \frac{2}{R}$$

$BB^* K(t_i,t_j)$ may be determined.

The norm of such a functional is given by

$$\| \bar{\ell}_i^* \|^2 = BB^* K(t_i,t_j)$$

$$= \frac{1}{4\pi} \sum_{n=0}^{\infty} (2n+1) \left(\frac{n-1}{R}\right)^2 = \infty.$$

when $r_s = r_t = R$.

So $\bar{\ell}_i^*$ is also unbounded on the sphere.

Since it will be required that $\langle \bar{\ell}_i, \bar{\ell}_j \rangle$ be evaluated on this boundary it is evident that the Poisson Kernel will be inadequate for the approximate solution of differential equations with boundary values.

A similar computation would reveal the same deficiency with the Dirichlet and Krarup Kernels.

Thus it is necessary to look to the general isotropic kernel for the evaluation and other functionals.

$$K(s,t) = \sum_{n=0}^{\infty} K_n^2 \left(\frac{R^2}{r_s r_t}\right)^{n+1} P_n(\cos\psi)$$

$$\| K^{t_i}(s) \|^2 = \sum_{n=0}^{\infty} K_n^2$$

and $$\| \bar{\ell}_i^* \|^2 = \| \bar{\ell}_i \|^2 = \sum_{n=0}^{\infty} K_n^2 \left(\frac{n-1}{R}\right)^2$$

ie. in this case it is not only required that ΣK_n^2 converges, but also that $\Sigma K_n^2 n^2$ converges too. Indeed, before choosing the Hilbert Space it may be necessary to look at the norms of all linear functionals which are to be used in the solution.

For all candidate operators B_i it would also be advisable to examine the norm $B_i f$ as well as that of f.

In the case above $\Delta g = Bf$, the norms are given by

$$\| f \|^2 = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{2n+1}{K_n^2} \sum_{m=0}^{n} (\bar{a}_{nm}^2 + \bar{b}_{nm}^2)$$

and $$\| \Delta g \|^2 = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{(2n+1)(n-1)}{K_n^2 R} \sum_{m=0}^{n} (\bar{a}_{nm}^2 + \bar{b}_{nm}^2)$$

In this case it can be seen that K_n^2 should converge at a slower rate than

$\sum_{n=0}^{\infty} n^2 \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]$, which is even slower than that required for $\| f \|^2$. So the choice of the constants K_n may be more severely limited when functionals of f are taken into account.

Now that the form of the linear functionals and a method of finding inner products between them has been determined, it is necessary to examine finite dimensional subspaces H_n of a reproducing kernel Hilbert Space H, for the test function in the approximate solution of differential equations is of the form

$$u^* = c_1\psi_1 + c_2\psi_2 + \ldots + c_n\psi_n ,$$

and $\psi_1, \psi_2, \ldots, \psi_n$ can only be the basis for some finite dimensional space.

1.6.5 Projections and Subspaces

Since the evaluation functionals are bounded, then if a subspace $H' \subset H$ is chosen, obviously they will remain bounded for this subspace:

Theorem 1.6.6

If H' is a closed subspace of H and H is a reproducing kernel Hilbert Space, then H' also has a reproducing kernel.

A subspace of a Hilbert Space will, of course, have the same inner product. Now if $\psi_1, \psi_2, \ldots$ is an orthonormal basis for H, then $\psi_1, \psi_2, \ldots \psi_n$ will be an orthonormal basis for H_N, an N-dimensional subspace of the separable space H. Thus any element $h_N \in H_N$ can be represented by

$$h_N(s) = \sum_{m=1}^{N} c_n \psi_m(s)$$

Now $$\left\langle \sum_{n=1}^{N} \psi_n(s)\psi_n(t), \sum_{m=1}^{N} c_m \psi_m(s) \right\rangle$$

$$= \sum_{m=1}^{N} c_m \left\langle \sum_{n=1}^{N} \psi_n(s)\psi_n(t), \psi_m(s) \right\rangle$$

$$= \sum_{m=1}^{N} c_m \psi_m(t)$$

So the reproducing kernel of the space H_N is given by

$$K'(s,t) = \sum_{n=1}^{N} \psi_n(s)\, \psi_n(t)$$

Now for some $h \in H$ it is seen that since $h(s) = \sum_{n=1}^{\infty} c_n \psi_n(s)$

$$\langle K'(s,t),\ \ell(s) \rangle = \langle \sum_{m=1}^{N} \psi_m(s)\ \psi_m(t),\ \sum_{n=1}^{\infty} c_n \psi_n(s) \rangle$$

$$= \sum_{m=1}^{N} \sum_{n=1}^{\infty} c_n \langle \psi_m(s)\ \psi_m(t), \psi_n(s) \rangle$$

$$= \sum_{m=1}^{N} c_m \psi_m(t)$$

But this is the projection of h onto the subspace H_N.

Theorem 1.6.7

If K' is the kernel of H' $\subset$ H and h is an element of h, then the projection P of h on H is found as

$$Ph(t) = \langle K'^{t}, h \rangle$$

That this is not restricted to separable spaces is shown in (LAURITZEN, 1973).

It can be seen that if

$$P_N h(t) = \sum_{m=1}^{N} c_m \psi_m(t)\ ,$$

as $N \to \infty$, $P_N h(t) \to h(t)$, since the constants c_i are the same for each i in both expressions.

Returning to the approximate solution of differential equations it is easily seen that the quantity u^* obtained is the projection of u on an n dimensional subspace H_n of H. This is because if the representors of evaluation functionals are chosen as the basis vectors for H_n, the other methods are equivalent to the least squares method. The least squares method minimises the norm

$$\| u-u^* \|$$

where $u^* = c_1\bar{\ell}_1 + c_2\bar{\ell}_2 + \ldots c_n\bar{\ell}_n$

Now $\langle u^*, u \rangle = \langle c_1\bar{\ell}_1 + c_2\bar{\ell}_2 + \cdots + c_n\bar{\ell}_n, f \rangle$

$$= \underset{\sim}{c}^T \begin{bmatrix} \langle \ell_1, f \rangle \\ \langle \ell_2, f \rangle \\ \vdots \\ \langle \ell_n, f \rangle \end{bmatrix} = \underset{\sim}{c}^T \underset{\sim}{f}$$

and $$\langle u^*, u^* \rangle = \underset{\sim}{c}^T \begin{bmatrix} \langle \bar{\ell}_1, \bar{\ell}_1 \rangle & \langle \bar{\ell}_2, \bar{\ell}_1 \rangle \cdots \langle \bar{\ell}_n, \bar{\ell}_1 \rangle \\ \langle \bar{\ell}_1, \bar{\ell}_2 \rangle & \langle \bar{\ell}_2, \bar{\ell}_2 \rangle \cdots \langle \bar{\ell}_n, \bar{\ell}_2 \rangle \\ & \\ \langle \bar{\ell}_1, \bar{\ell}_n \rangle & \langle \bar{\ell}_2, \bar{\ell}_n \rangle \cdots \langle \bar{\ell}_n, \bar{\ell}_n \rangle \end{bmatrix} \underset{\sim}{c}$$

$$= \underset{\sim}{c}^T G \underset{\sim}{c}$$

where G and $\underset{\sim}{f}$ are the same as in the section on least squares solutions.

Now when a least squares solution is obtained, it is given by the matrix equation

$$\underset{\sim}{c} = G^{-1}\underset{\sim}{f}$$

Thus $\langle u^*, u \rangle = \underset{\sim}{f}^T G^{-1} \underset{\sim}{f}$

and $\langle u^*, u^* \rangle = \underset{\sim}{f}^T G^{-1} G\, G^{-1} \underset{\sim}{f} = \underset{\sim}{f}^T G^{-1} \underset{\sim}{f}$

So $\langle u^*, u \rangle - \langle u^*, u^* \rangle = 0 = \langle u^*, u-u^* \rangle$

Thus $u-u^*$ is orthogonal to $u^* \in H_n$, and so u^* is the orthogonal projection of u on H_n.

So if $\bar{\ell}_1, \bar{\ell}_2, \ldots$ form a complete basis for H and $\bar{\ell}_1, \bar{\ell}_2, \ldots, \bar{\ell}_n$ form a linearly independent basis for H_n, in view of the fact that $u^* = \sum_{i=1}^{N} c_i'\psi_i$ where ψ_i is an orthonormal basis for H_n, and that as $N \to \infty$, $u^* \to \sum_{i=1}^{\infty} c_i'\psi_i = u$, then $u^* = \sum_{i=1}^{N} c_i\bar{\ell}_i \to u = \sum_{i=1}^{\infty} c_i\bar{\ell}$, as $N \to \infty$.

Now if the kernel functions $K^{t_i}(s)$ $i = 1,2, \ldots n$ are chosen as the basis of H_n, since it has been previously shown that if all $f \in H$ are continuous $K^{t_i}(s)$ $i = 1,2, \ldots \infty$ form a complete set of functions, then it must follow that as $n \to \infty$ $u^* = \sum_{i=1}^{n} c_i K^{t_i}(s)$ converges pointwise to u.

Naturally the same result will follow if $B_i^* K^{t_i}(s)$ $i = 1,2, \ldots \infty$ are chosen as the basis vectors for H.

It is also of importance that, as stated in (HEISKANEN & MORITZ, 1967), if u is a function harmonic outside a sphere of radius R and regular at infinity, it is continuous and infinitely differentiable, as are all harmonic functions.

1.7 Choice of H_n - a summary

The problem:

$Au = 0$ in D, is a partial differential equation

$Bu = g$ on ∂D, the boundary of D, is the boundary condition

where $\psi_1, \psi_2, \ldots \psi_n$ are solutions of the differential equation, the approximate solution is of the form:

$u^* = c_1\psi_1 + c_2\psi_2 + \ldots c_n\psi_n$ and this must satisfy the boundary conditions.

Let H be a reproducing kernel Hilbert Space with kernel K(s,t), $u \in H$, $\psi_i \in H$, $i = 1, \ldots, n$. Let H_n be an n dimensional subspace of H with a basis given by $\psi_1, \psi_2, \ldots \psi_n$.

Solutions are given by $\underset{\sim}{c} = G^{-1}\underset{\sim}{f}$ with G and $\underset{\sim}{f}$ given as

(i) Least Squares

$$g_{ij} = \langle B\psi_i, B\psi_j \rangle \, , \quad f_i = \langle B\psi_i, g \rangle$$

(ii) Galerkin's Method

$$g_{ij} = \langle \psi_i, B\psi_j \rangle \quad f_i = \langle \psi_i, g \rangle$$

(iii) Collocation

$$g_{ij} = \langle \ell_i, B\psi_j \rangle \quad f_i = \langle \ell_i, g \rangle$$

Now since H is a reproducing kernel Hilbert Space the evaluation functionals ℓ_i^* where $\ell_i^*(f) = f(t_i)$, have as their representers $K^{t_i}(s)$, the kernel functions.

Also since $K^{t_i}(s)$ $i = 1,2, \ldots, n$ form a complete set of functions for H if all $u \in H$ are continuous, $K^{t_1}(s)$, $K^{t_2}(s)$, ..., $K^{t_n}(s)$ may be chosen as the basis vectors for H_n and u^* will converge pointwise to u as $n \to \infty$.

This yields the solutions

(ia) Collocation

$$g_{ij} = \langle K^{t_i}, BK^{t_i} \rangle = BK(t_i,t_j) \quad , \quad f_i = \langle K^{t_i}, g \rangle = g(t_i) \quad = \langle K^{t_i}, Bu \rangle = Bu(t_i)$$

(iia) Galerkin's method

$$g_{ij} = \langle K^{t_i}, BK^{t_i} \rangle = BK(t_i,t_j) \quad , \quad f_i = \langle K^{t_i}, g \rangle = g(t_i) \quad = Bu(t_i)$$

(iiia) Least Squares

$$g_{ij} = \langle BK^{t_i}, BK^{t_j} \rangle = B^*B\,K(t_i,t_j) \quad f_i = \langle BK^{t_i}, g \rangle = B^*g(t_i) \quad = \langle K^{t_i}, B^*g \rangle = B^*Bu(t_i)$$

Now if B possesses an inverse, then so does B^* and B^*B so all three methods can be unified to give the same equation.

$$g_{ij} = K(t_i,t_j) \qquad f_i = u(t_i)$$

$$\text{ie} \quad g_{ij} = \langle K^{t_i}, K^{t_j} \rangle \qquad f_i = \langle K^{t_i}, u \rangle$$

On this equation which is an equation for pure collocation, least-squares, etc. the "mixed" methods or methods with derivitives or operators may be based.

So if the quantity $B_i u \quad i = 1,2, \ldots n$ is known at n points t_i of M where $H : M \to u$ is a reproducing kernel Hilbert Space, instead of choosing $K^{t_i}(s)$ as the basis, $\overline{\ell}_i(s) = B_i K^{t_i}(s)$ may be chosen.

The equations then are equivalent to

$$g_{ij} = \langle B_i K^{t_i}, B_j K^{t_j} \rangle \qquad f_i = \langle B_i K^{t_i}, u \rangle$$

ie the elements of the matrix G are

$$B_j^* B_i \, K(t_i,t_j)$$

and of the vector f are $B_i u(t_i)$.

There is an additional benefit in the use of a reproducing kernel Hilbert Space. Not only may boundary conditions on ∂D be used to determine G and $\underset{\sim}{f}$, but also other relationships in the space D may be used. There is a problem in some cases with this method. If M excludes the boundary ∂D of D where $H : M \to u$ is the Hilbert Space, there will be problems with unboundedness on ∂D.

Two possible solutions present themselves. One is to choose M such that ∂D is included within it as well as D, in which case there may be problems in relating the mathematical model to the physical reality. For example in Geodesy a "Bjerhammar Sphere" totally within the earth may be chosen, outside which the functions u are said to be harmonic and regular. There is no problem then with unboundedness of u on the surface of the earth which is then well within M, and thus outside

the sphere of radius R. But between the Bjerhammer Sphere and the surface of the earth there is a definite discrepancy between the mathematical model and the physical reality, since it assigns the property of harmonicity and regularity to potentials at points on and within the earth's crust.

The other approach is to attempt by some judicious choice of the reproducing kernel to make the norms of u, and $B_i u$, $i = 1,2, \ldots n$ converge as well as making the norms of the linear functionals $B_i K^{t_i}$ converge on ∂D.

This choice of reproducing kernel is really a choice of the inner product for the space H and is regarded as a "Norm Choice" problem.

So provided that the operators B_i are not singular, the only problem remaining in the choice of the space H_n and a basis for this space is this norm choice problem. To get some idea of what is involved, the concept of error bounds must be examined.

1.8 Error Bounds

It is possible to find bounds which the difference in value of some linear functional evaluated at a point and its estimated or approximated value cannot exceed.

Let $\bar{\bar{\ell}}_p$ be the representer of a functional $\bar{\bar{\ell}}_p^*$. This may or may not be some functional already used in determining the matrix G, such functionals being denoted $\bar{\ell}_i$, $i = 1,2, \ldots, n$.

Let the function u be estimated or approximated by the function u^* where

$$u^* = \sum_{i=1}^{n} c_i \bar{\ell}_i \quad , \text{ as before .}$$

So $\bar{\bar{\ell}}_p^*(u)$ is approximated by $\bar{\bar{\ell}}_p^*(u^*)$.

The absolute value of the difference is

$$|\bar{\bar{\ell}}_p^*(u) - \bar{\bar{\ell}}_p^*(u^*)| = |\bar{\bar{\ell}}_p^*(u-u^*)| = \langle \bar{\bar{\ell}}_p, u - u^* \rangle$$

Now $\langle \bar{\bar{\ell}}_p, u^* \rangle = \langle \bar{\bar{\ell}}_p, \sum_{i=1}^{n} c_i \bar{\ell}_i \rangle = \sum_{i=1}^{n} c_i \langle \bar{\bar{\ell}}_p, \bar{\ell}_i \rangle$

And where $\underset{\sim}{p}^T = [\langle \bar{\bar{\ell}}_p, \bar{\ell}_1 \rangle \langle \bar{\bar{\ell}}_p, \bar{\ell}_2 \rangle \ldots \langle \bar{\bar{\ell}}_p, \bar{\ell}_n \rangle]$,

$\langle \bar{\bar{\ell}}_p, u^* \rangle = \underset{\sim}{p}^T \underset{\sim}{c} = \underset{\sim}{p}^T G^{-1} \underset{\sim}{f}$ as before,

where $f^T = [\langle \bar{\ell}_1, u \rangle \langle \bar{\ell}_2, u \rangle \ldots \langle \bar{\ell}_n, u \rangle]$

Let $\underset{\sim}{p}^T G^{-1} = \underset{\sim}{d}^T$

Then $\langle \bar{\bar{\ell}}_p, u \rangle = \underset{\sim}{d}^T \underset{\sim}{f} = \sum_{i=1}^{n} d_i f_i = \langle \sum_{i=1}^{n} d_i \bar{\ell}_i, u \rangle$

Denoting $\sum_{i=1}^{n} d_i \bar{\ell}_i$ by $\hat{\ell}_p$ it is seen that

$$|\bar{\bar{\ell}}_p^*(u) - \bar{\bar{\ell}}_p^*(u^*)| = \langle \bar{\bar{\ell}}_p, u - u^* \rangle = \langle \bar{\bar{\ell}}_p - \hat{\ell}_p, u \rangle$$

$$\leq \|u\| \cdot \|\bar{\bar{\ell}}_p - \hat{\ell}_p\|$$

$$= \|u\| \cdot | \{ \|\bar{\bar{\ell}}_p\|^2 - \underset{\sim}{p}^T G^{-1} \underset{\sim}{p} \}^{\frac{1}{2}} |$$

Now if $\bar{\bar{\ell}}_p$ is denoted by $D^* \ell_p$ where D^* is an operator and ℓ_p is the representer of an evaluation functional,

$$|\bar{\bar{\ell}}_p^*(u) - \bar{\bar{\ell}}_p(u^*)| < \|u\| \cdot | \{ DD^* K(t_p, t_p) - \underset{\sim}{p}^T G^{-1} \underset{\sim}{p} \}^{\frac{1}{2}} | ,$$

$K(t,s)$ being the reproducing kernel of the space of functions $\{u\}$ and the operation DD^* being performed as previously described.

In particular for "pure" collocation without derivatives ie. prediction,

$$| u(t_p) - u^*(t_p) | < \|u\| \cdot | \{ K(t_p, t_p) - \underset{\sim}{p}^T G^{-1} \underset{\sim}{p} \}^{\frac{1}{2}} | = \|u\| \sigma .$$

This, with appropriate notation changes, is the expression given by (TSCHERNING, 1978). (DERMANIS, 1977) gives what he believes to be a better bound.

Since $u = u_o + u^*$, where u_o is the residual $u - u^*$,

$$\langle \bar{\bar{\ell}}_p - \hat{\ell}_p, u \rangle = \langle \bar{\bar{\ell}}_p - \hat{\ell}_p, u_o + u^* \rangle .$$

Now $$\langle \bar{\bar{\ell}}_p, u^* \rangle = \langle \sum_{i=1}^{n} c_i \bar{\ell}_i, \bar{\bar{\ell}}_p \rangle = \underset{\sim}{p}^T \underset{\sim}{c} = \underset{\sim}{p}^T G^{-1} \underset{\sim}{f} .$$

and $$\langle \hat{\ell}_p, u^* \rangle = \langle \sum_{i=1}^{n} d_i \bar{\ell}_i, \sum_{j=1}^{n} c_j \bar{\ell}_j \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \langle \bar{\ell}_i, \bar{\ell}_j \rangle$$

$$= \underset{\sim}{d}^T G \underset{\sim}{c} = \underset{\sim}{p}^T G^{-1} \underset{\sim}{f}$$

So $$\langle \bar{\bar{\ell}}_p - \hat{\ell}_p, u^* \rangle = 0$$

and $$\langle \bar{\bar{\ell}}_p - \hat{\ell}_p, u \rangle = \langle \bar{\bar{\ell}}_p - \hat{\ell}_p, u_o \rangle \leq \| u_o \| \cdot \| \bar{\bar{\ell}}_p - \hat{\ell}_p \|$$

$$\leq \| u \| \cdot \| \bar{\bar{\ell}}_p - \hat{\ell}_p \|$$

ie. it is a lower bound since by Pythagoras,

$$\| u \|^2 = \| u_o \|^2 + \| u^* \|^2 \quad \text{ie} \quad \| u_o \| \leq \| u \| .$$

Now $$\| u_o \|^2 = \langle u - u^*, u - u^* \rangle = \| u \|^2 - \underset{\sim}{f}^T G^{-1} \underset{\sim}{f}$$

So a bound is

$$| \bar{\bar{\ell}}_p^*(u) - \bar{\bar{\ell}}^*(u^*) | \leq | \{ \| u \|^2 - \underset{\sim}{f}^T G^{-1} \underset{\sim}{f} \}^{\frac{1}{2}} \{ DD^* K(t_p, t_p) - \underset{\sim}{p}^T G^{-1} \underset{\sim}{p} \}^{\frac{1}{2}} |$$

For a Hilbert space H(K) with a reproducing kernel which is isotropic ie

$K(\psi_{s,t}) = \sum_{n=0}^{\infty} K_n^2 \, P_n(\cos\psi_{s,t})$, the norm of u can be

obtained from the expression

$$\| u \|^2 \quad \frac{1}{4\pi} \sum_{n=0} \frac{(2n+1)}{K_n^2} \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2] .$$

Letting $\sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2] = S_n^2$ the following expression is obtained for the prediction case:

$$| u(t_p) - u^*(t_p) | \leq |\{\frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{(2n+1)S_n^2}{K_n^2} - \underset{\sim}{f}^T G^{-1} \underset{\sim}{f}\}^{\frac{1}{2}} \{ \sum_{n=0}^{\infty} K_n^2 - \underset{\sim}{p}^T G^{-1} \underset{\sim}{p}\}^{\frac{1}{2}}|$$

This function, according to Dermanis, is a tool for comparing kernels ie. if there are two kernels $K_1(s,t)$ and $K_2(s,t)$ then K_1 is better than K_2 for predicting a certain quantity if the expression above when computed for K_2 is small than that for K_1, for the particular data set which gives the vector $\underset{\sim}{f}$. An optimal kernel from a particular class of kernels would give the minimum for this expression.

He also makes the point that one is usually interested in not one, but a number of predictions and proposes the possible use of a risk function to allow for this, and mentions the possibility of using the density function of the mass of the earth in obtaining values for the bound.

However, in this project it is proposed to investigate a slightly different line of reasoning. Since it is mathematically easier to handle, only expressions for "pure" prediction, eg. potentials from potentials, gravity anomalies from gravity anomalies etc., will be dealt with. Examining the expression for $|u(t_p) - u^*(t_p)|$ it is seen that the terms $\underset{\sim}{f}^T G^{-1} \underset{\sim}{f}$ and $\underset{\sim}{p}^T G^{-1} \underset{\sim}{p}$ contain expressions for the kernel $K(s,t)$, for:

$$g_{ij} = K(t_i,t_j) \quad \text{and} \quad p_i = K(t_p,t_i).$$

What is required is that for a given data set (ie. a given $\underset{\sim}{f}$) the predictions should <u>in general</u> be in some way optimal.

It is sensible to assume a given or constant data set because the problem that one is usually posed is that of predicting some quantity or a number of quantities from the available data ie. finding the best predicted value from a given data set. $\underset{\sim}{f}$ can thus be regarded as a constant vector, and G can be regarded as a matrix whose terms depend on the coefficients K_n alone ie. in the isotropic case

$$g_{ij} = \langle \ell_i, \ell_j \rangle = K(\psi_{t_i,t_j}) = \sum_{n=0}^{\infty} K_n^2 \, P_n(\cos\psi_{t_i,t_j}) \text{ , varies}$$

with K_n, n=0,1,2, ... and not with ψ_{t_i,t_j} .

It is noticed at this stage that the term

$$\{ \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{(2n+1)}{K_n^2} - \underset{\sim}{f}^T G^{-1} \underset{\sim}{f} \} \text{ depends only on the } K_n\text{'s}$$

since the $S_n^2 = \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]$, n = 1,2, ... are constant for a particular u.

The term $\{ \sum_{n=0}^{\infty} K_n^2 - \underset{\sim}{p}^T G^{-1} \underset{\sim}{p} \}$, however, is dependent on the point t_p, ie. the particular prediction, as well as the K_n's. So it is conceivable that for two kernels K_1 and K_2, K_1 could give the best prediction at the point t_p whilst K_2 could be better at t_q.

It would seem more sensible to consider all possible predictions and choose a kernel such that the mean square error of prediction is minimised. Perhaps some expression such as

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \{ u(t_{p_i}) - u^*(t_{p_i}) \}^2$$

could be used, but a more fundamental expression is $\| u - u^* \|$.
In fact this expression has already been used as the minimisation criterion for least squares, etc., on the assumption that the kernel is known.

Now imagine that predictions have been carried out with a number of different kernels. Surely that kernel which yields the smallest value for $\| u - u^* \|$ is most preferable. Surely if the minimisation of $\| u - u^* \|$ is a good criterion for derivation of the prediction equations it is also a good criterion for choosing an optimal kernel. Here it will be assumed that it is.

Now the minimisation of $\| u - u^* \|$ is equivalent to the minimisation of its square ie. the function to be minimised is

$$\| u - u^* \|^2 = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{(2n+1)S_n^2}{K_n^2} - \underset{\sim}{f}^T \underset{\sim}{G}^{-1} \underset{\sim}{f}$$

As has been shown, for a particular data set this expression depends only on the kernel ie. the choice of K_n, $n = 0,1,2, \ldots$

An attempt will now be made to minimise this expression.

It will of course be required that $\sum_{n=0}^{\infty} K_n^2$ is finite, so the expression will be minimised subject to $\sum_{n=0}^{\infty} K_n^2 = K$, some constant.

It appears then, that a Lagrangian Multiplier technique will be appropriate.

ie. Minimise $\sum_{i=0}^{\infty} \frac{(2n+1)S_i^2}{K_i^2} - \underset{\sim}{f}^T \underset{\sim}{G}^{-1} \underset{\sim}{f}$

subject to $\sum_{i=0}^{\infty} K_i^2 = K$

Where λ is a Lagrangian Multiplier and where

$$F = \sum_{i=0}^{\infty} \frac{(2n+1)S_i^2}{K_i^2} - \underset{\sim}{f}^T \underset{\sim}{G}^{-1} \underset{\sim}{f} + \lambda \left(\sum_{i=0}^{\infty} K_i^2 - K \right) ,$$

the minimum will be obtained where $\frac{\partial F}{\partial K_i^2} = 0$, $i = 0,1,2, \ldots$

and $\frac{\partial F}{\partial \lambda} = 0.$

$$\text{Now } G = \begin{bmatrix} \sum_{i=0}^{\infty} K_i^2 P_i(\cos\psi_{11}) & \sum_{i=0}^{\infty} K_i^2 P_i(\cos\psi_{12}) & \cdots & \sum_{i=0}^{\infty} K_i^2 P_i(\cos\psi_{1n}) \\ & & & \\ \sum_{i=0}^{\infty} K_i^2 P_i(\cos\psi_{n1}) & & \cdots & \sum_{i=0}^{\infty} K_u^2 P_i(\cos\psi_{nn}) \end{bmatrix}$$

$$\left[\frac{\partial G}{\partial K_n^2}\right] = \begin{bmatrix} P_n(\cos\psi_{11}) & P_n(\cos\psi_{12}) & \cdots & P_n(\cos\psi_{1n}) \\ & & & \\ P_n(\cos\psi_{n1}) & & & P_n(\cos\psi_{nn}) \end{bmatrix}$$

$$= [P_n(\cos\theta)] \quad \text{(by definition)}$$

$$\text{Also } \frac{\partial G^{-1}}{\partial K_n^2} = -G^{-1}\left[\frac{\partial G}{\partial K_n^2}\right]G^{-1} = -G^{-1}[P_n(\cos\theta)]G^{-1}$$

$$\text{So } \frac{\partial F}{\partial K_n^2} = \frac{-(2n+1)S_n^2}{K_n^4} + \underset{\sim}{f}^T G^{-1}[P_n(\cos\theta)]G^{-1}\underset{\sim}{f} + \lambda = 0$$

$$\text{or } \frac{-(2n+1)S_n^2}{K_n^2} + \underset{\sim}{f}^T G^{-1}[K_n^2 P_n(\cos\theta)]G^{-1}\underset{\sim}{f} + \lambda K_n^2 = 0.$$

$$n = 0,1,2, \ldots$$

Summing over the range of n,

$$-\sum_{n=0}^{\infty} \frac{(2n+1)S_n^2}{K_n^2} + \underset{\sim}{f}^T G^{-1} G\, G^{-1}\underset{\sim}{f} + \lambda \sum_{n=0}^{\infty} K_n^2 = 0$$

$$\text{Now } \frac{\partial F}{\partial \lambda} = 0 \text{ implies } \sum_{n=0}^{\infty} K_n^2 = K$$

$$\text{So } \lambda = \frac{1}{K}\left\{\sum_{i=0}^{\infty} \frac{(2i+1)S_i^2}{K_i^2} - \underset{\sim}{f}^T G^{-1}\underset{\sim}{f}\right\}$$

$\psi_{1n})$

$\psi_{nn})$

$\psi_{1n})$

$\psi_{nn})$

...

Thus, letting $L = \sum_{i=0}^{\infty} \frac{(2i+1)S_i^2}{K_i^2}$,

$$L - \frac{K(2n+1)S_n^2}{K_n^4} + K\underset{\sim}{f}^T G^{-1}[P_n(\cos\theta)]G^{-1}\underset{\sim}{f} - \underset{\sim}{f}^T G^{-1}\underset{\sim}{f} = 0$$

$$n = 0,1,2, \ldots$$

This is an infinite set of equations in an infinite number of unknowns. However, for numerical purposes it may only be required to obtain, say, the first 201 coefficients K_n^2, $n = 0, 1, 2, \ldots, 200$. This would give a system of 201 equations in 201 unknowns, provided that one assigned values to L and K, or made them finite sums. The equations are however by no means linear and would be exceedingly difficult (if not impossible) to solve.

However, using the above equation and by dint of some perhaps doubtful manipulation it is possible to get a result which is interesting in terms of the convergence of the norm $\|u\|$ and the reproducing kernel $\sum_{n=0}^{\infty} K_n^2 P_n(\cos\psi)$.

Let $q_n = \frac{K(2n+1)S_n^2}{K_n^4} - L = K\underset{\sim}{f}^T G^{-1}[P_n(\cos\theta)]\, G^{-1}\underset{\sim}{f} - \underset{\sim}{f}^T G^{-1}\underset{\sim}{f}.$

In terms of the coefficient K_{n-1}^2, q_{n-1} may be formed:

$$q_{n-1} = \frac{K(2n-1)S_{n-1}^2}{K_{n-1}^4} - L = K\underset{\sim}{f}^T G^{-1}[P_{n-1}(\cos\theta)]\, G^{-1}\underset{\sim}{f} - \underset{\sim}{f}^T G^{-1}\underset{\sim}{f}$$

Thus $\underset{\sim}{f}^T[K\, q_{n-1}\, G^{-1}[P_n(\cos\theta)]G^{-1} - q_{n-1}\, G^{-1}]\underset{\sim}{f}$

$$= \underset{\sim}{f}^T[K\, q_n\, G^{-1}[P_{n-1}(\cos\theta)]G^{-1} - q_n G^{-1}]\underset{\sim}{f}$$

Noting that the converse is not claimed, it may be stated that this equation will be satisfied if

$$[Kq_{n-1}G^{-1}[P_n(\cos\theta)]G^{-1} - q_{n-1}G^{-1}] = [K\, q_nG^{-1}[P_{n-1}(\cos\theta)]G^{-1} - q_nG^{-1}]$$

ie. if $\quad (q_n - q_{n-1})G = K\{q_n[P_{n-1}(\cos\theta)] - q_{n-1}[P_n(\cos\theta)]\}$

From the definitions of G and $[P_n(\cos\theta)]$ it is seen that this condition is fulfilled if

$$(q_n - q_{n-1}) \sum_{n=0}^{\infty} K_n^2\, P_n(\cos\psi_{ij}) = K\{q_nP_{n-1}(\cos\psi_{ij}) - q_{n-1}P_n(\cos\psi_{ij})\}$$

for all ψ_{ij} in the matrix G.

Suppose $q_n = q_{n-1}$.

Then $\quad \dfrac{K(2n+1)S_n^2}{K_n^4} - L = \dfrac{K(2n-1)S_{n-1}^2}{K_{n-1}^4} - L$

$$\therefore \quad \frac{K_n^4}{K_{n-1}^4} = \frac{(2n+1)S_n^2}{(2n-1)S_{n-1}^2} \quad \text{or} \quad \frac{K_n^2}{K_{n-1}^2} = \sqrt{\frac{(2n+1)S_n^2}{(2n-1)S_{n-1}^2}}$$

A particular case where this will happen is where

$$K_n^2 = c\sqrt{(2n+1)S_n^2} \quad \text{and} \quad K_{n-1}^2 = c\sqrt{(2n-1)S_{n-1}^2}$$

where c is some constant.

But then $q_n = 0$ and $q_{n-1} = 0$ and the above condition holds for all ψ_{ij}, albeit trivially.

Now it is by no means claimed that setting $K_n^2 = c\sqrt{(2n+1)S_n^2}$ for n = 0, 1, 2, ... minimises $\|u-u^*\|$, for this would imply that

$$P_n(\cos\psi) = \frac{\sum_{n=0}^{\infty} K_n^2 P_n(\cos\psi)}{\sum_{n=0}^{\infty} K_n^2}$$

for all n, which only occurs when $\psi = 0$.

this

$-1 - q_n G^{-1}]$

t this

$_n(\cos\psi_{ij})\}$

ıll

ıat

Nevertheless, the choice of K_n^2 in such a manner does lead to interesting properties as far as the norm and the kernel and their convergence are concerned.

Now $$\|u\|^2 = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{(2n+1)S_n^2}{K_n^2}$$

with $$K_n^2 = c\sqrt{(2n+1)S_n^2},$$

$$\|u\|^2 = \frac{1}{4\pi c} \sum_{n=0}^{\infty} \sqrt{(2n+1)S_n^2}$$

And the kernel $$K(\psi) = \sum_{n=0}^{\infty} K_n^2 P_n(\cos\psi)$$

$$= c \sum_{n=0}^{\infty} \sqrt{(2n+1)S_n^2}\, P_n(\cos\psi)$$

$$\text{ie. } K(0) = c \sum_{n=0}^{\infty} \sqrt{(2n+1)S_n^2}$$

Here is a case where if the kernel is bounded then so is the norm and vice versa.

In section 2.9, this suggestion for the choice of norm is further discussed, in view of the recent tendency to use convenient expressions to approximate the empirical covariance function:

$$K(\psi) = \sum_{n=0}^{\infty} (2n+1)S_n^2 P_n(\cos\psi).$$

One thing which must be noted at this point is that although the empirical covariance kernel is bounded, the norm is not, as was pointed out by (TSCHERNING, 1977).

$$\text{ie} \quad \|u\|^2 = \sum_{n=0}^{\infty} \frac{(2n+1)S_n^2}{(2n+1)S_n^2} = \sum_{n=0}^{\infty} 1 = \infty.$$

So notwithstanding the lack of rigour in the derivation of this suggestion for the kernel, it may be worthwhile to use it with test data and compare the results with those obtained using the empirical covariance

function. It may also be interesting to consider $K_n^2 = \{(2n+1)S_n^2\}^{1/m}$, where m is some integer other than 2, in this regard.

1/m

CHAPTER 2

STATISTICAL METHODS

2.1 Introduction

To this stage, no probabilistic reasoning has been introduced into the solution of the problem, although mention has been made of covariance matrices in connection with the method known as least squares collocation.

Recapitulating, the equations were, for the "pure" case of least squares without derivatives

$$\hat{d}_p = \underset{\sim}{p}^T \underset{\sim}{c}$$

where $\underset{\sim}{c} = (H^T W H)^{-1} H^T W \underset{\sim}{f}$

and $\underset{\sim}{p}^T = [\, \langle \ell_p, B\ell_1 \rangle \quad \langle \ell_p, B\ell_2 \rangle \ \ldots \ \langle \ell_p, B\ell_n \rangle \,]$

Now since the functionals ℓ_i have been used for ψ_i, a least squares solution is also a collocation solution so $\underset{\sim}{c}$ may be expressed as

$$\underset{\sim}{c} = (L^T W H)^{-1} L^T W \underset{\sim}{f}$$

with $L = [\ell_{ij}]_{M\times n}$, $\ell_{ij} = \langle \ell_i, \ell_j \rangle \quad i = 1,2, \ldots M$

$H = [h_{ij}]_{M\times n}$, $h_{ij} = \langle \ell_i, B\ell_j \rangle \quad j = 1,2, \ldots n$

and $f_i = \langle \ell_i, g \rangle = g(t_i) \quad i = 1,2, \ldots, M.$

Now if $L^T W = [I_{n\times n} \vdots 0_{n\times M-n}]$

$L^T W H = G = [g_{ij}] \quad i = 1,2, \ldots, n, \quad j = 1,2, \ldots, n.$

$$L^T W \underset{\sim}{f} = \begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{bmatrix}$$

and the expression for $\underset{\sim}{c}$ becomes the same as that for the collocation, and thus the least squares and Galerkin cases. In effect, only the first n of the M observations of g are used.

The condition for this to happen is that

$$L^T W = [I_{n\times m} \vdots 0_{n\times m-n}]$$

Now let the matrix $S = [S_{ij}]_{M\times M}$ be given by

$$S_{ij} = \langle \ell_i, \ell_j \rangle \qquad i = 1,2, \ldots, M, \quad j = 1,2, \ldots, M.$$

The first n rows of S form the matrix L^T. So if $S = W^{-1}$ the expression for $L^T W$ holds.

In a purely deterministic solution one would, using knowledge of the properties of the differential equation to be solved, choose an appropriate reproducing kernel K and form the matrix S, letting

$$K(t_i,t_j) = S_{ij} .$$

W would then be the inverse of S, for the method to be equivalent to the ordinary least squares case.

However it is possible to examine the problem from a probabilistic viewpoint. If the observations $g(t_i)$ are regarded as random variables jointly distributed such that W is the inverse of the covariance matrix $C = [c_{ij}]$ where

$$c_{ij} = \mathrm{Cov}\{B^{-1}g(t_i), B^{-1}g(t_j)\} ,$$

the question is whether a reproducing kernel K may then be chosen such that $K(t_i,t_j) = c_{ij}$. If this is the case then $W^{-1} = S$ once more, and the method which is a statistical method known as "linear least squares prediction" will be equivalent to the solution of a differential equation using a particular and in some sense optimal reproducing kernel. That it is possible to do this is the subject of this chapter.

To further see the similarity between solution of differential equations by approximate methods and prediction it is instructive to examine, in a very elementary form, the concept of conditional expectation, especially in the case of jointly distributed normal random variables.

2.2 Conditional Expectation

Let the random variables $X_1, X_2, \ldots, X_n, Z$, be jointly distributed under some probability law with a density function $f(x_1, x_2, \ldots, x_n, z)$, which is assumed to be known.

Assume also that observations have been made of $X_1, X_2, \ldots, X_n$ but not Z, the observations being $x^1, x^2, \ldots x^n$. Then the conditional density function of Z given $X_1 = x^1$, $X_2 = x^2$, $X_n = x^n$ is

$$h(Z|\underset{\sim}{x}^{\cdot}) = h\,(Z|x^1, x^2, \ldots, x^n) = \frac{f(x^1, x^2, \ldots x^n, z)}{g(x^1, x^2, \ldots x^n)}$$

where $f(x^1, x^2, \ldots x^n, z)$ is the joint density function of $X_1, X_2, \ldots, X_n$, Z evaluated at $X_1 = x^1$, $X_2 = x^2$ etc, and $g(x^1, x^2, \ldots, x^n)$ is the joint marginal distribution of $X_1, X_2, \ldots, X_n$ evaluated at the same points.

The expectation of the variable Z with this density function is known as the conditional expectation of Z given $\underset{\sim}{x}^{\circ}$, or $E[Z|\underset{\sim}{x}^{\cdot}]$. It is also known as the regression of Z on $\underset{\sim}{x}^{\cdot}$.

Also if u(Z) is a function of Z the conditional expectation of u(Z) given $\underset{\sim}{x}^{\cdot}$ is, in the continuous case,

$$E[u(Z)|\underset{\sim}{x}^{\cdot}] = \int_{-\infty}^{\infty} u(z)\, h(z|\underset{\sim}{x}^{\cdot})dz.$$

The conditional variance is $E\{[Z - E[Z|\underset{\sim}{x}^{\cdot}]]^2|\underset{\sim}{x}^{\cdot}\}$ and is equal to

$$E[Z^2|\underset{\sim}{x}^{\cdot}] - \{E[Z|\underset{\sim}{x}^{\cdot}]\}^2 .$$

In the particular case when X_1, X_2, ... X_n, Z are jointly normally distributed, the conditional expectation which in a sense can be regarded as a predictor that is unbiassed and minimises the mean square error of prediction is given in a matrix form which one may recognise as being not unlike the equation for prediction of $g(t_p)$ in the solution of differential equations.

Using Theorem 3.10 (GRAYBILL, 1961) for the special case where Z is a single random variable, let $[X_1, X_2 \ldots X_n]^T = \underset{\sim}{X}$ and Z be all jointly normally distributed such that $E(\underset{\sim}{X}) = \underset{\sim}{\mu}_X$ and $E(Z) = \mu_Z$.

$$\text{Let } \underset{\sim}{Y} = \begin{bmatrix} Z \\ \hline \underset{\sim}{X} \end{bmatrix}, \quad \underset{\sim}{Y}^* = \begin{bmatrix} Z \\ \hline \underset{\sim}{x}^\cdot \end{bmatrix} \quad \underset{\sim}{\mu} = \begin{bmatrix} \mu_Z \\ \hline \underset{\sim}{\mu}_X \end{bmatrix}$$

Let the covariance matrix of $\underset{\sim}{Y}$ be V

where

$$V = \left[\begin{array}{c|c} \sigma_Z^2 & \underset{\sim}{P}^T \\ \hline \underset{\sim}{P} & G \end{array}\right]$$

Then the conditional distribution of Z given that $\underset{\sim}{X} = \underset{\sim}{x}^\cdot$ is the univariate normal distribution with mean $\mu_Z + \underset{\sim}{P}^T G^{-1}(\underset{\sim}{x}^\cdot - \underset{\sim}{\mu}_X)$ and variance given by

$$\sigma_Z^2 - \underset{\sim}{P}^T G^{-1} \underset{\sim}{P} .$$

In the case that $\underset{\sim}{\mu}_X = \underset{\sim}{0}$ and $\mu_Z = 0$ it is apparent that

$$E[Z|\underset{\sim}{x}^\cdot] = \underset{\sim}{P}^T G^{-1} \underset{\sim}{x}^\cdot$$

In the case of gravity measurements on the Earth for the prediction of some linear functional of the potential at point P, the elements of the matrix G, g_{ij} would be the covariances of the gravity anomaly measurements considered as normally distributed random variables at the points t_i, t_j $\quad i=1,2, \ldots, n, \quad j = 1,2, \ldots n$

and the elements of the vector $\underset{\sim}{P}$ would be the covariances between the particular linear functional of potential at the point P and the gravity anomaly at the points t_i, $i = 1,2, \ldots, n$ when both quantities are considered to be normally distributed random variables. The elements of the vector $\underset{\sim}{x}$ would be the actual measurements of gravity at the points t_i, $i = 1,2, \ldots n$ as they were in the case of the solution of differential equations.

So if some rationale could be arrived at to equate the $\underset{\sim}{P}$ vectors and the G matrices for both cases, then both methods would give identical results.

In the present context, however, this is not entirely possible due to the weakness of the assumption of normality. Even if it were possible to assume that the gravity anomaly measurements were normally distributed, it does not necessarily follow that the disturbing potential or a deflection of the vertical would also have this property. The nature of the measurements also seems to preclude the Gaussian distribution. For example a gravity anomaly less than about -990 gal would be extremely difficult to measure since this would imply that measured gravity at the point is negative and the observer and instrument would be accelerating spacewards. Very large positive anomalies would also be precluded so that the tail regions of the Normal curve would be greater than that of the actual distribution no matter how small the variance. This does tend to indicate, however, that whatever the true distribution, it will almost certainly have finite second moments if such moments exist. Consequently the assumption of Normality will be replaced by the assumption that all quantities ie. disturbing potential and its linear functionals are second order random variables with zero expectation.

The assumption of zero expectation is not completely necessary for it is always possible to construct random variables with zero expectation from the original random variables. A discussion of how this may be done will appear in section 2.10. For the time being, it will be assumed that the word "anomaly" means exactly what it appears to mean: something other than the expected value of zero.

In this case then the second moments will be variances and covariances and the problems considerably simplified.

The proper context for solution of this problem is undoubtedly a study of Hilbert Spaces spanned by second order random functions. The work of (PARZEN, 1959, 1960, 1961, 1962, 1963(a), (b)), (LAURITZEN, 1973), and (DERMANIS, 1977) will be closely followed in connection with this, reference being made to the particular problem of potential prediction where appropriate.

2.3 Second Order Random Functions

Consider a probability space (Ω, A, P). Let there be a family of random variables $\{X(t), t\varepsilon T\}$ where T is some index set such that the random variables are defined on the probability space for each point $t \varepsilon T$. The family of random variables is known as a random function. If in addition the nature of the probability space allows each of the random variables X(t) to have a finite second moment for any value of $t \varepsilon T$,

ie. $$E|X(t)|^2 = \int_{\Omega} |X(t)|^2 \, dP \quad \text{for all } t \varepsilon T,$$

then $\{X(t), t\varepsilon T\}$ is a second order random function.

In this definition the nature of the index set T has not been specified. In fact no restriction is placed on the nature of it.

If for example T is the integers or the natural numbers then $\{X(t), t\varepsilon T\}$ is known as a discrete parameter stochastic process, if T is the real line or a section of it a continuous parameter stochastic process results. When T is R^n the function is called an (n-dimensional) random field.

However what is of interest in this project is the case where T is either a sphere of radius R, the space outside this sphere, or the union of these two sets. The above definition does not preclude this. Whenever, for purposes of developing relationships, it is necessary to restrict the nature of T in future sections, the suitability of the modification to the special case in question will be discussed.

Now with the same probability space (Ω, A, P), $L_2(\Omega,A,P)$ will be defined to be the set of all random variables U, defined on this space, whose second moment is finite. The random variables comprising $\{X(t), t\varepsilon T\}$ will thus form a subset of this set.

2.4 Linear, Inner Product and Hilbert Spaces

Now if any two random variables U and V in $L_2(\Omega,A,P)$, are taken, it is easily seen that there is a unique random variable $W = U + V = V + U$ which must have finite variance. Also for random variables U, V and Y it is obvious that $U + (V + Y) = (U + V) + Y$ always holds. Now when a and b are real numbers and U and V are any random variables from $L_2(\Omega,A,P)$, aU is a unique random variable, $a(U + V) = aU + aV$, $(ab)U = a(bU)$, $(a+b)U = aU + bU$, $1.U = U$, when the scalar is unity, and for zero, $0.U = 0$, a random variable which is identically equal to zero. The zero random variable 0 is one for which

$$P\{\omega: 0(\omega) = 0\} = 1$$

and must be an element of $L_2(\Omega,A,P)$ since its variance is finite.

Obviously for any other U, $0 + U = U + 0 = U$. When uniqueness is mentioned in the above, it means that U and V may be regarded as the same vector if they differ only by a set of measure zero, for since (Ω, A, P) is a complete measure space, A must contain such sets. It is therefore seen that $L_2(\Omega, A, P)$ forms a linear space.

If a scalar product $\langle U,V \rangle$ between any two random variables is now defined as

$$\langle U,V \rangle = E[UV] = \int_{\Omega} UV dP$$

for all $U, V \in L_2(\Omega, A, P)$, then it can be seen that for $U,V,W, \in L_2(\Omega, A, P)$ and $a \in R$,

(i) $\langle aU,V \rangle = E[aUV] = aE(UV) = a\langle U,V \rangle$

(ii) $\langle U+V,W \rangle = E[(U+V)W] = E[UW+VW] = \langle U,W \rangle + \langle V,W \rangle$

(iii) $\langle U,V \rangle = E[UV] = E[VU] = \langle V,U \rangle$

and (iv) $\langle U,U \rangle = E[U^2] > 0$ if $U \neq 0$.

So $L_2(\Omega, A, P)$ is an inner product space.

Defining the norm of U as $\| U\| = \{E[u^2]\}^{\frac{1}{2}} \geq 0$, it is seen that

$$\begin{aligned} \| U+V\|^2 &= E(U+V)^2 = E[U^2] + E[V^2] + 2E[UV] \\ &\leq E[U^2] + E[V^2] + 2\left|\sqrt{E[U^2]}\right| \cdot \left|\sqrt{E[V^2]}\right| \\ &= \| U\|^2 + \| V\|^2 + 2\| U\| \cdot \| V\| \\ &= \{\| U\| + \| V\|\}^2 \end{aligned}$$

ie $\| U+V\| \leq \| U\| + \| V\|$.

So $L_2(\Omega, A, P)$ is a normed space under the norm induced by the inner product.

Now a series U_n of random variables in $L_2(\Omega, A, P)$ is said to converge (strongly) to U if

$$\lim_{n \to \infty} \| U_n - U\| = 0.$$

If U_n is convergent then the Cauchy sequence converges

ie, $\lim_{m,n\to\infty} \| U_n - U_m \| = 0$

Now $\| U_n - U_m \| \geq \left| \| U_n \| - \| U_m \| \right|$

and $\| U_n - U \| \geq \left| \| U_n \| - \| U \| \right|$

by Schwarz inequality.

So there exists some $\varepsilon_1 > 0$ such that for all $m,n \geq N$

$$\| U_n - U_m \| \geq \left| \| U_n \| - \| U_m \| \right| < \varepsilon_1$$

So let $n = N$ $m = N+k$, k an integer > 0.

For any value of k then $\| U_n \|$ cannot differ from $\| U_m \|$ by more than ε_1.

Also for $m = N+k$, $n = N$ there is some ε_2 such that

$$\| U_m - U \| \geq \left| \| U_m \| \quad \| U \| \right| < \varepsilon_2 .$$

Now $\| U_n \| < \infty$ by definition for a finite $n = N$.

So if $\| U \| = \infty$, for $\left| \| U_m \| - \| U \| \right| < \varepsilon_2$ to hold there must be some k such that $\| U_{n+k} \|$ becomes close to $\| U \|$. But then $\left| \| U_n \| - \| U_m \| \right| > \varepsilon_1$ contrary to the assumption that the Cauchy sequence converges. Thus $\| U \|$ cannot be infinite and thus it must be an element of $L_2(\Omega,A,P)$. So $L_2(\Omega,A,P)$ is complete.

Thus $L_2(\Omega,A,P)$ is a complete normed space with the norm generated by the inner product

$$\langle U,V \rangle = E(UV).$$

This is what is known as a Hilbert Space.

The completeness property of $L_2(\Omega,A,P)$ is of great consequence because if any linear subspace of it can be shown to have the same norm and thus the same inner product then merely by adding to this subspace the limits of its Cauchy sequences, which must exist because of this property, the subspace can be made into a Hilbert Space. Such a subspace is the topic of the next section.

2.5 Hilbert Space Spanned by a Random Function

Consider a random function of second order $\{X(t),\ t\varepsilon T\}$ consisting of random variables defined on the probability space (Ω,A,P). Now for each $t\varepsilon T$ $X(t)$ must lie in $L_2(\Omega,A,P)$. So also will linear combinations of the $X(t)$'s. It would appear then that in much the same way as a set of linearly independent vectors in R^n form a basis for a subspace of R^n the set of random variables $\{X(t),\ t\varepsilon T\}$ may serve a similar purpose for a subspace of $L_2(\Omega,A,P)$.

Define, therefore, a Linear Manifold, spanned by the random function $\{X(t),\ t\varepsilon T\}$, denoted $L\{X(t),\ t\varepsilon T\}$, to be the set of all random variables X which may be written in the form

$$X = \sum_{i=1}^{n} c_i\, X(t_i), \quad \text{for } n \text{ some integer,}$$

c_i $i = 1,2,\ldots n$, real constants

and $t_i\varepsilon T$, $i = 1,2,\ldots n$.

It thus consists of every possible finite linear combination of the random variables $X(t)$, $t\varepsilon T$. It does not, of course, include infinite linear combinations and a simple example shows why: Suppose T to be the real line and there is some random function which for all $i\varepsilon N$, $E(X(i))^2 > 1$ $E(X(i)\ X(j)) > 0$ and $E(X(t))^2 < \infty$, for all $t\varepsilon T$. The Linear Combination $Y = \sum_{i=1}^{\infty} X(i)$ will have an infinite second moment and thus not be an element of $L_2(\Omega,A,P)$.

However, finite linear combinations must have finite second moments so the linear manifold is restricted to only these. The properties of closure, associativity, commutativity, scalar multiplication and distributivity obviously hold, the zero element being any linear combination where $c_i = 0$ for all i. So the linear manifold $L\{X(t), t\varepsilon T\}$ is a linear space. Allowing the inner product to be the same as that of $L_2(\Omega,A,P)$ enables it to become an inner product subspace thereof, the norm being defined as before.

As mentioned before by simply allowing such limits of sequences of random variables X_n as are necessary to be adjoined to $L\{X(t), t\varepsilon T\}$ the completeness property for a larger space called $L_2\{X(t),t\varepsilon T\}$ is satisfied. It is not necessary to adjoin all limits of sequences in $L_2(\Omega,A,P)$ for this to happen, only the limits of sequences of random variables in $L\{X(t), t\varepsilon T\}$ and no others are required. So in fact $L_2\{X(t),t\varepsilon T\}$ is the <u>smallest</u> subset of $L_2(\Omega,A,P)$ spanned by the manifold which possesses the properties of a Hilbert Space, and is known as the Hilbert Space spanned by the random function $\{X(t), t\varepsilon T\}$. Note also that $\{X(t), t\varepsilon T\}$ does not necessarily form a <u>basis</u> for $L_2\{X(t), t\varepsilon T\}$. It is only a spanning set.

Now there are some points worth noting about the nature of the set T. As yet there have been no restrictions placed on it. So, for example, T could be the real line. It could also be a finite number m of points on the real line : T'. Now $T' \subset T$. Certainly any linear combination of random variables $X' = \sum_{i=1}^{m} c_i X(t_i)$, $t_i \varepsilon T'$ is also a linear combination for the same points $t_i \varepsilon T$. So $L\{X(t), t\varepsilon T'\}$ is a proper linear subspace of $L\{X(t), t\varepsilon T\}$. Since both manifolds will be given the same inner product and since limits of sequences of X_n' must be not only in $L_2\{X(t), t\varepsilon T'\}$ but also in $L_2\{X(t), t\varepsilon T\}$ then the first

is a Hilbert subspace of the second and is in fact a finite dimensional subspace, the dimension being the number of linearly independent elements in the spanning set $\{X(t),\ t\varepsilon T'\}$. There may be elements of $L_2\{X(t),\ t\varepsilon T\}$ which are not elements in $L_2\{X(t),\ t\varepsilon T'\}$ and a very important problem is to find that element in $L_2\{X(t),\ t\varepsilon T'\}$ which most closely approximates a given element in $L_2\{X(t),\ t\varepsilon T\}$. This will be the topic of much further discussion in section 2.6.

It may also be noted that the cases of most interest to this project are not precluded. So if gravity is measured at a number of points on the earth, the points where it is measured could be regarded as T' a subset of T the sphere of radius R, and $L_2\{X(t),\ t\varepsilon T'\} \subseteq L_2\{X(t),\ t\varepsilon T\}$.

But it is not only the prediction of gravity anomalies that is of interest. Linear functionals such as the disturbing potential, the geoid spheroid separation and the deflections of the vertical are also required.

That is to say if $\{X(t),\ t\varepsilon T\}$ are potential functions and thus $X(t_i)$ is the random variable at some point not only is $\ell_\omega^*\{X(t_i)\}$ required ie. the evaluation of $X(t_i)$ at some point $\omega \subseteq \Omega$ or "outcome", but the evaluation of certain linear operations on $\{X(t),\ t\varepsilon T\}$ is also required. If $L_2^*\{X(t),\ t\varepsilon T\}$ is the dual space of $L_2\{X(t),\ t\varepsilon T\}$ it consists of all possible bounded linear functionals of the space. There is a representer in $L_2\{X(t),\ t\varepsilon T\}$ for each functional.

So if $B : L_2\{X(t),\ t\varepsilon T\} \to L_2\{X(t),\ t\varepsilon T\}$ is a bounded linear operator and

$Y(t) = BX(t)$ is some other process then the outcome of $Y(t_i)$ is given by

$\langle \ell_\omega,\ Y(t_i) \rangle$, where ℓ_ω is the representer of the evaluation functional ℓ_ω^* .

But $\langle \ell_\omega, Y(t_i) \rangle = \langle \ell_\omega, BX(t_i) \rangle \quad \langle B^* \ell_\omega, X(t_i) \rangle$

where B^* is the adjoint of B. Now $B^* \ell_\omega \in L_2\{X(t), t\in T\}$

so letting $B^* \ell_\omega = \overline{\ell}_\omega$,

$$\ell_\omega^*\{Y(t_i)\} = \langle \overline{\ell}_\omega, X(t_i) \rangle = \overline{\ell}_\omega^*\{X(t_i)\}$$

which is simply a linear functional of $X(t_i)$, $t_i \in T$.

So it is only necessary to deal with the process $\{X(t), t\in T\}$ and not with a multitude of them. The Hilbert Space $L_2\{X(t), t\in T\}$ consists of all random variables which may be obtained by means of linear operations on the random variables $\{X(t), t\in T\}$. This becomes somewhat clearer when $L_2\{X(t), t\in T\}$ is represented by a Reproducing Kernel Hilbert Space. However at this stage the concept of Linear Least Squares prediction must be examined for one is now able to obtain an equation of the form

$$\hat{d}_p = \underset{\sim}{P}^T \underset{\sim}{G}^{-1} \underset{\sim}{x} .$$

2.6 The Projection Theorem and Linear Least Squares Prediction

The projection theorem is well known. The form in (PARZEN, 1959), theorem 6A will be used here with slightly altered notation. A proof is included in Parzen's article which will not be reproduced.

2.6.1 The Projection Theorem

Let H be some abstract Hilbert Space which has M as a subspace. Let $\underset{\sim}{z}$ be some vector in H. Let, for any vector $\underset{\sim}{z} \in H$:

$$d\{\underset{\sim}{z}|M\} = \inf_{\underset{\sim}{x}\in M} \| \underset{\sim}{x} - \underset{\sim}{z} \| .$$

Then there exists a unique vector $\hat{\underset{\sim}{z}} \in M$ which satisfies each of the following equivalent conditions

$$(1)\quad \| \hat{z} - z \| = d\{z|M\} = \inf_{x \in M} \| x - z \|$$

$$(2)\quad \langle \hat{z} - z, x \rangle = 0 \quad \text{for all } x \in M$$

$$(3)\quad \langle \hat{z}, x \rangle = \langle z, x \rangle \quad \text{for all } x \in M .$$

$\hat{z}$ may be regarded as a least squares estimator of z.
An exposition of the relationship between geometry and least squares estimation may be found in (DERMANIS, 1977).

Now suppose that the vectors $x_1, x_2, \ldots, x_n$ form a basis for M which is of dimension n. Any vector $h \in M$ may be expressed as

$$h = \sum_{i=1}^{n} c_i x_i \qquad \text{where } c_i \in R \qquad i = 1,2,\ldots, n.$$

Among all such vectors a vector $\hat{z}$ may be found which most closely approximates $z \in H$ in the least squares sense. So it is required to find constants $\hat{c}_i$, $i = 1,2,\ldots n$ such that

$$\hat{z} = \sum_{i=1}^{n} \hat{c}_i x_i .$$

From condition (3) above, then for $i = 1,2, \ldots, n$,

$$\langle \hat{z}, x_i \rangle = \langle z, x_i \rangle$$

$$\therefore \quad \langle \sum_{j=1}^{n} \hat{c}_j x_j, x_i \rangle = \langle z, x_i \rangle \qquad i = 1,2, \ldots, n.$$

$$\text{or } \hat{c}_1 \langle x_1, x_i \rangle + \hat{c}_2 \langle x_2, x_i \rangle + \ldots + \hat{c}_n \langle x_n, x_i \rangle = \langle z, x_i \rangle ,$$

$$i = 1,2, \ldots, n.$$

This is a system of n equations in n unknowns so $\hat{c}_i$ $i = 1,2, \ldots, n$ may be found by solving them simultaneously. The vector $\hat{z}$ is found immediately, for

$$\hat{z} = \sum_{i=1}^{n} \hat{c}_i x_i$$

Now let ℓ_p^* be an evaluation functional in the dual space M^* with representer $\ell_p \in M$.

$$\text{Let } \hat{d}_p = \ell_p^*\{\hat{z}\} = \langle \underset{\sim}{\ell}_p, \hat{\underset{\sim}{z}} \rangle$$

$$\text{Then } \hat{d}_p = \langle \underset{\sim}{\ell}_p, \sum_{i=1}^{n} \hat{c}_i \underset{\sim}{x}_i \rangle$$

$$= \sum_{i=1}^{n} \hat{c}_i \langle \underset{\sim}{\ell}_p, \underset{\sim}{x}_i \rangle$$

$$= \sum_{i=1}^{n} \hat{c}_i x_i^{\cdot} \qquad x_i^{\cdot} \in R \qquad (x_i^{\cdot} = \underset{\sim}{x}_i(p)).$$

Now the vectors $\underset{\sim}{x}$, $\underset{\sim}{z}$ etc. are vectors in abstract Hilbert Spaces M and H. Below, the vector notation will refer to vectors in E^n as defined.

$$\text{Let } \underset{\sim}{c}^T = [\hat{c}_1, \hat{c}_2, \ldots c_n]$$

$$\underset{\sim}{p}^T = [\langle \underset{\sim}{z}, \underset{\sim}{x}_1 \rangle, \langle \underset{\sim}{z}, \underset{\sim}{x}_2 \rangle, \ldots \langle \underset{\sim}{z}, \underset{\sim}{x}_n \rangle]$$

$$G = \begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_2, x_1 \rangle & \cdots & \langle x_n, x_1 \rangle \\ \langle x_1, x_2 \rangle & \langle x_2, x_2 \rangle & \cdots & \langle x_n, x_2 \rangle \\ \cdots & \cdots & & \cdots \\ \cdots & \cdots & & \cdots \\ \langle x_1, x_n \rangle & \langle x_2, x_n \rangle & \cdots & \langle x_n, x_n \rangle \end{bmatrix} = G^T$$

$$\underset{\sim}{x}^{\cdot T} = [x_1^{\cdot} \; x_2^{\cdot} \; \cdots \; x_n^{\cdot}]$$

Clearly, $G\underset{\sim}{c} = \underset{\sim}{p}$ or $\underset{\sim}{c} = G^{-1}\underset{\sim}{p}$

and thus $\hat{d}_p = \sum_{i=1}^{n} \hat{c}_i x_i^{\cdot} = \underset{\sim}{c}^T \underset{\sim}{x}^{\cdot} = \underset{\sim}{p}^T G^{-1} \underset{\sim}{x}^{\cdot}$.

However, this has only been shown for abstract spaces M and H.

To see even more similarity between this expression and conditional expectation for jointly normally distributed variables, the application to Hilbert Spaces of random variables will now be examined.

2.6.2 Application of the Projection Theorem to Random Variables

Let the space H be the Hilbert Space spanned by the random function $\{X(t),\ t\varepsilon T\}$ ie. $L_2\{X(t),\ t\varepsilon T\}$. Now let T' be the set $\{t_1, t_2, \ldots, t_n \mid t_i\varepsilon T\}$. The random variables $X(t_1), X(t_2), \ldots, X(t_n)$ will be assumed to form a basis for $L_2\{X(t) \mid t\varepsilon T'\}$. Let this space be M. Now let the random variable Z be an element of H. The inner product of both M and H is of course given by $\langle X,Y \rangle = E(XY)$ for all $X,Y\varepsilon M$ or H.

Note that $E(XY) = Cov(X,Y)$ since $E(X) = E(Y) = 0$, for all X,Y. The object is to choose the element of M which most closely approximates Z in the least squares sense. It is easy to see that $\hat{Z}$ is a least squares estimator in the usual sense of the word for one wishes to minimise

$$\| \hat{Z}-Z\| = \langle \hat{Z}-Z,\ \hat{Z}-Z \rangle^{\frac{1}{2}}$$

$$= \{Var(\hat{Z}-Z)\}^{\frac{1}{2}} = \{Var\{\hat{Z}\}\}^{\frac{1}{2}}$$

In other words $\hat{Z}$ possesses the properties of being a minimum variance linear unbiassed estimator, $E(\hat{Z})$ being equal to zero, and these are the basic properties of a least squares estimator.

The third condition of the projection theorem is once again invoked, ie. for all $X(t)\ t\varepsilon T'$

$$\langle \hat{Z},\ X(t) \rangle = \langle Z,\ X(t) \rangle$$

$$\text{or} \quad \langle \hat{Z},\ X(t_i) \rangle = \langle Z,\ X(t_i) \rangle \quad \text{for all} \quad t_i\varepsilon T' \quad \text{ie} \quad i = 1,2, \ldots$$

A random variable in M will have the form

$$Y = \sum_{i=1}^{n} c_i X(t_i) .$$

In particular the one which most closely approximates Z will be

$$\hat{Z} = \sum_{i=1}^{n} \hat{c}_i X(t_i)$$

So for $i = 1,2, \ldots, n$

$$\langle \sum_{j=1}^{n} \hat{c}_j X(t_j),\ X(t_i) \rangle = \langle Z, X(t_i) \rangle$$

$$\therefore \sum_{j=1}^{n} \hat{c}_j \langle X(t_j), X(t_i) \rangle = \langle Z, X(t_i) \rangle \qquad i = 1,2, \ldots, n$$

$$\text{or} \quad \sum_{j=1}^{n} \hat{c}_j \operatorname{Cov}\{X(t_j), X(t_i)\} = \operatorname{Cov}\{Z, X(t_i)\} \qquad i = 1,2, \ldots, n$$

In the prediction case $Z = X(t_p)$.

Letting the vector of covariances $\operatorname{Cov}\{Z, X(t_i)\}$ $i = 1,2, \ldots,n$ be $\underset{\sim}{p}$, the covariance matrix of the $X(t_i)$'s and $X(t_j)$'s be G and the vector of estimated constants $\hat{c}_i$ be $\underset{\sim}{c}$, again the following equation is obtained:

$$G\underset{\sim}{c} = \underset{\sim}{p} \qquad \text{or} \qquad \underset{\sim}{c} = G^{-1}\underset{\sim}{p} .$$

If $X(t_1)$, $X(t_2)$ ... $X(t_n)$ had only been a spanning set rather than a basis for $L_2\{X(t), t\varepsilon T'\}$ it would not be possible to use G^{-1} and generalised inverses would have to be used. However, it is not proposed to deal with this aspect in this project. The relevant details are in (DERMANIS, 1977), section 3.5.

$$\text{Now} \quad \hat{Z} = \sum_{i=1}^{n} c_i X(t_i) = X(t_p)$$

This gives the random variable $\hat{Z}$ in terms of the random variables $X(t_i)$. However, for prediction, what is of importance is the value of $\hat{Z}$ for a particular realisation of the process. ie for the sampled values of $X(t_i)$ $i = 1,2, \ldots, n$ which may be written $X(t_i,\omega)$, $\omega\varepsilon A$. In effect $Z(\omega)$ for some particular ω is required.

This is done by defining the evaluation functional ℓ_ω^* such that

$$\ell_\omega^*\{\hat{Z}\} = \langle \ell_\omega, \hat{Z} \rangle = \hat{Z}(\omega)$$

and $$\ell_\omega^*\{X(t_i)\} = \langle \ell_\omega, X(t_i) \rangle = X(t_i, \omega)$$
$$= x_i^\cdot \qquad i = 1,2, \ldots, n.$$

It is not intended to construct one of these functionals for it is only its effect which is of importance. It is known that such things exist and that ℓ_ω is in fact a random variable since M and H are Hilbert spaces and the Riesz Representation Theorem holds. This is one of the benefits of abstract reasoning, for such random variables may be somewhat difficult to construct.

So armed with such functionals it is possible to obtain a prediction for $\hat{Z}$ given that certain values have been observed at the points t_i.

Since

$$\hat{Z} = \sum_{i=1}^{n} \hat{c}_i X(t_i)$$

$$\hat{d}_p = \hat{Z}(\omega) = \langle \ell_\omega, \hat{Z} \rangle = \langle \ell_\omega, \sum_{i=1}^{n} \hat{c}_i X(t_i) \rangle$$

$$= \sum_{i=1}^{n} \hat{c}_i \langle \ell_\omega, X(t_i) \rangle$$

or $$\hat{X}^\omega(t_p) = \sum_{i=1}^{n} \hat{c}_i x_i^\cdot$$ where $x_i^\cdot$, $i = 1,2, \ldots, n$ are the values observed.

In matrix notation then

$$\hat{d}_p = \underset{\sim}{c}^T \underset{\sim}{x}^\cdot \qquad \text{where } \underset{\sim}{x}^{\cdot T} = [x_1^\cdot, x_2^\cdot, \ldots x_n^\cdot]$$

$$\hat{d}_p = \underset{\sim}{p}^T G^{-1} \underset{\sim}{x}^\cdot$$ an identical expression to that previously obtained. However in addition to being the same expression in terms of symbols, the symbols mean exactly the same things as they did in the expression for the conditional expectation of Z given $\underset{\sim}{x}^\cdot$ in the case of jointly normally distributed random variables.

ie. in the normal case

$$E[Z|\underset{\sim}{x}^{\bullet}] = \hat{d}_p$$

(PARZEN, 1959) gives as lemma 6a a necessary and sufficient condition for the least squares estimator or projection to be identical to a true conditional expectation and because of the similarity uses the notation $E^*[Z|M]$ to denote $\hat{d}p$. Here, however projections are of interest and if they happen to coincide with conditional expectations this will merely be regarded as fortunate. It has only been mentioned to emphasise that the coincidence of the two concepts does not belong exclusively to the normal distribution.

2.6.3 Further Properties of Projections

The smoothing property is given by Theorem 6B (PARZEN, 1959) and put simply it states that if M and M_1 are subspaces of the Hilbert Space H such that $M_1 \subseteq M$ then the prediction obtained by projecting first on M then onto M_1 is the same as if one projected first onto M_1 and then onto M. In short it is simply the projection onto M_1.

Now if there were a second subspace of M called M_2 and M_1 and M_2 were orthogonal complements then the addition of data could be dealt with by determining the projection on M_2 and adding it to the projection on M_1. Unfortunately, however in gravity anomaly prediction it is never the case that $M_1 \perp M_2$. However the procedure is not much more complicated if the spaces are not orthogonal. Simple matrix manipulation gives an expression for the quantity to be added to the projection on M_1. The procedure in Geodesy is called "Stepwise Collocati and is used as a computational aid to prevent computer storage problems where large matrices are involved, as well as for updating or improving prediction as more gravity and other data becomes available.

As there is extensive literature on this aspect, for example (MORITZ, 1980) section 19, it will not be dealt with here.

A more interesting property is given by Theorem 6D (PARZEN, 1959) which, in part, states:

Let M_n, $n = 1, 2, \ldots$, be a sequence of Hilbert Subspaces of H which are monotone non-decreasing (ie. M_n is a subspace of M_{n+1}). Define M_∞ to be the Hilbert Space spanned by the union $\bigcup_{n=1}^{\infty} M_n$. Let $Z_1, Z_2, \ldots$, be a sequence of vectors in H such that for every integer m and n

$$E^*[Z_n|M_m] = Z_m \quad \text{if } m \leq n.$$

Then there is a unique vector X in M_∞ such that

$$Z_n = E^*[X|M_n] \quad \text{for every } n,$$

$$\lim_{n\to\infty} \|Z_n - X\| = 0$$

if and only if $\lim_{n\to\infty} \|Z_n\|^2 < \infty$

If Z is a vector in H such that $Z_n = E^*[Z|M_n]$, $n = 1,2, \ldots,$ then

$$X = E^*[Z|M_\infty]$$

In the case of linear least squares prediction the Hilbert Space H would be $L_2\{X(t), t\varepsilon T\}$ for some set T. For example, T may be the real line, T_n could be a finite subset of T, perhaps $\{N-n, N-n+1, \ldots, N-1, N\}$ and T_n could be the integers less than or equal to some number N. Thus with H as above M_n would be $L_2\{X(t), t\varepsilon T_n\}$ and $M_\infty = L_2\{X(t), t\varepsilon T_\infty\}$. Since the elements of the Hilbert Spaces are random variables with finite second moments and because of the completeness property the theorem holds. In other words if a sequence of predictions utilised an increasing amount of available information then the prediction

based on a finite past period would tend to the prediction over the infinite past.

For a random field with index set T the surface of a sphere of radius R a similar effect occurs. Here one would have observations of the gravity anomaly at n points on the surface of the earth. The positions at which these measurements were made would constitute a finite subset of the elements of T and this would be T_n. The set T_∞ would be the set of all points with rational values for the coordinates (θ, λ) where θ is the co-latitude and λ is the longitude ie T_∞ is a countable set dense in T. Again assuming that the random variables have finite second moments, the theorem holds and the predictions based on an increasing amount of gravity data will tend to the predictions based on gravity data made at all points on T_∞. Now since in practice there is absolutely no difference between having gravity data at all points in T_∞ and at all points in T, as far as gravity anomaly prediction is concerned this is a useful theorem indeed for it confirms the belief that the more gravity data there is available the better will be the prediction.

However, when T is the space in R^3 outside a sphere of radius R it is quite possible to have a countably infinite set which is not dense in the set T. For example, rational values or even integral values on a ray in a particular direction from the origin outside a sphere of radius R. This aspect will be further dealt with when reproducing kernels and their relationship to linear least squares prediction is examined.

Before going on to this topic, by way of summary it is worthwhil noting two important points which are perhaps not stressed in the Geodetic literature and could cause some conjecture.

The first is that in all the above, absolutely no restriction has been placed on the nature of T. T may be the real line, a sphere of radius R, the space in R^3 outside such a sphere, the integers or a subset of the integers. This means that the method is not merely applicable to discrete or continuous time stochastic processes but may be used on random fields based on sets which need not even be of a Euclidean type.

The second point is that if we define the function $C(t_1, t_2) = \text{cov}[X(t_1), X(t_2)]$ $t_1, t_2 \in T$ no restriction such as stationarity, homogeneity or isotropy has been placed on such a function, which may be referred to as the covariance function. It is only required that the elements of the matrix G be known ie. $g_{ij} = \text{cov}[X(t_i),X(t_j)]=C(t_i,t_j)$ and prediction is possible by the linear least squares projection method described. (Note that the projection theorem nowhere requires $E[X(t)]=0$).

It is noted that the form of the equation for the prediction by projection methods and that using approximate solutions of differential equations is much the same. In the solution of differential equations the equation was

$$\hat{d}_p = \underset{\sim}{p}^T G^{-1} \underset{\sim}{f}$$

$\underset{\sim}{f}$ was a vector of known or observed values, $f_i = \langle K^{t_i}, u \rangle$

G was the matrix $[g_{ij}]$ whose i, j th element was given by

$$g_{ij} = K(t_i, t_j) \qquad t_i, t_j \in T$$

where $K(t_i, t_j)$ was a reproducing kernel, and the i th element of $\underset{\sim}{p}$ was given by

$$p_i = K(t_p, t_i) \qquad t_p, t_i \in T$$

where t_p was the point where the differential equation was to be evaluated.

The function K(s,t) was the reproducing kernel of the reproducing kernel Hilbert Space H(K). This is the equations for "pure" collocation (etc) or prediction of like quantities. Slight modifications are made when "unlike" quantities (eg. potential from gravity anomaly etc.) are required.

In linear least squares prediction, however, the elements p_i and g_{ij} are given as covariances ie.

$$g_{ij} = \mathrm{Cov}[X(t_i), X(t_j)] = E[X(t_i), X(t_j)] = C(t_i, t_j)$$
$$p_i = \mathrm{Cov}[X(t_p), X(t_i)] = E[X(t_p), X(t_i)] = C(t_p, t_i)$$

with the elements of the vector $\underset{\sim}{x}^{\bullet}$ being the observed quantities once again, the equation of prediction being

$$\hat{d}_p = \underset{\sim}{p}^T G^{-1} \underset{\sim}{x}^{\bullet} .$$

Now $C(t_i, t_j) = E[X(t_i), X(t_j)] = E[X(t_j), X(t_i)]$

$= C(t_j, t_i) \qquad t_i, t_i \in T.$

C is defined on the set T X T.

It is also non-negative due to Theorem 2A (PARZEN, 1959) which states:

C is the covariance kernel of a random function, if and only if, C is a symmetric non-negative kernel.

It thus possesses many of the properties of the reproducing kernel K(s,t). In fact, recalling Theorem 1.6.3, if T is some arbitrary set and the function C : T X T → R is positive and symmetric, there exists a Hilbert Space, unique up to isomorphism with C as its reproducing kernel. (This is called the Moore-Aronszajn -Loève theorem

One then may ask if it is possible to use the covariance kernel to generate a Hilbert Space H(C) and simply regard the problem as the solution of a differential equation by approximate (MWR) methods in the

special case where C happens to be the reproducing kernel. In other words, are the deterministic and statistical methods in a sense equivalent? This will be the topic of discussion in section 2.7. (DERMANIS, 1977) has dealt with this problem using the orthogonal expansion known as the Karhunen-Loève expansion. This requires the use of Mercer's Theorem. According to (PARZEN, 1959) and (ZAANEN, 1953), Mercer's Theorem holds if T is assumed to be a closed bounded subset of a Euclidean space. It is intended, however, in this project to place as few restrictions as possible on the nature of T. According to (PARZEN, 1960) the notion of the representation of a time series by a reproducing kernel Hilbert Space is the natural setting in which to solve problems of statistical influence. His work is closely followed in the next section, and this is where the proofs of theorems and lemmas may be found.

2.7 Relationship Between Approximation and Statistical Methods

2.7.1 Definitions

An isomorphism between two Hilbert Spaces H_1 and H_2 is a mapping $\psi : H_1 \to H_2$ which is a one to one correspondence (ie. such that an inverse $\psi^{-1} : H_2 \to H_1$ exists), and satisfies the following properties.

For any vectors u_1 and $u_2 \in H_1$ and for any real number a:

1) $\psi(u_1 + u_2) = \psi(u_1) + \psi(u_2)$

$\psi(au_1) = a\psi(u_1)$

A mapping ψ is an isometry if it preserves inner products.

ie. Denote the inner product between two vectors $u_1, u_2 \in H_1$ as $\langle u_1, u_2 \rangle_1$ and the inner product between $v_1, v_2 \in H_2$ as $\langle v_1, v_2 \rangle_2$

Then ψ is an isometry if

$$\langle u_1, u_2 \rangle_1 = \langle \psi(u_1), \psi(u_2) \rangle_2$$

A congruence is a mapping which is both an isomorphism and an isometry.

So two Hilbert Spaces H_1 and H_2 are said to be congruent if there exists such a mapping between them. A congruence not only maps linear combinations of vectors in H_1 to corresponding linear combinations in H_2 but maps limits into limits ie.

$$u = \lim_{n\to\infty} u_n \text{ if and only if } \psi(u) = \lim_{n\to\infty} \psi(u_n).$$

In much the same way as was done for random functions a Hilbert Space H with index set T may be considered and the linear manifold $L\{u(t), t\varepsilon T\}$ spanned by the family of vectors $\{u(t), t\varepsilon T\}$ is defined as the set of all finite linear combinations of vectors in this family. By adjoining limits of cauchy sequences as before the Hilbert Space $L^*\{u(t), t\varepsilon T\}$ spanned by the family may be formed.

If $L^*\{u(t), t\varepsilon T\}$ coincides with H it may be said that $\{u(t), t\varepsilon T\}$ spans H.

Furthermore if $\{u(t), t\varepsilon T\}$ spans H but no subset of it spans H then it is a basis for H, and the number of vectors in $\{u(t), t\varepsilon T\}$ is the dimension of H. Such a basis is not unique, but all bases for H must have the same number of vectors.

2.7.2 The Basic Congruence Theorem

Let H_1 and H_2 be two abstract Hilbert Spaces. Let T be an index set. Let $\{u(t), t\varepsilon T\}$ be a family of vectors which span H_1 and $\{v(t), t\varepsilon T\}$ a family of vectors which span H_2. Suppose that, for every s and t in T,

$$\langle u(s), u(t) \rangle_1 = \langle v(s), v(t) \rangle_2.$$

Then the spaces H_1 and H_2 are congruent, and one can define a congruence ψ from H_1 to H_2 which has the property that

$$\psi(u(t)) = v(t) \quad \text{for all } t \in T.$$

For a proof see (PARZEN, 1959).

Using the Gram-Schmidt orthonormalisation process it may be shown that any Hilbert Space possesses an orthonormal basis ie. if $\{w(t), t \in T\}$ is a basis for H it is an orthonormal basis if for all $t_i, t_j, \in T$,

$$\langle w(t_i), w(t_i) \rangle = 1$$

$$\langle w(t_i), w(t_j) \rangle = 0 \quad \text{if } t_i \neq t_j$$

Now if H_1 and H_2 are Hilbert Spaces of the same dimension their bases may be put in a one-to-one correspondence. Thus there exists an index set T and families of vectors $\{u(t), t \in T\}$ and $\{v(t), t \in T\}$ which are orthonormal bases for H_1 and H_2 respectively.

ie. $\psi(u(t_i)) = v(t_i)$

$\psi(u(tj)) = v(t_j)$

for all $t_i, t_j \in T$ where ψ is the one-to-one correspondence.

Now since the families are orthonormal sets

$$\langle u(t_i), u(t_j) \rangle_1 = 0 \qquad i \neq j$$

$$\langle u(t_i), u(t_i) \rangle_1 = 1$$

and

$$\langle v(t_i), v(t_j) \rangle_2 = 0 \qquad i \neq j$$

$$\langle v(t_i), v(t_i) \rangle_2 = 1$$

But $\langle v(t_i), v(t_j) \rangle_2 = \langle \psi(u(t_i)), \psi(u(t_j)) \rangle_2 = 0$

and $\langle v(t_i), v(t_i) \rangle_2 = \langle \psi(u(t_i)), \psi(u(t_i)) \rangle_2 = 1$

ie. when $i \neq j$ $\langle u(t_i), u(t_j) \rangle_1 = \langle \psi(u(t_i)), \psi(u(t_j)) \rangle_2$

and

$$\langle u(t_i), u(t_i) \rangle_1 = \langle \psi(u(t_i)), \psi(u(t_i)) \rangle_2$$

for all $t_i, t_j \in T$.

So the inner product is preserved for all basis vectors and by the Basic Congruence Theorem H_1 and H_2 are congruent. Thus the useful result is obtained:

Any two abstract Hilbert Spaces of the same dimension are congruent.

In fact there are a multitude of congruences between H_1 and H as is easily seen: the particular correspondence between basis vectors in the above was not specified.

Now in the section on Linear Least Squares prediction the Hilbert Spaces used were $L_2\{X(t), t\in T\}$ and subspaces thereof. The elements of these spaces were random variables. Now if Hilbert Spaces of the same dimension as these are chosen, it should be possi to work in the congruent spaces instead of the original ones althoug the elements of such spaces do not look like and in fact are not random variables at all. The elements of such spaces will thus represent random variables. Any Hilbert Spaces of appropriate dim could be used to do this but it is best to choose spaces which are most suitable for the purposes in mind. Reproducing kernel Hilber Spaces appear to be best for the representation of random functions especially if the covariance function is chosen as the reproducing kernel.

2.7.3 Representation of Random Functions by Reproducing Kernel Hilbert Spaces.

The properties of reproducing kernel Hilbert Spaces have been dealt with quite comprehensively in section 1.6. Nevertheless there are some additional facts worth noting.

Let H be a Hilbert Space and let $M = L^*\{u(t), t\varepsilon T\}$ be a subset of it. Let g be a vector in H. Let $\langle g, u(t) \rangle = 0$ for all $t\varepsilon T$. By the projection theorem

$$\langle E^*[g|M], u(t) \rangle = \langle g, u(t) \rangle \quad \text{for all } t \text{ in } T.$$

Thus $\langle g, u(t) \rangle = 0$ implies $E^*[g|M] = 0$. Clearly unless $g = 0$, $g \neq E^*[g|M]$ ie. $g\varepsilon M$ if and only if $g = 0$. So $M = H$ if and only if the only vector satisfying $\langle g, u(t) \rangle = 0$ for all t is $g = 0$, or put another way:

The family $\{u(t), t\varepsilon T\}$ spans H if $g = 0$ is the only vector in H satisfying $\langle g, u(t) \rangle = 0$ for all t in T.

Now let K be a reproducing kernel for some Hilbert Space of functions f(s). $K(t,s)$ is a function of $s\varepsilon T$ for some fixed $t\varepsilon T$. So the set $\{K(t,s), t\varepsilon T\}$ is a family of functions of s, one for each $t\varepsilon T$. Now if $\langle f(s), K(t,s) \rangle_s = 0$ for each $t\varepsilon T$ (the inner product being taken over the index s), by the reproducing property $\langle K(t,s), f(s) \rangle_s = f(t)$ for all t in T. ie. $f(t) \equiv 0$ on T or $f = 0$ so the family of functions $\{K^t(s), t\varepsilon T\}$ spans H. (The notation $K^t(s)$ is often used in preference to $K(t,s)$ to emphasise that it is a function of one variable $s\varepsilon T$ for (various) fixed $t\varepsilon T$).

Now suppose that there are two Hilbert Spaces H and H' of the same dimension. There is a congruence between them, ψ, and they may be based on the same index set T. If K is the reproducing kernel for both of them and $g(t) \varepsilon H$, $g'(t) \varepsilon H'$, then

$$g(t) = \langle g(s), K(s,t) \rangle_H = \langle g'(s), K(s,t) \rangle_{H'} = g'(t)$$

by the definition of congruence and the reproducing property.

So H and H' are identical.

Now let $\{X(t), t\varepsilon T\}$ be a random function with covariance function $C(s,t)$ which must be symmetric and non negative. It is possible to form a Hilbert Space of functions $g(t)$, $t\varepsilon T$ which will be called $H(C)$ in the following manner.

For some unique random variable U in $L_2\{X(t), t\varepsilon T\}$, let

$$g(t) = E[UX(t)].$$

An inner product between the functions

$$g(t) = E[UX(t)] \quad \text{and} \quad h(t) = E[VX(t)]$$

can be defined as

$$\langle g, h \rangle = E[UV] ,$$

and the norm defined in the usual manner.

Obviously,

$$C(s,t) = C^s(t) = E[X(s)\, X(t)]$$

is in $H(C')$ and

$$\langle g, C^s \rangle = \langle g(t), C(s,t) \rangle = E[UX(s)] = g(s)$$

verifies that $C(s,t)$ is a reproducing kernel and it is quite simple to verify that $H(C)$ is a Hilbert Space. The family of functions $\{C^s(t), s\varepsilon T\}$ span $H(C)$ and only Hilbert Spaces identical to $H(C)$, so $H(C)$ is unique. In fact the above reasoning has shown that the Moore - Aronszajn - Loève Theorem, (Theorem 1.6.3) holds for random functions.

Now since $\langle C^{s_i}(t), C^{s_j}(t) \rangle_{H(C)} = C^{s_i}(s_j) = C(s_i,s_j)$

and $E[X(s_i), X(s_j)] = C(s_i,s_j)$ also,

and since $E[X(s_i)\ X(s_j)]$ is the inner product between the elements $X(s_i)$ and $X(s_j)$ in the Hilbert Space $L_2\{X(t),\ t\varepsilon T\}$, by the Basic Congruence theorem there is a congruence between $H(C)$ and $L_2\{X(t),\ t\varepsilon T\}$. Before going into detail, however it is necessary to define what is meant by a representation.

A Hilbert Space H is said to be a representation of a random function $\{X(t),\ t\varepsilon T\}$ if H is congruent to $L_2\{X(t),\ t\varepsilon T\}$. Or, because of the basic congruence theorem:

A family of vectors $\{u(t),\ t\varepsilon T\}$ in a Hilbert Space H is a representation of a random function $\{X(t),\ t\varepsilon T\}$ if, for every s and t in T

$$\langle u(s),\ u(t)\rangle_H = C(s,t) = E[X(s)X(t)]$$

In the special case of reproducing kernel Hilbert Spaces the family of vectors will, of course, be the family of functions $\{C^t(s),\ t\varepsilon T\}$ in the Hilbert Space H(C). So $\{C^t(s),\ t\varepsilon T\}$ is a representation of $\{X(t),\ t\varepsilon T\}$.

Any linear map ψ from H(C) to $L_2\{X(t),\ t\varepsilon T\}$ which has the property that for any $g\varepsilon H(C)$ and any $t\varepsilon T$,

$$E[\psi(g)\ X(t)] = g(t)$$

is the congruence which maps $C^t(s)$ to $X(t)$

ie $E[\psi(C^t(s))\ X(t)] = C^t(t) = C(t,t)$

which is true if $\psi(C^t(s)) = X(t)$.

If $\psi(g) = U$ then the equation

$E[UX(t)] = g(t)$ is obtained as before, so the inverse mapping ψ^{-1} can be denoted as an inner product in $L_2\{X(t),\ t\varepsilon T\}$ if one allows t to take any value in T.

It is also possible in view of Theorem 7C (PARZEN, 1959) to represent the congruence $\psi(g) = U$ as an inner product in the space H(C).

if $\langle X, g \rangle_C = 0$

for some function $X \in H(C)$ such that

$\langle X, C^t \rangle_C = X(t)$, the subscript C denoting that the inner product is taken with respect to the space H(C).

Now suppose that $\{X(t), t \in T\}$ is a random function with $L_2\{X(t), t \in T\} = H$ the Hilbert Space spanned by it. Let the covariance kernel be C. Now if T' is a subset of T then $L_2\{X(t), t \in T'\}$ is the subspace $H' \subseteq H$. Let R be the covariance kernel of H' such that $R(t,s) = C(t,s)$ for $s,t \in T' \subseteq T$.

Let Z be some random variable in H which is to be approximated by its projection on H'.

The projection theorem may be applied directly to H and H' to get the results previously obtained. However, the congruent spaces H'(R) and H(C) may be used instead and the projection theorem applied to $\psi^{-1}(Z) = h(t)$ where

$$h(t) = E[ZX(t)].$$

Let the projection of Z on H' be $\hat{Z}$ which maps under the congruence to $\hat{h}(t)$, where

$$\hat{h}(t) = E[\hat{Z}X(t)].$$

By the projection theorem, and since $R^{t_i}(t) = C^{t_i}(t)$ for any $t_i \in T' \subseteq T$,

$$\langle \hat{h}(t), C^{t_i}(t) \rangle_C = \langle h(t), C^{t_i}(t) \rangle$$

or

$$\langle \hat{h}, C^{t_i} \rangle_C = \langle h, C^{t_i} \rangle \qquad i = 1,2, \ldots n.$$

Now $\hat{h} = \sum_{j=1}^{n} \hat{d}_j C^{t_j}$ since $\hat{h} \in H'(R)$ and because of the reproducing property, $\langle h, C^{t_i} \rangle = h(t_i)$.

Thus

$$\sum_{j=1}^{n} d_j \langle C^{t_j}, C^{t_i} \rangle_C = h(t_i) \qquad i = 1,2, \ldots n.$$

Now $\langle C^{t_j}, C^{t_i} \rangle = C(t_i, t_j) = E[X(t_i)\ X(t_j)]$.

So in matrix form $G\underset{\sim}{d} = \underset{\sim}{h}$ where G is the same as in the statistical case. So $\underset{\sim}{d} = G^{-1}\underset{\sim}{h}$.

Now $$\hat{h} = \sum_{j=1}^{n} d_j C^{t_j}$$

or $$\hat{h}(t) = \sum_{j=1}^{n} d_j C^{t_j}(t)$$

At this stage it is worthwhile to compare this expression with the one obtained by application of the projection theorem to the statistical case ie. the spaces H and H'.

If one is only interested in prediction ie. X(t) is to be predicted at the point $t_p \varepsilon T$ rather than BX(t) where B is a linear operator, then $Z = X(t)$ and $\hat{Z} = \hat{X}(t)$.

Also $\psi(h) = X$.

Writing the above expression in matrices after rearranging,

$$\hat{h}(t) = [C^{t_1}(t)\ C^{t_2}(t) \ldots C^{t_n}(t)]\ G^{-1} \begin{bmatrix} h(t_1) \\ h(t_2) \\ \vdots \\ h(t_n) \end{bmatrix}$$

whereas in the statistical case

$$\hat{X}(t) = [C^{t_1}(t)\ C^{t_2}(t) \ldots C^{t_n}(t)]\ G^{-1} \begin{bmatrix} X(t_1) \\ X(t_2) \\ \vdots \\ X(t_n) \end{bmatrix}$$

Since $\psi(h) = X$ it follows from these expressions that $\psi(\hat{h}) = \hat{X}$ so that the predictions are congruent as would be expected.

There is a difference, however and that is that the next step in the statistical method is to apply the evaluation functional at a particular point $\omega \in A$ and obtain $\hat{X}^{\omega}(t)$ from the values $X^{\omega}(t_i)$ $i = 1,2,\ldots,r$ of a particular realisation.

In the deterministic case however the problem may be regarded as the interpolation of the function h(t) at the point t_p, the best value for this being $\hat{h}(t_p)$. So in the above $h(t_i)$ are values which may be considered to be already known - ie. h(t) is known at points t_i $i = 1,2, \ldots, n$ and its value at t_p is to be estimated. So the next step is to apply the evaluation functional with respect to a particular point $t_p \varepsilon T$.

Also in the statistical method the point t_p has been predetermined In the above expressions $C^{t_1}(t)$, $C^{t_2}(t)$ etc. are in fact $C^{t_1}(t_p)$, $C^{t_2}(t_p)$ and these values are already known numerically as $C^{t_1}(t_p)=C(t_1,t_p)=E[X(t_1$ etc.

So in effect the problem is being attacked from "opposite sides" although, as will be seen, the destinations are identical.

This is because $\langle \ell^{\omega}, X(t_i) \rangle_H = X^{\omega}(t_i) = h(t_i)$ $i = 1,2, \ldots$ in the statistical case using the inner product in H, and when the final operation is performed in the deterministic case, $\langle C^{t_p}, C^{t_i} \rangle_C = C^{t_i}(t_p) = C(t_i, t_p) = E[X(t_i)\, X(t_p)]$ $i = 1,2,\ldots,n$.

Thus applying the evaluation functionals to both cases

$$\hat{h}(t_p) = \langle \hat{h}, C^{t_p} \rangle_C = \langle \ell^{\omega}, \hat{X}(t_p) \rangle_H = \hat{X}^{\omega}(t_p) ,$$

and the two approaches are identical.

Now it is not a difficult matter to show that the approaches are identical for cases other than the pure prediction case ie. if so quantity BX(t) were to be predicted at $t = t_p$ or if instead of having the quantities $E[X(t_i)\, X(t_j)]$ the quantities $E[L_i X(t_i)\, J_j X(t_j)]$ only

were available for various L_i J_j where B L_i and J_j are linear operators. (DERMANIS, 1977) deals with this quite adequately.

The main point is that if the Covariance function C(t,s) is used as a reproducing kernel in the space H(C) one may regard the procedure of solving a partial differential equation where there is incomplete boundary value data and the procedure known as linear least squares prediction as essentially the same. The reason for this is that the Basic Congruence Theorem applies, but a more practical reason is that they always will give identical results.

Some discussion of (LAURITZEN, 1973) is in order: In Chapter 5 examples 4.1 and 4.2 he cites the cases of the Wiener process where $T = [0,1]$, $C(s,t) = \min(s,t)$ and $E(X(t)) = 0$ for all $t \in T$, and pure white noise where $T = R$, $C(s,t) = \delta_{st}$ (the Kronecker delta) and $E(X(t)) = 0$. In the case of the Wiener process he states: "... the sample functions of the Wiener process indeed behave very wildly! It can be shown that the sample functions cross any level more than a finite number of times in any small interval ... But note that the kernel space of the Wiener process entirely consists of functions which are <u>absolutely continuous</u> with square-integrable derivatives ... There is in fact a remarkable difference between the sample functions and the functions in the kernel space!" In the case of pure white noise: "The sample functions behave as wildly as functions possibly can. The kernel space consists of functions that are identically zero except in a denumerable set of points, and in such a way that $\sum_{t \in R} f(t)^2 < \infty$. The sample functions and the functions in the kernel space do not look like each other at all!" An example is also given where the "sample function space and the kernel space are identical", where $T = R$, $C(s,t) \equiv 1$ and $E(X(t)) \equiv 0$ with the comment: "But the process is not very interesting at all."

Now this demonstrates that the congruences between the spaces $L_2\{X(t), t\varepsilon T\}$ with covariance kernel C and the spaces H(C) are certainly not identities. That is to say, the congruences are of the form

$$\psi(C^t) = X(t) \quad \text{and not}$$

$$C^t = X(t).$$

It also serves to demonstrate the usefulness of working in the kernel space for one is often able to use well behaved functions rather than ones which "behave very wildly".

In Chapter 6, Lauritzen makes use of the concept of the Hilbert valued random variable, a mapping $X : \Omega \to U$ where U is a Hilbert Space and where for all $u^* \varepsilon U^*$, the dual space, the mappings $u^* \circ X : \Omega \to R$ are real valued random variables, to derive a Karhunen-Loève expansion in the case where the random variables $u^* \circ X$ are Gaussian. (DERMANIS, 1977) following Lauritzen's work with some minor changes of emphasis arrives at the same result without the Gaussian restricti Whilst Lauritzen does not mention the index set T, Dermanis stresses that his exposition is general and without restrictions on its nature for the case where $X(t), t\varepsilon T$ has sample functions in a Hilbert Space and can be regarded as a Hilbert-valued random variable. Then, by requiring U to be a reproducing kernel Hilbert Space with the reprod kernel K(s,t) ie. such that $K(s,t) = \sum_{n\varepsilon N} e_n(s)\, e_n(t)$ where $\{e_n(t)\}$ an orthonormal system in U, Lauritzen derives this result: $R(s,t) = E[X(s)\, X(t)]$ is the covariance function of a Gaussian proc with mean value zero the sample functions of which belong to U with probability one if and only if -

$$R(s,t) = \Sigma\sigma_n^2 \, e_n(s) \, e_n(t) \quad \text{for all } s,t, \varepsilon T$$

where $\Sigma\sigma_n^2 < +\infty$ are eigenvalues

of the operator $R : U \to U$ where

$$R(s,t) = \langle R K^t, K^s \rangle$$

Dermanis, on the other hand, derives Mercer's Theorem without the requirement that U be a reproducing kerne' Hilbert Space or that $X(t)$, $t\varepsilon T$ be Gaussian.

These results are quite remarkable, and perhaps a little brave, for usually Mercer's Theorem and the Karhunen-Loève expansion are derived for a finite closed interval $[a,b] \varepsilon R$ although (PARZEN, 1959) and (ZAANEN, 1953) state that the theory can be extended to the case where T is a closed bounded subset of a Euclidean space.

Indeed, Dermanis uses these results to show that the deterministic and statistical approaches are identical which indicates that he is not making use of the basic properties of the congruence between $L_2\{X(t), t\varepsilon T\}$ and H(C). It is even more obvious that Lauritzen is not for it would be somewhat difficult to find a reproducing kernel for $L_2\{X(t), t\varepsilon T\}$. In fact by noting that $R(s,t) = K(s,t)$ implies $\sigma_n^2 = 1$ for all $n \varepsilon N$, Lauritzen shows there is <u>no</u> Gaussian process with mean value zero and covariance function K(s,t) the sample functions of which belong to U with probability one. If use had been made of the congruence between $L_2\{X(t), t\varepsilon T\}$ and H(C) it would never have been proposed that H(C) comprised of all the sample functions of X(t). The congruence is not an identity mapping.

It is proposed to attempt something a little less ambitious than trying to show that Mercer's Theorem holds regardless of the nature of T. Instead,an attempt will be made to derive an orthogonal expansion for a random function on a specific T for a particular type

of covariance function, namely the case in question: the harmonic functions outside a sphere of radius R. First it is necessary to discuss the nature of the covariance function.

2.8 The Covariance Function

In this section it is really only necessary to deal with the ca where the index set T is a sphere. This is because any harmonic funct outside a sphere may be expressed in the form

$$h(\theta,\lambda,r) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{1}{r^{n+1}} [\bar{a}_{nm} \bar{R}_{nm}(\theta,\lambda) + \bar{b}_{nm} \bar{S}_{nm}(\theta,\lambda)]$$

whilst continuous functions on the sphere may be expressed

$$f(\theta,\lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\bar{a}_{nm} \bar{R}_{nm}(\theta,\lambda) + \bar{b}_{nm} \bar{S}_{nm}(\theta,\lambda)]$$

There is thus a one-one correspondence between continuous functions on the sphere and harmonic functions outside it. In other words for different realisations of a stochastic process it is only the $\bar{a}_{nm}$ and $\bar{b}_{nm}$ values which change regardless of which type of process is bei dealt with. Thus, for all intents and purposes the process on the sphere may be dealt with and simple modifications made when processes harmonic outside the sphere are of interest.

This is useful, for a sphere of radius R is a closed, bounded subset of E^3 and according to (PARZEN, 1959) and (ZAANEN, 1953), Merc Theorem holds and Karhunen-Loève expansions are possible.

It is thus not necessary to use more general results such as in (LAURITZEN, 1973) or (DERMANIS, 1977).

Now since the results obtained for deterministic and statist methods are identical where the reproducing kernel in the determinis method happens to be the true covariance function in the statistical case, it would appear to follow that the covariance kernel must be o of the permissible reproducing kernels of the deterministic case.

Otherwise, since the kernels in the deterministic case have been determined with a view to solution of Laplace's equation, the probabalistic case would not provide such a solution.

For the sphere, permissible reproducing kernels are of the form

$$K(s,t) = \sum_{n=0}^{\infty} k_{no}^2 P_n(\cos\psi) + \sum_{n=1}^{\infty} 2 \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_{nm}(\cos\theta_s) P_{nm}(\cos\theta_t) [(k_{nm}^2 - k_{no}^2)\cos m\lambda_s \cos m\lambda_t - (h_{nm}^2 - k_{no}^2)\sin m\lambda_s \sin m\lambda_t]$$

for two points $s : (\theta_s, \lambda_s)$ and $t : (\theta_t, \lambda_t)$ where ψ is the spherical distance between them. For each set of k_{nm}, $m,n = 0,1,2, \ldots$ h_{nm}, $m,n = 1,2, \ldots$ there is a different reproducing kernel. One such set of constants should yield the true covariance function.

This is not an isotropic kernel, but when the values of k_{nm} and h_{nm} are the same for all values of m at any particular value of n, the anisotropic term is zero and the isotropic kernel is obtained:

$$K(s,t) = \sum_{n=0}^{\infty} K_n^2 P_n(\cos\psi) .$$

If this were a covariance function for some random field on the sphere it would imply that such field were homogeneous and isotropic. In fact it is enough simply to say that it is homogeneous, for on the sphere homogeneity implies isotropy, a fact which is not difficult to show. (see also (OBUKHOV, 1947) and (MORITZ, 1980)).

However, a more fundamental expression for such a kernel is of greater interest, since by its use the Karhunen-Loève expansion of a random function on the sphere may be derived.

For the above kernel it may be recalled that the norm was given by

$$\| f \|_H = \frac{1}{4\pi} \sum_{j=1}^{\infty} \frac{2n+1}{k_j^2} \bar{c}_j^2 = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{2n+1}{K_n^2} \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]$$

An orthonormal basis for H was given by

$$\psi_j = k_j \sqrt{\frac{4\pi}{2n+1}}\, \phi_j$$

where the expression ϕ_j corresponds to $\bar{R}_{nm}$ or $\bar{S}_{nm}$ using the numbering system in (MORITZ, 1966)

Thence by Theorem 1.6.5

$$\bar{K}(s,t) = \sum_{n=0}^{\infty} \psi_n(s)\, \psi_n(t) .$$

For a continuous function on the sphere

$$f(\theta,\lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\bar{a}_{nm}\, \bar{R}_{nm}(\theta,\lambda) + \bar{b}_{nm}\, \bar{S}_{nm}(\theta,\lambda)]$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\hat{a}_{nm}\, \hat{R}_{nm}(\theta,\lambda) + \hat{b}_{nm}\, \hat{S}_{nm}(\theta,\lambda)]$$

where $\hat{R}_{no} = k_{no} P_n(\cos\theta)$

$$\hat{R}_{nm} = k_{nm} \sqrt{\frac{2(n-m)!}{(n-m)!}}\, \bar{R}_{nm} \; ; \; \hat{S}_{nm} = h_{nm} \sqrt{\frac{2(n-m)!}{(n+m)!}}\, \bar{S}_{nm}$$

$$\hat{a}_{nm} = \frac{\bar{a}_{nm}}{k_{nm}} \sqrt{\frac{2n+1}{4\pi}} \quad ; \quad \hat{b}_{nm} = \frac{\bar{b}_{nm}}{h_{nm}} \sqrt{\frac{2n+1}{4\pi}} ,$$

and where $\hat{R}_{nm}$ is an abreviation for $\hat{R}_{nm}(\theta,\lambda)$ etc.

The expressions $\hat{R}_{no}$, $\hat{R}_{nm}$, $\hat{S}_{nm}$ form an orthonormal set for H which correspond to the ψ_j under the (MORITZ, 1966) numbering system.

Hence

$$K(s,t) = \sum_{j=0}^{\infty} \psi_j(s)\ \psi_j(t)$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\hat{R}_{nm}(\theta_s,\lambda_s)\ \hat{R}_{nm}(\theta_t,\lambda_t) + \hat{S}_{nm}(\theta_s,\lambda_s)\ \hat{S}_{nm}(\theta_t,\lambda_t)]$$

Here it must again be stressed that $\hat{S}_{no}$ does not exist, but the above equation holds if by convention $\hat{S}_{no} \equiv 0$. ($\hat{R}_{nm}(s)$, $\bar{S}_{nm}(t)$ etc. will be used to denote $\hat{R}_{nm}(\theta_s,\lambda_s)$, $\bar{S}_{nm}(\theta_t,\lambda_t)$ etc. when a point t or s is specified and simply $\bar{R}_{nm}$, $\hat{S}_{nm}$ etc. when no particular point is specified).

Now it is known that

$$\int_\sigma \phi_m(t)\ \phi_n(t)d\sigma = \delta_{nm}$$

ie
$$\int_0^{2\pi} \int_0^{\pi} \bar{R}_{nm}(\theta,\lambda)\ \bar{R}_{pq}(\theta,\lambda)\ \sin\theta d\theta d\lambda = \delta_{np}\delta_{mq}$$

$$\int_0^{2\pi} \int_0^{\pi} \bar{R}_{nm}(\theta,\lambda)\ \bar{S}_{nm}(\theta,\lambda)\ \sin\theta d\theta d\lambda = 0$$

$$\int_0^{2\pi} \int_0^{\pi} \bar{S}_{nm}(\theta,\lambda)\ \bar{S}_{pq}(\theta,\lambda)\ \sin\theta d\theta d\lambda = \delta_{np}\delta_{mq}$$

Now

$$\int_0^{2\pi} \int_0^{\pi} \{\sum_{n=0}^{\infty} \sum_{m=0}^{n} [\hat{R}_{nm}(\theta_s,\lambda_s)\ \hat{R}_{nm}(\theta_t,\lambda_t)+\hat{S}_{nm}(\theta_s,\lambda_s)\hat{S}_{nm}(\theta_t,\lambda_t)]\}R_{pq}(\theta_s,\lambda_s)\sin\theta_s d\theta_s d\lambda_s$$

$$= \int_0^{2\pi} \int_0^{\pi} \bar{R}_{pq}(\theta_s,\lambda_s)\ \hat{R}_{pq}(\theta_s,\lambda_s)\ \hat{R}_{pq}(\theta_t,\lambda_t)\ \sin\theta_s d\theta_s d\lambda_s$$

$$= \hat{R}_{pq}(\theta_t,\lambda_t)\ \ k_{pq}\sqrt{\frac{4\pi}{2p+1}} \int_0^{2\pi} \int_0^{\pi} \bar{R}_{pq}(\theta_s,\lambda_s)\ \bar{R}_{pq}(\theta_s,\lambda_s)\ \sin\theta_s d\theta_s d\lambda_s$$

$$= \frac{4\pi k_{pq}^2}{2p+1}\ \bar{R}_{pq}(\theta_t,\lambda_t). \qquad p,q = 0,1,2, \ldots$$

ie. $$\int_{s\epsilon\sigma} K(s,t)\ \bar{R}_{pq}(s)\ ds = \frac{4\pi k_{pq}^2}{2p+1}\ \bar{R}_{pq}(t)$$

Similarly $$\int_{s\epsilon\sigma} K(s,t)\ \bar{S}_{pq}(s)\ ds = \frac{4\pi h_{pq}^2}{2p+1}\ \bar{S}_{pq}(t)$$

$$p,q = 1,2, \ldots$$

Or using the (MORITZ, 1966) numbering,

$$\int_{s\epsilon\sigma} K(s.t)\ \phi_i(s)\ ds = \frac{4\pi k_i^2}{2n+1}\ \phi_i(t) = \lambda_i \phi_i(t).$$

So $\phi_i(t)$ $i = 1,2, \ldots$ form an orthonormal set of eigenfunctions of the kernel K and λ_i $i = 1,2, \ldots$ are the corresponding non-negative eigenvalues.

By Mercer's Theorem then,

$$K(s,t) = \sum_{n=1}^{\infty} \lambda_n\ \phi_n(s)\ \phi_n(t)\ ,$$

the series converging absolutely and uniformly on $\sigma \times \sigma$.

Or, in the $\bar{R}_{nm}$, $\bar{S}_{nm}$ notation,

$$K(s,t) = \sum_{n=0}^{\infty} \frac{4\pi}{2n+1} \sum_{m=0}^{n} [k_{nm}^2\ \bar{R}_{nm}(s)\ \bar{R}_{nm}(t) + h_{nm}^2\ \bar{S}_{nm}(s)\ \bar{S}_{nm}(t)]\ ,$$

as is easily seen on substitution for $\hat{R}_{nm}$, $\hat{S}_{nm}$ in terms of $\bar{R}_{nm}$ and $\bar{S}_{nm}$ in the original expression.

It is noted that the eigenvalues are positive and that

$$\sum_{n=0}^{\infty} \frac{4}{2n+1} \sum_{m=0}^{n} [k_{nm}^2 + h_{nm}^2] < \infty\ .\quad (h_{no}^2 = 0 \text{ for all } n).$$

Now let $\{X(t),\ t\epsilon\sigma\}$ be a mean square continuous process which is measurable, with covariance kernel K(s,t), which has orthonormal eigenfunctions $\phi_n(t)$ and eigenvalues λ_n as described above.

Define the random variables

$$X_n = \frac{1}{\sqrt{\lambda_n}} \int_{t\epsilon\sigma} \phi_n(t)\, X(t)\, dt.$$

Then $E(X_n X_m) = \delta_{nm}$

$$X_{nm} = \sqrt{\frac{2n+1}{4\pi k_{nm}^2}} \int_0^{2\pi} \int_0^{\pi} X(\theta,\lambda)\, \overline{R}_{nm}(\theta,\lambda)\, \sin\theta d\theta d\lambda.$$

$$Y_{nm} = \sqrt{\frac{2n+1}{4\pi h_{nm}^2}} \int_0^{2\pi} \int_0^{\pi} X(\theta,\lambda)\, \overline{S}_{nm}(\theta,\lambda)\, \sin\theta d\theta d\lambda$$

$$E(X_{nm}X_{pq}) = \frac{2n+1}{4\pi k_{nm} k_{pq}} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_0^{\pi} E[X(\theta,\lambda)X(\theta',\lambda')]\overline{R}_{nm}(\theta,\lambda)\overline{R}_{pq}(\theta',\lambda')$$

$$\times \sin\theta' \sin\theta d\theta' d\theta d\lambda' d\lambda,$$

by Fubini's theorem.

$$= \frac{k_{pq}^2}{k_{nm} k_{pq}} \int_0^{2\pi} \int_0^{\pi} \overline{R}_{nm}(\theta,\lambda)\, R_{pq}(\theta,\lambda)\, \sin\theta d\theta d\lambda$$

$$= \delta_{np}\, \delta_{mq}$$

Similarly $E(X_{nm}\, Y_{pq}) = 0$ always

and $E(Y_{nm}\, Y_{pq}) = \delta_{np}\, \delta_{mq}$ $n \neq 0$.

Again using Fubini's theorem,

$$E(X(t)\, X_n) = \frac{1}{\sqrt{\lambda_n}} \int_{t\epsilon\sigma} E(X(t)\, X(t))\, \phi_n(t)\, dt$$

$$= \frac{1}{\sqrt{\lambda_n}} \int_{t\epsilon\sigma} K(t,t)\, \phi_n(t)\, dt$$

$$= \sqrt{\lambda_n}\, \phi_n(t) \qquad t\epsilon\sigma.$$

So $$E[X(t) - \sum_{n=1}^{N} \sqrt{\lambda_n}\, \phi_n(t)\, X_n]^2$$

$$= E[X(t)]^2 + \sum_{n=1}^{N} \lambda_n [\phi_n(t)]^2 - 2 \sum_{n=1}^{N} \phi_n(t) \sqrt{\lambda_n}\, E\{X(t)\, X_n\}$$

$$= K(t,t) - \sum_{n=1}^{N} \lambda_n [\phi_n(t)]^2$$

which tends to zero as $N \to \infty$, by Mercer's Theorem.

So $X(t)$ has the orthogonal expansion

$$X(t) = \lim_{N \to \infty} \text{in q.m.} \sum_{n=1}^{N} \sqrt{\lambda_n}\, \phi_n(t)\, X_n$$

This expansion is known as the Karhunen-Loève expansion and in terms of $\bar{R}_{nm}$ and $\bar{S}_{nm}$ is expressed as

$$X(t) = \sum_{n=0}^{\infty} \sqrt{\frac{4\pi}{2n+1}} \sum_{m=0}^{n} [k_{nm}\, X_{nm}\, \bar{R}_{nm}(\theta,\lambda) + h_{nm}\, Y_{nm}\, \bar{S}_{nm}(\theta,\lambda)].$$

For a particular realisation $X^{\omega}(t)$, X^{ω}_{nm} and Y^{ω}_{nm} are constants so the right hand side is a continuous function. But the sample functions are also continuous, hence the equality.

The expansion does not rely on homogeneity of the covariance function, just as in the one dimensional case stationarity is not necessary. It is not difficult to obtain the expansion for the case where the covariance function is homogeneous and isotropic.

Here $k_{nm} = k_{no} = K_n$, $n,m, = 0,1,2, \ldots$ and

$$h_{nm} = K_n, \quad n,m, = 1,2, \ldots$$

Thus $$X(t) = \sum_{n=0}^{\infty} K_n \sqrt{\frac{4\pi}{2n+1}} \sum_{m=0}^{n} [X_{nm}\, \bar{R}_{nm} + Y_{nm}\, \bar{S}_{nm}]$$

$$= \sum_{n=0}^{\infty} K_n \sum_{m=0}^{n} [X_{nm}\, \hat{R}_{nm} + Y_{nm}\, \hat{S}_{nm}] .$$

Also because $E[X_{nm}X_{pq}] = \delta_{np}\delta_{mq} = E[Y_{nm}Y_{pq}]$ and $E[X_{nm}Y_{pq}] = 0$ it is not difficult to show that

$E[X(s)\ X(t)] = K(s,t)$ for both the isotropic and anisotropic cases.

Evidently, under the congruence J from $L_2\{X(t),\ t\varepsilon\sigma\}$ to $H(K)$, the following relationships hold:

$$J[X_{nm}] = \hat{R}_{nm}(t) = k_{nm}\sqrt{\frac{4\pi}{2n+1}}\,\overline{R}_{nm}(t)$$

$$J[Y_{nm}] = \hat{S}_{nm}(t) = h_{nm}\sqrt{\frac{4\pi}{2n+1}}\,\overline{S}_{nm}(t)$$

This follows from (PARZEN, 1959) Theorem 5D, since

$$E[J^{-1}[\hat{R}_{nm}(t)]X(t)] = E[X_{nm}X(t)] = \hat{R}_{nm}(t)$$

$$E[J^{-1}[\hat{S}_{nm}(t)]X(t)] = \hat{S}_{nm}(t)$$

Also,

$$J[X(t_i)] = K^{t_i}(t)$$

$$= \sum_{n=0}^{\infty}\frac{4\pi}{2n+1}\sum_{m=0}^{n}[k_{nm}^2\overline{R}_{nm}(t_i)\ \overline{R}_{nm}(t)+h_{nm}^2\overline{S}_{nm}(t_i)\ \overline{S}_{nm}(t)],$$

in the case of an inhomogeneous kernel, and

$$J[X(t_i)] = \sum_{n=0}^{\infty}K_n^2\,P_n(\cos\psi_{t_i,t})$$

in the case of a homogeneous one, where $\psi_{t_i,t}$ is the spherical angle between some point $t_i\varepsilon\sigma$ and a variable point t. It is thus a function of t.

It is easily verified that

$$E[\sum_{i=1}^{n} a_i X(t_i) \sum_{j=1}^{m} b_j X(t_j)] = \langle \sum_{i=1}^{n} a_i K^{t_i}, \sum_{j=1}^{m} b_j K^{t_j} \rangle_{H(K)}$$

for $t_i t_j \varepsilon \sigma$,

and that the results hold for limits of Cauchy sequences.

At the risk of being accused of repetitiveness it is once more noted that

$$\delta(s,t) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\overline{R}_{nm}(s)\, \overline{R}_{nm}(t) + \overline{S}_{nm}(s)\, \overline{S}_{nm}(t)]$$

$$= \frac{1}{4\pi} - \frac{1}{4\pi} \sum_{n=0}^{\infty} (2n+1)\, P_n(\cos\psi_{s,t})$$

and where $s = t$ (or $\psi_{s,t} = 0$)

$$\delta(s,s) = \frac{1}{4\pi} + \frac{1}{4\pi} \sum_{n=0}^{\infty} (2n+1) = \infty$$

That is to say, the space of sample functions X(t) regarded as real-valued functions on σ ie. $L_2[\sigma]$ do not form a reproducing kernel Hilbert Space. In other words the space H(K) is not a space of sam functions. (This is the significance of the statement by (LAURITZE 1973): "there is no Gaussian process with mean value zero and covar function r the sample functions of which belong to U with probabili one".)

(t)],

2.9 Estimating the Covariance Function: Ergodicity

The Karhunen-Loève expansion may be written for X(t) in a slightly modified form

$$X(t) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [A_{nm}\overline{R}_{nm}(\theta,\lambda) + B_{nm}\overline{S}_{nm}(\theta,\lambda)]$$

where $A_{nm} = \sqrt{\frac{4\pi}{2n+1}}\, k_{nm} X_{nm}$, $B_{nm} = \sqrt{\frac{4\pi}{2n+1}}\, h_{nm} Y_{nm}$

are random variables such that

$$E[A_{nm}\, A_{pq}] = k_{nm} k_{pq} \sqrt{\frac{4\pi}{2n+1}}\, \delta_{np} \delta_{mq} \text{ , } E[A_{nm}\, B_{pq}] = 0 \text{ always,}$$

and $$E[B_{nm}\, B_{pq}] = h_{nm} h_{pq} \sqrt{\frac{4\pi}{2n+1}}\, \delta_{np} \delta_{mq}.$$

A particular realisation is given by

$$X^{\omega}(t) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\bar{a}_{nm}\, \bar{R}_{nm}(\theta,\lambda) + \bar{b}_{nm}\, \bar{S}_{nm}(\theta,\lambda)]$$

ie. $$\bar{a}_{nm} = k_{nm}\, X^{\omega}_{nm} \sqrt{\frac{4\pi}{2n+1}} = A^{\omega}_{nm};\ \bar{b}_{nm} = h_{nm}\, Y^{\omega}_{nm} \sqrt{\frac{4\pi}{2n+1}} = B^{\omega}_{nm} \text{ ,}$$

X^{ω}_{nm}, Y^{ω}_{nm} being realisations of X_{nm} and Y_{nm} .

Attempts have been made to utilise a realisation or a "partial" realisation to estimate the covariance function K(s,t). The most common method is quite similar to that used in the estimation of the covariance function of a one dimensional stationary time series where ergodicity is assumed. It is called the "empirical covariance function", and mainly for computational reasons it is usually assumed that the process is a homogeneous random field. For an inhomogeneous and anisotropic field it would be impossible to compute a sensible estimate, just as it would be for a non-stationary time series if only a single record were obtainable. After trend removal it may be possible to regard the resulting field as homogeneous, however. On the plane, or in R^n it is possible to have anisotropic random fields which are homogeneous but this type of field is precluded when the sphere is the index set, since here homogeneity implies isotropy. It is, of course, possible to have an inhomogeneous random field which is isotropic at two points diametrically opposite each other on the sphere

but at no other point, as mentioned by (JONES, 1963) in connection with meteorological forecasting. In terms of the gravity field of the earth one would expect gravity measurements to be of this nature. However by subtracting normal gravity, as determined by a model, to obtain the gravity anomaly this trend is deemed to have been removed.

In effect then, the assumption of homogeneity is the same as the assumption that all trends likely to cause inhomogeneity have been removed. If the trends are expressed as the first few terms of an expansion in spherical harmonies, the resultant or residual field will still be harmonic outside the sphere or continuous on it.

The assumption of homogeneity therefore should not be regarded as highly controversial. If there is any doubt as to its validity the remedy is to use better models for trend removal rather than invalid concepts such as "local anisotropy" within a homogeneous field. If a random field is locally inhomogeneous it can hardly be globally homogeneous. Also the benefits computationally are of great importance: in fact they are the difference between being able at least to attempt to estimate the covariance function and not being able to do it at all.

The formula for the empirical covariance function is

$$R(s,t) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} X^{\omega}(\theta_s,\lambda_s)\, X^{\omega}(\theta_t,\lambda_t)\, \sin\theta d\theta d\lambda .$$

$X^{\omega}(\theta_s,\lambda_s)$, being the realisation of the function in question, (usually gravity anomaly), at the point (θ_s,λ_s) etc.

As shown by (HEISKANEN & MORITZ, 1967), for an isotropic homogeneous random field the expression in terms of spherical harmonics is a function of $\psi_{s,t}$, the spherical distance from s to t and may be written

$$R(s,t) = R(\psi_{s,t}) = \sum_{n=0}^{\infty} R_n P_n(\cos\psi_{s,t})$$

where $$R_n = \frac{1}{4\pi} \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]$$

Usually the index n runs from 3 to ∞ since the terms with index n=0,1,2 represent known Geodetic parameters, ie. some trend removal has taken place in the calculation of gravity anomalies.

The R_n terms are known as "degree variances" or variances at degree n because of the similarity to the expression

$$\frac{1}{4\pi} \sum_{m=0}^{n} \{E[A_{nm}^2] + E[B_{nm}^2]\} = \frac{1}{4\pi} \sum_{m=0}^{n} \frac{4\pi}{2n+1} [k_{nm}^2 + h_{nm}^2]$$

$$= K_n^2 \quad \text{since } k_{nm}^2 = h_{nm}^2 = K_n^2 ,$$

$$h_{no}^2 = 0, \text{ for an isotropic kernel.}$$

The expression differs from Moritz's by a factor of $\frac{1}{4\pi}$ but this is only because the inner product was chosen such that $\int_\sigma \bar{R}_{nm}\bar{R}_{pq} d\sigma = \delta_{np}\delta_{mq}$ rather than $\frac{1}{4\pi}\int_\sigma \bar{R}_{nm}\bar{R}_{pq} = \delta_{np}\delta_{mq}$ etc. It is in fact the (TSCHERNING, 1970) notation.

So $\bar{a}_{nm}^2$ corresponds to $E[A_{nm}^2]$ and $\bar{b}_{nm}^2$ to $E[B_{nm}^2]$ and they may thus be regarded as estimates of variances, hence the term "degree variance". They are estimates and not estimators, for the estimator of the covariance function is in fact

$$\hat{K}(s,t) = \sum_{n=0}^{\infty} \hat{K}_n^2 P_n(\cos\psi_{s,t})$$

where $$\hat{K}_n^2 = \frac{1}{4\pi} \sum_{m=0}^{n} [A_{nm}^2 + B_{nm}^2]$$

which is unbiassed, since $E[\hat{K}_n^2] = K_n^2$.

For a particular realisation, $A^{\omega}_{nm} = \bar{a}_{nm}$ and $B^{\omega}_{nm} = \bar{b}_{nm}$, an estimate is obtained, so the empirical covariance function may be deemed unbiassed.

Now it should be remembered that the formula for the empirical covariance function is just that: a formula. It has not been derived from inviolable precepts by a rigorous statistical treatment but is merely proposed as a method of estimating the true covariance function from a single record or realisation. In this it is not altogether different from most estimators, for, in general, one proposes a theory and derives an estimator using it. The next step is to examine the various properties of the estimator, often in comparison with other estimators, and see whether it adequately suits the purposes for which it is intended.

In the case of the empirical covariance function it is seen that it possesses one desirable property, that of unbiassedness, but it appears to be somewhat difficult to find other properties which are not undesirable. The formula seems to have been proposed because of its similarity to the one used in computing the covariance function on the real line for a time series, which is given by the time average

$$\hat{R}_T(\lambda) = \frac{1}{2T} \int_{-T}^{T} X(t + \lambda)\, X(t)\, dt.$$

In turn, this formula has been widely used for temporal processes possibly because of its resemblance to the following formula for the sample correlation coefficient of two variables X and Y:

$$\hat{R} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}\ \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

where $\overline{X}$, $\overline{Y}$ are the respective sample means. The average here is over the probability space rather than over the time axis and is equivalent to estimating the covariance between X and Y in the case where they are known to have zero means by

$$\hat{C} = \sum_{i=1}^{n} X_i Y_i$$

A slight modification of the variables X_i and Y_i to X_{t_i}, $X_{t_i+\lambda}$ followed by conversion to the continuous analogue gives $\hat{R}_T(\lambda)$.

Now $\hat{R}$ was derived using the method of maximum likelihood on the assumption that X and Y were normally distributed. See, for example, (HOEL, 1954). It would appear therefore, that some assumption to the effect that the process is Gaussian has influenced the choice of $\hat{R}_T(\lambda)$ in computing the sample covariance function of stationary time series.

For this estimator to be of use, an ergodicity assumption must also be made ie. that time averages equal ensemble averages. In the case of a time series, for ergodicity of the autocorrelation the following is required:

$$\lim_{T \to \infty} \hat{R}_T(\lambda) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} X(t+\lambda)\, X(t)\, dt$$

$$= E[X(t+\lambda)\, X(t)] = R(\lambda).$$

According to a theorem in (PAPOULIS, 1965) this is true if and only if the variance of the random variable $\hat{R}_T(\lambda)$ tends to zero as T tends to infinity: ie. if

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

where $R_{\phi\phi}(\tau) = E[X(t+\lambda+\tau)\, X(t+\tau)\, X(t+\lambda)\, X(t)]$

ie. $[R_{\phi\phi}(\tau) - R^2(\lambda)]$ is the autocovariance of the process $X(t+\lambda)\ X(t)$, so to test for ergodicity of the autocorrelation, knowledge of fourth order moments is required.

Now in the formula for the empirical covariance function, $\frac{1}{4\pi}$ outside the integral sign corresponds to $\frac{1}{2T}$ outside the integral sign in the time series case ie.

$$R(s,t) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} X^{\omega}(\theta_s,\lambda_s)\ X^{\omega}(\theta_t,\lambda_t) \sin\theta d\theta d\lambda$$

or for the estimator $\hat{K}(s,t)$ of which $R(s,t)$ is a particular estimate:

$$\hat{K}(s,t) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} X(\theta_s,\lambda_s)\ X(\theta_t,\lambda_t)\ \sin\theta d\theta d\lambda, \quad \text{and}$$

$$\int_0^{\pi} \int_0^{\pi} \sin\theta d\theta d\lambda = 4\pi,$$

just as $\int_{-T}^{T} dt = 2T$ in the case of the autocovariance function $\hat{R}_T(\lambda)$

Now it is difficult to see exactly how one can perform for the empirical covariance function, a limiting analogy to that of the time series case where the variance of $\hat{R}_T(\lambda)$ is a function of T and T is allowed to go to infinity. 4π is not a variable and obviously cannot do the same thing. A misconception is that the process of "filling the gaps" of a discrete record of gravity anomalies so that gravity is known all over the earth and a summation becomes an integral is the same as allowing T to go to infinity. This is definitely not the case; just as a time series of, say, electrical current through a resistor measured at discrete intervals over a period T, will not yield a particularly different estimate of the covariance function to that calculated from a continuous record over exactly the same period. The fact that there is a

continuous record does not alter the fact that the record is not over the entire real line.

One may thus doubt whether ergodicity is a particularly appropriate concept or even a reasonable assumption for an index set such as a sphere or interval of the real line. For time averages to equal ensemble averages in such cases, then with probability one the same empirical covariance function would be obtained for each sample chosen. In other words, if there were a multitude of readily available planets with exactly the same geological or cosmological history as the earth it would be possible to get ensemble averages and thus a good estimate of the true covariance function. It does not seem likely, however, that in each case the empirical covariance function would be the same and thus be the same as the true covariance function. That is not to say that it is impossible, just highly unlikely. The prospect of actually obtaining such repeated trials is even more unlikely.

Ergodicity, therefore, is much more restrictive in the case of a finite interval or a sphere. The limit as T tends to infinity of the variance is unobtainable, so it must be demanded that the variance of $\hat{K}(s,t)$ itself be equal to zero.

(LAURITZEN, 1973) has shown that in the case where the process is Gaussian the variance of $\hat{K}(s,t)$ is indeed not zero. To do this he used topological group theory but it does not seem necessary to go to such lengths, for elementary probability concepts appear to be quite adequate in this regard.

In general the estimator corresponding to the empirical covariance function may be expressed as

$$\hat{K}(s,t) = \sum_{n=0}^{\infty} \hat{K}_n^2 P_n(\cos\psi_{s,t})$$

where $\hat{K}_n^2 = \frac{1}{4\pi} \sum_{m=0}^{n} [A_{nm}^2 + B_{nm}^2]$

and, as stated before

$$A_{nm} = \sqrt{\frac{4\pi}{2n+1}}\, k_{nm} X_{nm} \; ; \quad B_{nm} = \sqrt{\frac{4\pi}{2n+1}}\, h_{nm} Y_{nm}$$

(Note that Y_{nm} and hence B_{nm} do not exist for $n = 0$).

In the case of a Gaussian process X_{nm} and Y_{nm} are identically and independently distributed as N(0,1).

Thus, $$\hat{K}_n^2 = \frac{1}{4\pi} \left(\frac{4\pi}{2n+1}\right) \sum_{m=0}^{n} [k_{nm}^2 X_{nm}^2 + h_{nm}^2 Y_{nm}^2]$$

Now the covariance function is isotropic and homogeneous and thus $K_n^2 = h_{nm}^2 = k_{nm}^2$, so

$$\hat{K}_n^2 = \frac{K_n^2}{2n+1} \sum_{m=0}^{n} [X_{nm}^2 + Y_{nm}^2]$$

$$= \frac{K_n}{2n+1} \sum_{i=1}^{2n+1} U_i^2$$

where U_i, $i = 1,2, \ldots 2n+1$ are identically and independently distributed as N(0,1) and thus U_i^2, $i = 1,2, \ldots, 2n+1$ are independently chi-squared with one degree of freedom. Thus by the additive property of the chi-squared distribution,

$$\hat{K}_n^2 = \frac{K_n^2}{2n+1} Q^2 \quad \text{where } Q^2 : \chi^2(2n+1)$$

Hence $$V(\hat{K}_n^2) = \frac{K_n^4}{(2n+1)^2} V(Q^2)$$

$$= \frac{2K_n^4}{2n+1} .$$

As a direct consequence of the fact that if X and Y are independent and g(X) and h(Y) are functions of these variables then $E[g(X)\,h(Y)] = E[g(X)]E[h(Y)]$, it is easily seen that

$$\mathrm{Cov}\,(\hat{K}_m, \hat{K}_n) = 0 \quad \text{for } m \neq n.$$

Thus
$$V[\hat{K}(s,t)] = \sum_{n=0}^{\infty} V[\hat{K}_n^2\, P_n(\cos\psi_{s,t})]$$

$$= 2\sum_{n=0}^{\infty} \frac{K_n^4}{2n+1} [P_n(\cos\psi_{s,t})]^2$$

This is obviously not zero, so the process is not ergodic.

The expression in (LAURITZEN, 1973) for the empirical covariance function of the gravity anomaly is

$$R(\Delta g(s),\Delta g(t)) = \sum_{n=3}^{\infty} (2n+1)(n-1)^2 \sigma_n^2 \frac{R^{2n+2}}{(r_s r_t)^{n+2}} P_n(\cos\psi_{s,t})$$

Converting to the present notation and letting $r_s r_t = r^2$ the expression becomes

$$\hat{K}(\Delta g(s),\Delta g(t)) = \sum_{n=3}^{\infty} \hat{K}_n^2 \left[(2n+1)(n-1)^2 \left(\frac{R}{r}\right)^{2n} \left(\frac{R}{r^2}\right)^2 P_n(\cos\psi_{s,t})\right]$$

Thus $V[\hat{K}(\Delta g(s),\Delta g(t))] = \sum_{n=3}^{\infty} \left[(2n+1)(n-1)^2 \left(\frac{R}{r}\right) \left(\frac{R}{r^2}\right)^2 P_n(\cos\ s,t)\right]^2 V(\hat{K}_n^2)$,

$$= 2\left(\frac{R}{r^2}\right)^4 \sum_{n=3}^{\infty} (n-1)^4 (2n+1)\, K_n^4 \left(\frac{R}{r}\right)^{4n} [P_n(\cos\psi_{s,t})]^2$$

which with σ_n^4 substituted for K_n^4 is Lauritzen's expression for the variance. So Lauritzen's expression is verified by methods not involving topological group theory.

This result led Lauritzen to say that it seems to be a more difficult problem to find the true covariance function than to find the value of the potential (or gravity anomaly) at a point, since

if gravity were known at all points on the earth's surface, the potential could be found but the true covariance function could not, and that the problem does not seem suited to statistical treatment.

Perhaps it is not suited to a classical statistical treatment, ie. where probability is defined in a relative frequency sense, but it would be an interesting exercise to apply Bayesian methods to the problem. It does not appear that this has been done in the literature and unfortunately it is a little beyond the scope of this project.

Nevertheless, Lauritzen concluded that it would be better to choose a kernel that reflects the observed behaviour of the actual potential rather than an arbitrary one, and for this reason it is desirable to get a covariance function which is easy to handle and not too far away from the empirical one. It was found that a potential covariance function

$$R(s,t) = \sum_{n=3}^{\infty} \frac{A}{(n-1)(n-2)} \left(\frac{R^2}{r_s r_t}\right)^{n+1} P_n(\cos\psi_{s,t})$$

gave a good approximation to the empirical covariance function calculated from the data in (KAULA, 1966). Closed expressions for this are also given in (LAURITZEN, 1973). This covariance function corresponds to

$$\hat{K}_n^2 = \sigma_n^2 = \begin{cases} \dfrac{A}{(n-1)(n-2)(2n-1)} & \text{if } n \geq 3 \\ 0 & \text{if } n \leq 2 \end{cases}$$

(LAURITZEN, 1975) and (TSCHERNING, 1976) showed that it is possible to obtain this expression from the point of view of mass density anomaly functions, and similar methods of computing values for the various covariance functions are still in use today. For example (TSCHERNING & FORSBERG, 1981) quote an expression for the covariance function of the disturbing potential used in computation of the geoic in the Norwegian-Greenland Sea area as

$$\text{Cov}\,(T(P),T(Q)) = \sum_{i=2}^{180} (2i+1)(3\times10^{-9})\left(\frac{GM}{R}\right)^2 + \sum_{i=181}^{\infty} \frac{300\ \text{mgal}^2/R^2}{(i-1)(i-2)(i+24)} \left(\frac{R_b^2}{r_P r_Q}\right)^{i+1} P_i(\cos\psi).$$

It was chosen to approximate an empirically determined covariance function.

However it is somewhat difficult to understand why there is so much emphasis placed upon obtaining something not quite as good as an empirical covariance function, which is itself an inconsistent estimator of the true covariance function in any case. The lack of the ergodic property is, however, a comparatively minor problem when compared with the fact that the values predicted using the empirical covariance function are not elements of the Hilbert Space being used (ie. $L_2\{X(t),\ t\varepsilon\sigma\}$ or $H(K)$).

The norm squared of an element f of this space is given by

$$\|f\|^2 = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{2n+1}{K_n^2} \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]$$

As was shown in (TSCHERNING, 1977) if K_n^2 is estimated by

$$\hat{K}_n^2 = \frac{1}{4\pi} \sum_{m=0}^{n} [A_{nm}^2 + B_{nm}^2]$$

then for the particular realisation of the earth

$$A_{nm}^{\omega} = \bar{a}_{nm} \quad \text{and} \quad B_{nm}^{\omega} = \bar{b}_{nm}$$

ie. $R(s,t) = \sum_{n=0}^{\infty} R_n^2 P_n(\cos\psi_{s,t})$ where

$$R_n^2 = \frac{1}{4\pi} \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2],$$ and thus the norm squared is

$$\|f\|^2 = \sum_{n=0}^{\infty} (2n+1) = \infty$$

This is basically the same as saying that the variance of the predicted quantity is infinite.

If the expression for K_n^2 as suggested by Lauritzen is used,

ie. $$K_n^2 = \frac{A}{(n-1)(n-2)(2n-1)} ,$$

the norm squared of f becomes

$$\| f\|^2 = \sum_{n=3}^{\infty} \frac{(n-1)(n-2)(4n^2-1)}{A} \sum_{n=0}^{m} (\bar{a}_{nm}^2 + \bar{b}_{nm}^2).$$

If the sequence of numbers $\sum_{n=0}^{m} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]$ is of the order $\frac{1}{n^{5+\varepsilon}}$ then $\| f\|^2$ will be finite where $\varepsilon > 0$. It is not known if this can be guaranteed, but one would expect that such would not be the case, if the expression gives a close approximation to the empirical covariance function. It is to be noted also that this divergence happens, no matter what the distribution of the process is, whereas the ergodicity problem was a consequence of the process being Gaussia

Referring to the derivation of the variance of $\hat{K}(s,t)$ it is seen that use is made of the fact that if X and Y are independent and g and h are functions then Cov $(g(X)\ h(Y)) = 0$.

However this cannot be guaranteed when X and Y are merely uncorrelated. The expression for the variance would become much more complicated because of cross product terms, and will not be reproduced here. It is certain however, that it is not impossible for the variance to be zero.

Now according to (GIKMAN & SKOROKHOD, 1969) independence of the random variables X_{nm}, Y_{nm} implies that they are uncorrelated, orthogonal and Gaussian. Thus the lack of ergodicity follows directly from the fact that the process is assumed to be Gaussian.

If the process is not assumed to be Gaussian of course this does not mean that it is automatically ergodic, merely that it may be. In fact (MORITZ, 1980) gives two examples of processes which are ergodic, both of which are only of interest because of this fact. They have little chance of occurring in practice, and one of them is trivial merely being the same sample rotated by arbitrary angles in space to make the samples appear different.

The problem then appears to be that the formula used to compute the empirical covariance function is somewhat inappropriate. It is based on a Gaussian assumption which, if it were true, would preclude ergodicity. More importantly when used in linear least squares prediction it gives results which have infinite variance (when regarded as estimators) and are thus not realisations of second order random functions at the particular point in space. In (MORITZ, 1980) it is suggested that instead of

$$R_n = \frac{1}{4\pi} \sum_{m-0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]$$

the expression $R_n(1 + \varepsilon)^n$ where ε is arbitrarily small be used in the expression for the empirical covariance function, giving

$$R(s,t) = \frac{1}{4\pi} \sum_{n=0}^{\infty} (1 + \varepsilon)^n \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2] \; P_n(\cos\psi_{s,t}) \; .$$

Substituting this in the expression for $\| f \|^2$,

$$\| f \|^2 \sum_{n=0}^{\infty} \frac{2n+1}{(1+\varepsilon)^n} \quad \text{is obtained.}$$

Using d'Alembert's ratio test it is seen that $\| f \|^2$ is finite. However there is some difficulty with the boundedness of the kernel $R(s,t)$. When $\psi_{s,t} = 0$, (ie. $P_n(\cos\psi_{s,t}) = 1$),

$$R(0) = \frac{1}{4\pi} \sum_{n=0}^{\infty} (1+\varepsilon)^n \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2].$$

The expression which fits the empirical covariance function well, according to Lauritzen, is of the order of $\frac{1}{n^4}$ so it is well to examine the convergence of the series

$$\sum_{n=0}^{\infty} \frac{(1+\varepsilon)^n}{n^4} \quad , \quad \text{using Cauchy's } n^{th} \text{ root test.}$$

Now $\left\{\frac{(1+\varepsilon)^n}{n^4}\right\}^{\frac{1}{n}} = \frac{1+\varepsilon}{n^{4/n}}$ and $\lim_{n \to \infty} \frac{1+\varepsilon}{n^{4/n}} = 1+\varepsilon > 1,$

so it would appear that the kernel is unbounded.

Regarded as an estimator it is biassed:

$$E\{(1+\varepsilon)^n \frac{1}{4\pi} \sum_{m=0}^{n} [A_n^2 + B_n^2]\} = (1+\varepsilon)^n K_n^2 .$$

So unless the true values of K_n^2 in the true covariance function converge very rapidly the estimator is quite biassed, no matter how small ε is.

$$\text{ie.} \quad E[(1+\varepsilon)^n \hat{K}(s,t)] = \sum_{n=0}^{\infty} (1+\varepsilon)^n K_n^2 P_n(\cos\psi_{s,t})$$

A suggestion made in section 1.8 of this project was to use some function involving the square root or some power of R_n.

For example if $C(s,t) = \sum_{n=0}^{\infty} c_n P_n(\cos\psi_{s,t})$ where

$$c_n = \frac{1}{4\pi} \sqrt{(2n+1) \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]}$$ the following results are obtained

$$\|f\|^2 = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{2n+1}{\frac{1}{4\pi}\sqrt{(2n+1)\sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]}} \sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]$$

$$= \sum_{n=0}^{\infty} \sqrt{2n+1} \sqrt{\sum_{m=0}^{n} [\bar{a}_{bn}^2 + \bar{b}_{nm}^2]}$$

Now if $\sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]$ is of the order $\frac{1}{n^4}$, then the nth term of $\|f\|^2$ is of the order $\left(\frac{R}{n^4}\right)^{\frac{1}{2}}$ or $\frac{1}{n^{3/2}} = \frac{1}{n^{1+\frac{1}{2}}}$

and $\sum_{n=0}^{\infty} \frac{1}{n^{1+\frac{1}{2}}}$ is finite.

By the same token for $s = t$ ie. $\psi_{s,t} = 0$

$$C(0) = \frac{1}{4\pi} \sum_{n=0}^{\infty} \sqrt{2n+1} \sqrt{\sum_{m=0}^{n} [\bar{a}_{nm}^2 + \bar{b}_{nm}^2]}$$

which also converges. So a bounded kernel is obtained which also gives predicted values whose norms are bounded. C(s,t) is not to be seriously regarded as an estimate of the true covariance function. It would be difficult to find a distribution, for example, where the corresponding estimator

$$\hat{C}(s,t) = \frac{1}{4\pi} \sum_{n=0}^{\infty} \sqrt{2n+1} \left[\sum_{m=0}^{n} [A_{nm}^2 + B_{nm}^2] \right]^{\frac{1}{2}} P_n(\cos\psi_{s,t})$$

is unbiassed.

Indeed C(0) would not be a good estimate of the variance of the process. Perhaps if C(s,t) were multiplied by R(0), something resembling a covariance function would be obtained, and this would give the same results for prediction as if C(s,t) were used. But it is best to regard this kernel as an interpolation kernel rather than a covariance function. It is however, not an arbitrary kernel for it makes use of known values of gravity, albeit in a somewhat indirect way. The coefficients c_n could be computed from Lauritzen's expression or any other expression which fits the empirical covariance function well, simply by multiplying by (2n+1) and taking the square root. It would be preferable to obtain a closed expression for this kernel but in view of the fact that at this stage it is but a tentative suggestion, and

that it is beyond the scope of this project to perform the necessary numerical work to see what results are obtained, this will not be done.

Since the kernel C(s,t) is regarded as an interpolation kernel rather than a covariance function, the prediction obtained using it is more in the nature of a deterministic approximation solution. The basic problem with such solutions is that it is somewhat difficult to deal with observational noise. To see why this is so the concept of Geodetic Least Squares Collocation will be examined.

2.10 Observation Noise and Trend Removal: Geodetic Least Squares Collocation.

The definitive work on this topic is (MORITZ, 1972) although (MORITZ, 1980) includes all the material therein plus extensions. Consequently it is well known and will only be briefly dealt with for illustrative purposes. (DERMANIS, 1977) also gives a geometrical approach to the method.

Here only gravity prediction will be dealt with, but it must be remembered that it is possible to predict any linear functional of the earth's gravity field by using cross covariances in much the same manner as in Chapter 1 where linear operators were applied to kernels. For further information see (MORITZ, 1980), (DERMANIS, 1977).

Any observation of gravity may be said to be comprised of normal gravity, ie. the gravity value determined from a mathematical model, plus a signal caused by geoidal undulation or divergence from the chosen model plus observational error or noise. If gravity is measured at m points on the earth's surface this may be represented by the vector equation

$$\underset{\sim}{L} = \underset{\sim}{\mu}_{\gamma} + \underset{\sim}{s} + \underset{\sim}{n}$$

ie. Observed gravity = "True" normal gravity + signal + noise.

The signal and noise in the above equations are m×1 random vectors with zero expectation and given covariance matrices $E(\underset{\sim}{s}\underset{\sim}{s}^T) = C_{ss}$, $E(\underset{\sim}{n}\underset{\sim}{n}^T) = C_{nn}$. Signal and noise are assumed to be independent, ie. $E(\underset{\sim}{n}\underset{\sim}{s}^T) = \underset{\sim}{0}_{m\times m}$.

Now it is possible to parameterise the normal gravity such that

$$\underset{\sim}{\gamma} = A\underset{\sim}{X} \; ; \quad E(\underset{\sim}{\gamma}) = AE(\underset{\sim}{X}) = A\underset{\sim}{\mu}_X = \underset{\sim}{\mu}_\gamma$$

where $\underset{\sim}{\gamma}_{m\times 1}$ and $\underset{\sim}{X}_{q\times 1}$ are random vectors and A is an m×q matrix of constants. Some linearisation is necessary for this.

The model thus becomes

$$\underset{\sim}{L} = A\underset{\sim}{X} + \underset{\sim}{s} + \underset{\sim}{n}$$

If m observations of gravity are taken on the earth's surface it is possible therefore to estimate $\underset{\sim}{X}$ by least squares.

Let $\underset{\sim}{s} + \underset{\sim}{n} = \underset{\sim}{Z}$

$$E(\underset{\sim}{Z}) = \underset{\sim}{0} \qquad E(\underset{\sim}{Z}\underset{\sim}{Z}^T) = C_{ss} + C_{nn} = \overline{C}$$

Thus $\underset{\sim}{Z} = -A\underset{\sim}{X} + \underset{\sim}{L}$, and minimising $\underset{\sim}{Z}^T\overline{C}^{-1}\underset{\sim}{Z}$ the estimator of $\underset{\sim}{X}$ is obtained as

$$\hat{\underset{\sim}{X}} = (A^T\overline{C}^{-1}A)^{-1}A^T\overline{C}^{-1}\underset{\sim}{L}$$

Also, $\hat{\underset{\sim}{Z}} = \underset{\sim}{L} - A\hat{\underset{\sim}{X}}$

If the observations were assumed to be error free, ie. $C_{nn} = 0_{m\times m}$ then appropriate adjustments could be made giving

$$\hat{\underset{\sim}{X}} = (A^TC_{ss}^{-1}A)^{-1}A^TC_{ss}^{-1}\underset{\sim}{L}$$

$$\hat{\underset{\sim}{s}} = \underset{\sim}{L} - A\hat{\underset{\sim}{X}}$$

The use of $\hat{\underset{\sim}{s}}$ would give gravity anomalies where indeed $E(\hat{\underset{\sim}{s}}) = 0$ rather than having to make the assumption that $E(\underset{\sim}{s}) = 0$ as would be necessary if some arbitrary model were used to compute $\underset{\sim}{\gamma}$.

In this way trends would be removed in a much more satisfactory manner than determining the average gravity anomaly from an arbitrary normal gravity model, taking the average and subtracting it from each gravity anomaly value. It is also better than using an arbitrary model and assuming it is correct, ie. that

$$\underset{\sim}{\gamma} = \underset{\sim}{\mu}_{\gamma} \quad \text{ie.} \quad E(\underset{\sim}{s}) = 0$$

The $\hat{\underset{\sim}{s}}$ values could then be used in linear least squares prediction as previously described.

However it is not wished to make the assumption that there is no observational error at this point, and it also desired to predict the value of the signal at some point P which has not been previously observed. Let the signal s at point P be denoted by d, and form the vector

$$\underset{\sim}{v} = [d \; Z_1 \; Z_2 \; \cdots \; Z_m]^T = [d \; \vdots \; \underset{\sim}{Z}^T]^T$$

Let $B = [0_{m\times 1} \; \vdots \; I_{m\times m}]$ where $I_{m\times m}$ is an identity matrix.

Form the model

$$A\underset{\sim}{X} + B\underset{\sim}{V} - \underset{\sim}{L} = 0$$

It is to be noted that

$$A\underset{\sim}{X} + B\underset{\sim}{V} - \underset{\sim}{L} = A\underset{\sim}{X} + 0_{m\times 1}\,\underset{\sim}{p}_{1\times 1} + I_{m\times m}\,\underset{\sim}{Z}_{m\times 1} - \underset{\sim}{L}_{m\times 1}$$

$$= A\underset{\sim}{X} + \underset{\sim}{s} + \underset{\sim}{n} - \underset{\sim}{L} = \underset{\sim}{0}$$

ie. $\underset{\sim}{L} = A\underset{\sim}{X} + \underset{\sim}{s} + \underset{\sim}{n}$ as before, so it will be expected that $\hat{\underset{\sim}{X}}$ when estimated from the new model will be the same as previously est

To estimate $\hat{\underset{\sim}{X}}$ one first places

$$Q = \left[\begin{array}{c|c} C_{dd} & C_{ds} \\ \hline C_{sd} & \overline{C} \end{array}\right] = E[\underset{\sim\sim}{V}\underset{\sim\sim}{V}^T]$$

where $C_{dd} = E(\underset{\sim}{d}\underset{\sim}{d}^T)$ is a 1×1 matrix

$C_{ds} = E[\underset{\sim}{d}\underset{\sim}{Z}^T] = E[\underset{\sim}{d}(\underset{\sim}{s}^T+\underset{\sim}{n}^T)]_{1\times m}$

$= E[\underset{\sim}{d}\underset{\sim}{s}^T]$, for signal and noise are independent,

$C_{sd} = C_{ds}^T$ is $m\times1$

and $\overline{C}_{m\times m}$ is, as before, $E[\underset{\sim}{Z}\underset{\sim}{Z}^T]$.

Then the function $\underset{\sim}{V}^T Q^{-1} \underset{\sim}{V}$ is minimised giving the standard estimators for this type of least squares problem:

$$\hat{\underset{\sim}{X}} = (A^T(BQB^T)^{-1}A)^{-1} A^T(BQB^T)^{-1} \underset{\sim}{L}$$

$$\hat{\underset{\sim}{V}} = QB^T(BQB^T)^{-1} (\underset{\sim}{L} - A\hat{\underset{\sim}{X}})$$

Substituting for B and Q ,

$$\hat{\underset{\sim}{X}} = (A^T\overline{C}^{-1}A)^{-1} A^T\overline{C}^{-1}\underset{\sim}{L}$$

$$V = \begin{bmatrix} \hat{\underset{\sim}{d}} \\ \hat{\underset{\sim}{Z}} \end{bmatrix} = \begin{bmatrix} C_{ds}\,\overline{C}^{-1}(L-A\hat{X}) \\ \underset{\sim}{L} - A\hat{\underset{\sim}{X}} \end{bmatrix} .$$

So the same results are obtained for $\hat{\underset{\sim}{X}}$ and $\hat{\underset{\sim}{Z}}$ as before.

Now $\hat{d} = C_{ds}\overline{C}^{-1}(\underset{\sim}{L} - A\hat{\underset{\sim}{X}})$ and it is interesting to note that for error free observations, $C_{nn} = 0_{m\times m}$ and

$$\hat{\underset{\sim}{s}} = \underset{\sim}{L} - A\hat{\underset{\sim}{X}}$$

$$\overline{C}^{-1} = (C_{ss} + C_{nn})^{-1} = C_{ss}^{-1}$$

$$\hat{d} = C_{ds}\, C_{ss}^{-1}\, \hat{\underset{\sim}{S}} .$$

In the notation of the section on linear least squares prediction $\hat{\underset{\sim}{s}}$ corresponds to $\underset{\sim}{x}^{\cdot}$, C_{ss} to G and C_{ds} to $\underset{\sim}{p}$, so

$$\hat{d} = \underset{\sim}{p}^T G^{-1} \underset{\sim}{x}^{\cdot} .$$

So the model reduces to the linear least squares prediction model in this case, which is identical to the deterministic model when the reproducing kernel is the covariance function.

However it is to be noted that the model where noise was included was derived from the point of view that the signal was random. That is to say $\underset{\sim}{V}^T Q^{-1} \underset{\sim}{V}$ was minimised and Q was a covariance matrix. So if observational noise is to be included it appears than an interpolation kernel cannot be used for C_{ss}. A least squares estimate is only obtained when C_{ss} is the covariance matrix of the signal. If something else is used an optimal solution is not obtained. It is easy to see why. The diagonal elements of C_{ss} are the variances K(0) of the signal process. If no noise is included it does not matter if the interpolation kernel C(s,t) at s = t, ie. C(0), is equal to MK(0) where K(0) is the true covariance function, for the factor M cancels out in the prediction formula. But if one were to use $(MC_{ss} + C_{nn})^{-1}$ instead of $(C_{ss} + C_{nn})^{-1}$ in the case where observational noise is taken into account, the cancellation would not occur and the balance between signal and noise would be drastically altered. If M were small, this could lead to very strange results, for the impression gained would be one of a process similar to a white noise process. For large M the effect would be similar to not including observational error at all. Now it may be argued that this would not matter particularly since the accuracy of a single gravity observation is such that its variance is minute compared to the variations in the signal. It is in fact true that the precision of gravity observations is [illegible] but it must be remembered that if single gravity observations were used for prediction, the storage and computational problems in determining global

estimates would be immense. Usually what is done is to try to obtain a value which is representative of the gravity anomaly in a particular area, say a $5^{\circ} \times 5^{\circ}$ or $1^{\circ} \times 1^{\circ}$ square. Whereas the precision of the observations made to obtain the value is very good it may not be possible to say the same of the accuracy or "error of representation" of the value finally chosen to represent an area. In addition, whereas gravity observations are not going to differ greatly in precision at various points on the earth's surface, there may be quite large variations of the accuracies of anomalies used to represent $1^{\circ} \times 1^{\circ}$ squares, simply because the density of gravity coverage varies greatly over the earth's surface.

So there are certain disadvantages in using non-probabilistic kernels for prediction. The problems may not be insurmountable but should not be simply ignored. As stated in (DERMANIS, 1977) it is necessary to extend the deterministic approach either by introducing bounds for observational errors or by means of some combination of probabilistic and deterministic concepts, to remove this inherent defect of the deterministic approach.

CONCLUSION

It has been shown in this project that it is possible to take two entirely different points of view in deriving suitable equations for prediction of linear functionals of the earth's anomalous potential field. In Chapter 1 the equations were derived by various mean weighted residuals methods, particularly least squares, Galerkin's Method and collocation. These were shown to yield identical results if a reproducing kernel Hilbert Space were chosen and n kernel functions were used as ψ_i in the test function

$$u^* = c_1\psi_1 + c_2\psi_2 + \ldots c_n\psi_n.$$

It is quite valid to use such functions for suitable choices of reproducing kernels, namely those kernels K(s,t) where the function $K^s(t)$ are solutions of Laplace's equation. The general form of such kernels has been derived, with the isotropic kernel shown as a special case.

In Chapter 2 the problem has been approached from a statistical or probabilistic point of view. Here, no reference at all has been made to the solution of Laplace's equation. The observed gravity anomalies have been assumed to be values obtained at various points on a single realisation of a second order random field on a sphere. The method of least squares linear prediction has been used to derive equations using the projection theorem on a Hilbert Space of random functions. This space is a reproducing kernel Hilbert Space, but the kernel is not chosen with a view to satisfying Laplace's equation. It is chosen as the covariance function of the random field. This is a positive definite function and if the random field is homogeneous may be expressed in the form

$$C(\psi) = \sum_{n=0}^{\infty} g_n P_n(\cos\psi)$$

where $\sum_{n=0}^{\infty} g_n < \infty$. (OBUKHOV, 1947)

In addition, the Moore-Aronszajn-Loève theorem (theorem 1.6.3) shows that there exists a Hilbert Space unique up to isomorphism with $C(\psi)$ as its reproducing kernel, which is called H(C). By use of the Basic Congruence Theorem it is shown that the results obtained using H(C) in a deterministic sense are identical to results obtained using linear least squares prediction.

But $C(\psi)$ is of the same form as isotropic kernels giving rise to a solution of Laplace's equation, so the statistical method solves Laplace's equation with the given boundary conditions.

It is therefore not just some method which has been arbitrarily chosen merely on statistical grounds, although it appears to be when only the statistical treatment in terms of matrices is given - especially when the empirical covariance function is used. It is a special case of the solution of Laplace's equation by approximation with insufficient boundary data. Now all kernels of the above form (and of course of the more general anisotropic form) give such an approximation. But a particular choice of the coefficients in the expansion, such that the true covariance function is used as the kernel, gives the solution attractive statistical properties and may allow the computation of confidence bands with appropriate probability statements. For example the error bound shown in section 1.8 would, due to the congruence, be a variance in the statistical method. This aspect, however, has not been further amplified in this project because of the difficulty or impossibility of obtaining the true covariance function. The problems of estimation of such a function discussed in section 2.9 appear to preclude the eventual attainment of even a consistent estimator, no matter how many observations of gravity are made in the future.

(DERMANIS, 1977) section 3.2.4 shows that the choice of the true covariance function leads to an unbiassed estimator which has minimum variance among all linear estimators, so from the probabilistic point of view, the choice of reproducing kernel should approximate the true covariance function as closely as possible.

Nevertheless, it is interesting to consider the following: suppose the true covariance function were known, and that the gravity anomalies were known at all points on the earth. Now suppose that the record of all but n "saved" values had been misplaced. If an infinite array of earths were taken, each of which had anomaly values at the n locations identical to the "saved" values, and the mean value of the anomalies at all unknown points were computed, these values should equal the values predicted using the true covariance function in conjunction with the n "saved" values.

Now imagine that the lost record had become found again, and the missing anomalies were available. The predicted values would differ from the original record, albeit optimally. But a question may be posed: Is it possible to choose coefficients in a kernel function such that when it is used with the prediction formula and the n "saved" values it yields smaller discrepancies (in mean square, perhaps) from the actual record than those yielded by the true covariance function?

Put another way, the stochastic nature of the problem has more to do with the fact that the unknown values have not been measured than to some variation between samples.

The problem then, is to predict what has occurred in a <u>single sample,</u> rather than what happens <u>on the average.</u>

Let us suppose that it is possible to obtain a function, albeit with foreknowledge of all values, which is at least competitive with the true covariance function. The next step would be to obtain an estimator which may be used to estimate this function from a limited record, in such a way that if information were increased it would approach the function.

An attempt was made to do something of this nature in section 1.8, where the mean square error of prediction was examined with the intention of choosing coefficients in the kernel function to minimise prediction error for a particular limited record. Unfortunately it led to a rather uncomfortable set of equations and more thought will be required if this avenue is to be adequately examined.

Notwithstanding this, a result was obtained which, although it by no means minimises the mean square error of prediction, at least has the property that if the kernel is bounded then so is the norm of the function. Here the error bound of the predicted value is not infinite, a property not shared with the empirical covariance function. The expression for the kernel so obtained is related to the empirical covariance function so it is not difficult to enumerate.

It should not be thought of as an estimator of the true covariance function but more as an interpolation kernel.

The use of such deterministic kernels has the disadvantage that there does not appear to be a way of adequately combining them with the effects of observation noise, an important aspect if mean anomalies are being used to represent large areas away from the computation point.

Some remarks should be made on anisotropy and local prediction. Firstly it must be emphasised that the method works for an anisotropic

kernel, and that nowhere has isotropy been assumed in the derivation of the prediction equations by either the approximation or statistical approach. The difficulty with anisotropic kernels is that they are extremely difficult to estimate from a single record in the global sense.

The problem is similar to that faced in a non-stationary time series. Here, covariance values are estimated by taking a mean square of values sampled at two points a constant distance apart over the entire record. If the sample is non-stationary this cannot be done because covariance ceases to be independent of the starting point. So it is necessary to remove all trends, linear, periodic etc. by some filter and work with what remains, which can then be assumed stationary. Something similar would need to be done for a random field on the sphere rather than attempting to obtain an anisotropic kernel. This has been briefly discussed in section 2.10. As is the case in time series analysis the method of achieving this trend removal would be basically trial and error, but more difficult. In time series it largely amounts to repeated examination of the correlogram, after each attempt to remove trend, until it looks like that of a stationary time series. There are systematic methods of doing this: Box-Jenkins for example. However, when the index set is a sphere difficulties arise. On a plane it would be possible to use a two dimensional covariance function, such as the one in (KEARSLEY, 1977), and draw the covariance function as a contour map. Repeated attempts to remove trend could be judged by the closeness of the contour lines to circles. Perhaps the sphere or part of it could be projected onto a plane in such a way that distortions in shape did not occur and the same procedure followed.

It may also be possible to remove trend iteratively: First use an isotropic kernel, with normal gravity to remove trend, and determine a geoid. Then use this geoid model to determine "normal" gravity, subtracting the values so obtained from the original gravity values to obtain new anomalies. Then determine the empirical covariance function using these values. If the function so obtained does not resemble one which could be attributed to white noise, repeat the procedure until such a function is obtained. The penultimate covariance function is then isotropic. But this method would tend to depend on the original kernel used, and may not converge.

As far as applying collocation in a local sense one needs to be careful. It is either possible to make the assumption that globally the field is isotropic and homogeneous, or it is not. If it is not, some attempt should be made to remove trends globally until the residual field is isotropic. For example this is done where gravity anomaly rather than gravity itself is used: the north-south trend due to the ellipsoidal shape of the earth being removed. In any case, the field used in global prediction should be isotropic if only because of the difficulty in estimation of an anisotropic global covariance function.

But no matter how sophisticated the trend removal is globally, it will always be possible to find areas on the earth's surface where anisotropy seems to prevail. Such areas could be "detected" using a two-dimensional covariance function. One may then ask whether a special anisotropic covariance function should be used locally for interpolation in such an area, possibly in conjunction with the global covariance function for observations in areas remote from the computation point.

To this, surely the answer is no, for the following reasons:

1. Such an interpolation is likely to be on a local plane using a two dimensional covariance function. If so,

the kernel can no longer be regarded as one whose functions satisfy Laplace's equation.

2. Statistically it is usually an absurdly small sample for covariance function estimation. It is similar having a fifty year record of electrical impulses measured hourly which exhibit stationarity, then selecting from this record a period of a week, say, where some small trend appears (a trend probably balanced by subsequent weeks) and using this small sample to compute the covariance function.
3. If the field is globally homogeneous, how can it be inhomogeneous in a number of locations?
4. The computation of the anisotropic two dimensional covariance function carries an inherent assumption of stationarity, at least in two directions.
5. If such a function is not used in conjunction with the global covariance function for points outside the area this is tantamount to an implication that values outside a small area do not contribut to N, (although this is not so serious when deflections of the vertical are being considered). If the function is used in conjunction with the global one, the computations become quite involved.
6. If the global homogeneous covariance function is used, the peculiarities of the particular area are reflected in the $\underset{\sim}{p}$ and $\underset{\sim}{f}$ vectors in the prediction formula, in any case.

When one adds to the above objections the problems of even obtaining a good estimate of the global covariance function using all available data, problems associated with non-ergodicity and non-converge

of the norm, which will be greatly exagerated for smaller samples, one must conclude that the method should only be applied globally. The problems associated with global estimation are simply not avoided by using local interpolation and prediction.

In fact the only inherent difficulty in the problem is lack of knowledge of the true covariance function, the rest of the theory being quite sound. But then all that is lost is the possibility of making probability statements about confidence intervals or bands, for as long as the kernel chosen is of the form derived in section 1.6.3 the solution obtained is always an approximate solution of Laplace's equation with the relevant boundary condition. One point to note is that even if the true covariance function were known, it would still be necessary to know or assume something about the distribution of the random function before very efficient probability statements could be made. So really all that is sacrificed is the best linear unbiassed property of the estimator and compatibility with the random noise of observations.

Finally it should be noted that the methods examined in this project are not the only ones which may be applied. Other interpolation methods include the use of spline functions as in (BHATTACHARYYA, 1969), (PRETLOVA, 1976), (SUNKEL, 1975), the use of piecewise quadratic smooth functions (GERSTL, HEINDL & REINHART, 1979), multiquadric harmonic function interpolation (HARDY, 1971, 1979), (HARDY & GÖPFERT, 1975) and the methods used in Geostatistics such as Kriging and interpolation using the Variogram (MATHERON, 1963), (MONGET & ALBUISSON, 1971).

Mean weighted residual methods are by no means the only way to approximately solve partial differential equations, either Other approaches include finite-element methods and methods using the calculus

of variations (DAHLQUIST & BJÖRK, 1974), (FINLAYSON, 1972). Indeed even the least squares principle could be varied and the minimax principles of Tchebychev Approximation (DAVIS, 1963) applied. Bayesian statistical methods could also be examined.

It would be instructive to compare such methods theoretically if possible. For example (FINLAYSON, 1972) relates variational principles to the Galerkin and finite-element methods. In cases where such relationships cannot be found there would be great value in a comparison of numerical results.

In this way a fuller appreciation and understanding of the approximate solution of the Stakesian geodetic boundary value problem would be obtained. Similar sentiments apply to modern non-Stokesian methods where integral rather than differential equations must be solved.

ACKNOWLEDGEMENTS

The author is indebted to Dr. P.J. Cooke for his valuable advice and constant encouragement during the preparation of this dissertation.

The valuable assistance of Mrs. B. Golding in the typing and presentation of this project is greatly appreciated.

In addition, the part that the late Dr. R.S. Mather played in initially attracting the author's attention to this area of research, and his encouragement, is also greatfully acknowledged.

BIBLIOGRAPHY

1. ARONSZAJN, N. (1950), *The Theory of Reproducing Kernels*, Trans. Amer. Math. Soc., Vol 68, pp.337-404.

2. BERNSTEIN, U. (1975), *High-Degree Geopotential Coefficient Behaviour From Local Gravity Statistics*, J. Geophysical Research, Vol 80, No.20.

3. BHATTACHARYYA, B.K. (1969), *Bicubic Spline Interpolation as a Method for Treatment of Potential Field Data*, Geophysics Vol 34 No.3, pp.402.

4. BJERHAMMAR, A. (1969), *Theory of a new Geoid*, Bull. Geod. No.92, pp.173-203.

5. BJERHAMMAR, A. (1974), *Discrete Approaches to the Solution of the Boundary Value Problem in Physical Geodesy*, International School of Geodesy, Eriche, 1974.

6. BRENNECKE, J., GROTEN, E., & HEIN, G. (1976), *Some Problems of Digital Representation of Geodetic Functions and of Integral Transforms*, Manuscripta Geodaetica, Vol 1, pp.143-171.

7. COLLATZ, L. (1966), *The Numerical Treatment of Differential Equations*, Berlin, Springer-Verlag, 568 pp.

8. CRAMER, H. & LEADBETTER, M.R. (1967), *Stationary and Related Stochastic Processes*, New York, John Wiley.

9. CURTAIN, R.F. & PRITCHARD, A.J. (1977), *Functional Analysis in Modern Applied Mathematics*, London, Academic, 339 pp.

10. DAHLQUIST, G. & BJÖRCK, A. (1974), trans. ANDERSON, N., *Numerical Methods*, Englewood Cliffs, N.J. Prentice-Hall 573 pp

11. DAVENPORT, W.B. & ROOT, W.L. (1958), *Random Signals and Noise*, New York, McGraw-Hill, 393 pp.

12. DAVIS, P.J. (1963), *Interpolation and Approximation*, New York, Blaisdell, 393 pp.

13. DE GRAAFF-HUNTER, J. (1935), *The Figure of the Earth from Gravity Observations and the Precision Obtainable*, Phil. Trans. Roy. Soc. London, Ser.A, Vol 234 pp.377-431.

14. DERMANIS, A. (1976), *Probabilistic and Deterministic Aspects of Linear Estimation in Geodesy*, Ohio State University, Reports of the Department of Geodetic Science, No.244, 167 pp.

15. DERMANIS, A. (1977) *Geodetic Linear Estimation Techniques and the Norm Choice Problem*, Manuscripta Geodaetica, Vol 2, pp.15-97.

16. DOOB, J.L. (1953), *Stochastic Processes*, New York, John Wiley 654 p.

17. FINLAYSON, B.A. (1972), *The Method of Weighted Residuals and Variational Principles*, New York, Academic, 412 pp.

18. FORSBERG, R. & TSCHERNING, C.C. *The Use of Height Data in Gravity Field Approximation by Collocation*, (Preprint, 1980)

19. FORSBERG, R. & TSCHERNING, C.C. (1981), *Deflection and Gravity Anomaly Prediction for Inertial Position using Collocation*, Second International Symposium on Inertial Technology for Surveying and Geodesy, Banff, June 1981.

20. FREUND, J.E. (1962), *Mathematical Statistics*, Englewood Cliffs N.J., Prentice-Hall, 390 pp.

21. GELFAND, I.M. & VILENKIN, N. Ya. (1964), *Generalised Functions*, Vol 4, New York, Academic, 384 pp.

22. GERSTL, M., HEINDL, G. & REINHART, E. (1979), *Interpolation and Approximation by Piecewise Quadratic Smooth Functions of two Variables*, XVIIth I.U.G.G. General Assembly, Canberra, December 1979.

23. GIKHMAN, I.I. & SKOROKHOD, A.V. (1969), *Introduction to the Theory of Random Processes*, Philadelphia, W.B. Saunders, 516 pp.

24. GRAFAREND, E.W. (1976), *Geodetic Applications of Stochastic Processes*, Physics of the Earth and Planetary Interiors, 12, pp.151-179.

25. GRAYBILL, F.A. (1961), *An Introduction to Linear Statistical Models*, Vol 1, New York, McGraw-Hill, 463 pp.

26. GRENANDER, U. (1963), *Probabilities on Algebraic Structures*, New York, John Wiley, 218 pp.

27. GRENANDER, U. & ROSENBLATT, M. (1957) *Statistical Analysis of Stationary Time Series*, New York, John Wiley, 300 pp.

28. GROTEN, E. (1969), *On the Covariance Function of the Terrain Correction*, Symposium on Physical Geodesy, Prague, 1969.

29. GROTEN, E. (1972), *Some Numerical Aspects of Linear Regression Prediction*, Martin Hobine Symposium on Mathematical Geodesy, Florence, 1972.

30. GROTEN, E. (1975), *Some Problems of Large-Scale Gravity Interpretation*, J. Geophys., 41, pp.659-679.

31. HARDY, R. (1971), *Multiquadric Equations of Topography and Other Irregular Surfaces*, J. Geophysical Research, Vol 76, No.8, pp.1905-1915.

32. HARDY, R. & GÖPFERT, W.M. (1975), *Least Squares Prediction of Gravity Anomalies, Geoidal Undulations, and Deflections of the Vertical with Multiquadric Harmonic Functions*, XVIth I.U.G.G. General Assembly, Grenoble, August 1975.

33. HARDY, R. (1979) *Multiquadric Representation of the Earth's Gravitational Field*, XVIIth I.U.G.G. General Assembly, Canberra, December 1979.

34. HEATHCOTE, C.R. (1971), *Probability: Elements of the Mathematical Theory*, London, George Allen & Unwin, 267 pp.

35. HEISKANEN, W.A. & MORITZ, H. (1967), *Physical Geodesy*, San Francisco, W.H. Freeman, 364 pp.

36. HENRIKSON, P. & NASH, R.A. (1970), *A Statistical Analysis of Errors in Gravimetrically Computed Vertical Deflections*, J. Geophysical Research, Vol 75, No.20, pp.4017-4028.

37. HIRVONEN, R.A. (1956) *On the Precision of the Gravimentric Determination of the Geoid*, Trans. Amer. Geoph. Union, Vol 37, No.1, pp.1-8.

38. HIRVONEN, R.A. (1962), *On the Statistical Analysis of Gravity Anomalies*, Ann. Acad. Scient. Fenn., Ser A, III Geol.-Geog., No.58, pp.1-24.

39. HOEL, P. (1954), *Introduction to Mathematical Statistics*, New York, John Wiley

40. HOGG, R.V. & CRAIG, A.T. (1965), *Introduction to Mathematical Statistics*, Toronto, Collier-MacMillan, 383 pp.

41. JEFFREYS, H. (1941), *The Determination of the Earth's Gravitational Field*, M.N.R.A.S. Geophys. Suppl. Vol 5, No.1, pp.1-22.

42. JEFFREYS, H. (1943), *The Determination of the Earth's Gravitational Field (Second Paper)*, M.N.R.A.S. Geophys. Suppl. Vol 5, No.3, pp.55-66.

43. JONES, R.H. (1963), *Stochastic Processes on a Sphere as Applied to Meteorological 500-Millibar Forecasts*, in *"Procs. of Symposium on Time Series Analysis"*, Brown University June 1963, Rosenblatt, M., ed., New York, John Wiley, pp.119-124.

44. KAULA, W.M. (1957), *Accuracy of Gravimetrically Computed Deflections of the Vertical*, Trans. Amer. Geoph. Union, Vol 38, No.3, pp.297-305.

45. KAULA, W.M. (1959), *Statistical and Harmonic Analysis of Gravity*, J. Geophys. Res., Vol 64, No.12, pp.2401-2421.

46. KAULA, W.M. (1963), *Determination of the Earth's Gravity Field*, Revs. Geophys., Vol 1, No.4, pp.507-551.

47. KAULA, W.M. (1966), *Tests and Combination of Satellite Determinations of the Gravity Field with Gravimetry*, J. Geophs. Res. Vol 71, p.5503.

48. KAULA, W.M. (1967), *Theory of Statistical Analysis of Data Distributed over a Sphere*, Revs. Geophys., Vol 5, No.1, pp.83-107.

49. KEARSLEY, W. (1977), *Non-Stationary Estimation in Gravity Prediction Problems*, Sc.Rep. No.1, Air Force Geophysics Lab. U.S.A.F., Hanscom A.F.B., Massachusetts.

50. KINGMAN, J.F.C. & TAYLOR, S.J. (1966), *Introduction to Measure and Probability*, Cambridge, University Press, 401 pp.

51 KRARUP, T. (1969), *A Contribution to the Mathematical Foundation of Physical Geodesy*, Danish Geodetic Inst., Meddelelse No.44, Copenhagen, 80 pp.

52. LACHAPELLE, G. (1975), *Astrogravimetric Levelling by Least Squares Collocation*, XVIth I.U.G.G. General Assembly, Grenoble, August 1975.

53. LACHAPELLE, G. (1977), *Estimation of Disturbing Potential Components using a Combined Integral Formulae and Collocation Approach*, Second International Summer School in the Mountains, Ramsau, August 1977.

54. LAURITZEN, S.L. (1973), *The Probabilistic Background of some Statistical Methods in Physical Geodesy*, Danish Geodetic Inst., Meddelelse No.48, Copenhagen, 96 pp.

55. LAURITZEN, S.L. (1975) *Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth*, Stoch. Proc. Appls., Vol 3, pp.65-72.

56. LONSETH, A.T. (1977), *Sources and Applications of Integral Equations*, Siam Review, Vol 19, No.2, pp.241-278.

57. LOWREY, B. & ARGENTIERO, P. (1977), *On Estimating Gravity Anomalies - A Comparison of Least Squares Collocation with Conventional Least Squares Techniques*, Bull. Geod., 51, pp.119-12

58. MACMILLAN, W.D. (1930), *The Theory of the Potential*, New York, Dover, 467 pp.

59. MATHER, R.S. (1967), *The Extension of the Gravity Field in South Australia*, Uniciv Report No.R-19, University of N.S.W. Kensington, N.S.W. Australia, 26 pp.

60. MATHER, R.S., LERCH, F.J., RIZOS, C., MASTERS, E.G. & HIRSCH, B. (1978), *Determination of Some Dominant Parameters of the Global Dynamic Sea Surface Topography from GEOS-3 Altimetry*, N.A.S.A. Technical Memorandum 79558, Goddard Space Flight Centre, Greenbelt, Maryland, U.S.A. 40 pp.

61. MATHERON, G. (1963), *Principles of Geostatistics*, Economic Geology, Vol 58, pp.1246-1266.

62. MEISSL, P. (1971), *A Study of Covariance Functions Related to the Earth's Disturbing Potential*, Ohio State University, Department of Geodetic Science, Report No.151, 88 pp.

63. MEISSL, P. (1976), *Hilbert Spaces and Their Application to Geodetic Least Squares Problems*, Boll. Geod. Scien. Affini, No.1, 1976, pp.49-80.

64. MONGET, J.M. & ALBUISSON, M. (1971), *A New Statistical Treatment of Gravity Data*, Bull. Geod., No.102, pp.451-466.

65. MORITZ, H. (1962), *Interpolation and Prediction of Gravity and their Accuracy*, Ohio State University, Dept. of Geodetic Science, Report No.24, 69 pp.

66. MORITZ, H. (1964), *Accuracy of Mean Gravity Anomalies obtained from Point and Profile Measurements*, Ann. Acad. Scient. Fenn., Ser A, III Geol.-Geog. No.75, pp.1-16.

67. MORITZ, H. (1966), *An Extension of Error Theory with Application to Spherical Harmonies*, Bull. Geod. No.81, pp.225-234.

68. MORITZ, H. (1972), *Advanced Least-Squares Methods*, Ohio State University, Dept. of Geodetic Science, Report No.175, 129 pp.

69. MORITZ, H. (1973), *The Method of Least-Squares Collocation in Geometrical Geodesy*, International Symposium on Computational Methods in Geodesy, Oxford, September 1973.

70. MORITZ, H. (1976), *Integral Formulas and Collocation*, Manuscripta Geodetica, Vol 1, pp.1-40.

71. MORITZ, H. (1977) *Least Squares Collocation and the Gravitational Inverse Problem*, J. Geophys., 43, pp.153-162.

72. MORITZ, H. (1978), *Statistical Foundations of Collocation*, Ohio State University, Department of Geodetic Science, Report No.272, 75 pp.

73. MORTIZ, H. (1980), *Advanced Physical Geodesy*, Karlsruhe, Wichmann Verlag, 500 pp.

74. MORTIZ, H. & SUNKEL, H. eds,/(1978) *Approximation Methods in Geodesy*, Karlsruhe, Wichmann Verlag, 284 pp.

75. OBUKHOV, A.M. (1947), *Statistically Homogeneous Random Fields on a Sphere*, Uspekhi. Mat. Nauk., Vol 2, No.2, pp.196-198 (in Russian; English translation available from School of Surveying Collection of Papers, University of N.S.W., Kensington).

76. OBENSON, G. (1972), *Comparison of Prediction Accuracies of* $5^{o} \times 5^{o}$ *Mean Gravity Anomalies from Two Methods*, Bull. Geod., No.106, pp.425-429.

77. OBENSON, G. (1973), *Error Analysis of Deflections of the Vertical and Undulations from the Accuracies of Gravity Anomalies*, Bull. Geod., No.108, pp.141-156.

78. PAPOULIS, A. (1965) *Probability, Random Variables and Stochastic Processes*, New York, McGraw-Hill, 583 pp.

79. PARZEN, E. (1959), *Statistical Inference on Time Series by Hilbert Space Methods, I*, Technical Report No.23, Applied Mathematics and Statistics Laboratory, Stanford University, California, 131 pp.

80. PARZEN, E. (1960), *An Approach to Time Series Analysis*, The Annals of Mathematical Statistics, Vol.32, No.4, pp.951-989.

81. PARZEN, E. (1961), *Regression Analysis of Continuous Parameter Time Series*, from "Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability", Vol.1, Berkeley, University of California Press, pp.469-489.

82. PARZEN, E. (1962), *Extraction and Detection Problems and Reproducing Kernel Hilbert Spaces*, J.SIAM Control, Ser. A, Vol 1, No.1, pp.35-62.

83. PARZEN, E. (1963,a), *Probability Density Functionals and Reproducing Kernel Hilbert Spaces*, from: "Time Series Analysis", Rosenblatt, M. ed., New York, John Wiley, pp.155-169.

84. PARZEN, E. (1963,b), *A New Approach to the Synthesis of Optimal Smoothing and Prediction Systems*, from: "Mathematical Optimization Techniques", University of California Press, Berkeley, pp.75-108.

85. PELLINEN, L.P. (1970), *Estimation and Application of Degree Variances of Gravity*, Studia Geoph. et. Geod., No.14, pp.168-173.

86. PRETLOVA, V. (1976), *Bicubic Spline Smoothing of Two-Dimensional Geophysical Data*, Studia Geoph. et Geod. No.20, pp.168-179.

87. RAPP, R.H. (1964,a) *Statistical Analysis of Gravity Anomalies and Elevations by Long Profiles and by Areas*, Ann. Acad. Scient. Fenn., Ser.A, III Geol.-Geog., No.79, pp.1-11.

88. RAPP, R.H. (1964,b), *The Prediction of Point and Mean Gravity Anomalies through the Use of a Digital Computer*, Ohio State University, Dept. of Geodetic Science, Report No.43, 177 pp.

89. RAPP, R.H. (1972), *Results from a Combination of Satellite and Gravimetric Data Considering the Collocation Concept*, V Symposium on Mathematical Geodesy, Florence, October 1972, 25 pp.

90. RAPP, R.H. (1973), *Improved Models for Potential Coefficients and Anomaly Degree Variances*, J. Geophys. Res., Vol.78, No.17, pp.3497-3500.

91. RUMMEL, R. (1976), *A Model Comparison in Least Squares Collocation*, Bull. Geod., 50, pp.181-192.

92. RUMMEL, R., HAJELA, D.P., & RAPP, R.H. (1976), *Recovery of Mean Gravity Anomalies from Satelite -Satelite Range Rate Data Using Least Squares Collocation*, Ohio State University, Department of Geodetic Science, Report No.248, 61 pp.

93. SCHWARZ, K.P. (1976), *Least Squares Collocation for Large Systems*, Boll. Geod. Scien. Affini, No.3, pp.311-324.

94. SCHWARZ, K.P., (1977), *On the Nonhomogeneity of the Global Covariance Function*, Bull. Geod., 51, pp.93-103.

95. SCHWARZ, K.P. & KRYNSKI, J. (1977), *Improvement of the Geoid in Local Areas by Satellite Gradiometry*, Bull. Geod., 51, pp.163-176.

96. STAKGOLD, I. (1967), *Boundary Vajue Problems of Mathematical Physics*, Vol 1, New York, MacMillan, 340 pp.

97. SUNKEL, H. (1975), *Reconstruction of Functions from Discrete Mean Values Using Cubic Spline-Functions*, XVIth I.U.G.G. General Assembly, Grenoble, August 1975.

98. TSCHERNING, C.C. (1970), *Determination of the Geoid by Collocation*, VIth Nordic Meeting of Geodisists, Helsinki, May 1970.

99. TSCHERNING, C.C. (1971), *Collocation Methods in Harmonic Spaces*, Meeting on Free Boundary Value Problems, Bonn 1971.

100. TSCHERNING, C.C. (1972,a), *An Algol-Program for Prediction of Height Anomalies, Gravity Anomalies and Deflections of the Vertical*, Danish Geodetic Institute, Internal Report No.2, 40 pp.

101. TSCHERNING, C.C. (1972,b), *Representation of Covariance Functions Related to the Anomalous Potential of the Earth using Reproducing Kernels*, Danish Geodetic Institute, Internal Report No.3, 63 pp.

102. TSCHERNING, C.C. (1972,c), *On the Relation Between the Variation of the Degree-Variances and the Variation of the Anomolous Potential*, Fifth Symposium on Mathematical Geodesy.

103. TSCHERNING, C.C. (1973), *Application of Collocation. Determination of a Local Approximation to the Anomalous Potential of the Earth, using "Exact" Astrogravimetric Collocation*, published in: Brosowski, B. and Martensen, E., eds. "Methoden und Verfahren der mathematischen Physik/ Bond 14", Bibliographisches Institut, Manheim.

104. TSCHERNING, C.C. (1974), *A Fortran IV Program for the Determinati of the Anomalous Potential using Stepwise Least Squares Collocation*, Ohio State University, Department of Geodetic Science, Report No.212, 125 pp.

105. TSCHERNING, C.C. (1976,a), *A Mass Density Covariance Function Consistent with the Covariance Function of the Anomalous Potential*, Boll. Geod. Scien. Affini, N.2, 1976, pp.161-172.

106. TSCHERNING, C.C. (1976,b), *Determination of Datum-Shift Parameters using Least Squares Collocation*, Boll. Geod. Scien. Affini, N.2 1976, pp.173-183.

107. TSCHERNING, C.C. (1976,c), *Covariance Expressions for Second and Lower Order Derivatives of the Anomalous Potential*, Ohio State University, Department of Geodetic Science, Report No.225.

108. TSCHERNING, C.C. (1977), *A Note on the Choice of Norm when using Collocation for the Computation of Approximations to the Anomalous Potential*, Bull. Geod., 51, pp.137-147.

109. TSCHERNING, C.C. (1978,a), *Introduction to Functional Analysis with a view to its Application in Approximation Theory*, from "Approximation Methods in Geodesy", Moritz, H. & Sunkel, H. eds., Karlsruhe, Wichmann Verlag, pp.157-191.

110. TSCHERNING, C.C. (1978,b), *On the Convergence of Least Squares Collocation*, Boll. Geod. Scien. Affini, N.2-3, 1978, pp.509-516.

111. TSCHERNING, C.C. (1978, c), *Collocation and Least Squares Methods as a tool for handling Gravity Field Dependent Data obtained through Space Research Techniques*, Bull. Geod., 52, pp.199-212.

112. TSCHERNING, C.C. (1978,d), *A Users Guide to Geopotential Approximation by Stepwise Collocation on the RC400 - Computer*, Danish Geodetic Institute, Meddelelse No.53, Copenhagen, 61 pp.

113 TSCHERNING, C.C. (1979,a), *Gravity Prediction using Collocation and taking Known Mass Density Anomalies into Account*, Geophys. J.R. Astr. Soc., 59, pp.147-153.

114. TSCHERNING, C.C. (1979,b), *Comparison of Some Methods for the Detailed Representation of the Earth's Gravity Field*, XVIIth I.U.G.G. General Assembly, Canberra, December 1979.

115. TSCHERNING, C.C. (1980), *The Role and Computation of Gravity for the Processing of Levelling Data*, Second International Symposium on Problems Related to the Redefinition of North American Vertical Geodetic Networks, Canada 1980.

116. TSCHERNING, C.C. & FORSBERG, R. (1981), *Geoid-determinations in the Norwegian-Greenland Sea. An Assessment of Recent Results*, Fourth Discussion Meeting on the Blue Road Geotraverse, Berlin Free University, February 1981.

117. TSCHERNING, C.C. & RAPP, R.H. (1974), *Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations, and Deflections of the Vertical implied by Degree Variance Models*, Ohio State University, Department of Geodetic Science, Report No.208, 89 pp.

118. UOTILA, U.A. (1960), *Investigations on the Gravity Field and Shape of the Earth*, Ann. Acad. Scient. Fenn., Ser.A, III.Geol.-Geog., No.55, 92 pp.

119. UOTILA, U.A. (1962), *Harmonic Analysis of World-Wide Gravity Material*, Ann. Acad. Scient. Fenn., Ser.A, III. Geol., No.67, 17 pp.

120. VYSKOCIL, V. (1969), *Statistical Characteristics of Linear Transformations of Gravity Anomalies*, Studia Geoph. et Geod., 13 (1969), pp.252-258.

121. WONG, E. (1971), *Stochastic Processes in Information and Dynamical Systems*, New York, McGraw-Hill, 308 pp.

122. YAGLOM, A.M. (1962), *An Introduction to the Theory of Stationary Random Functions*, trans. R.A. Silverman, Englewood-Cliffs, Prentice-Hall, 235 pp.

123. ZAANEN, A.C. (1953), *Linear Analysis*, Amsterdam, North Holland, 600 pp.

Publications from
THE SCHOOL OF SURVEYING, THE UNIVERSITY OF NEW SOUTH WALES
P.O. Box 1, Kensington, N.S.W. 2033
AUSTRALIA

Reports

1.* G.G. Bennett, "The discrimination of radio time signals in Australia", Uniciv Rep. D1, 88 pp. (G 1)

2.* J.S. Allman, "A comparator for the accurate measurement of differential barometric pressure", Uniciv Rep. D3, 9 pp. (G 2)

3. R.S. Mather, "The establishment of geodetic gravity networks in South Australia", Uniciv Rep. R-17, 26 pp. (G 3)

4.* R.S. Mather, "The extension of the gravity field in South Australia", Uniciv Rep. R-19, 26 pp. (G 4)

5.* J.S. Allman, "An analysis of the reliability of barometric elevations", Unisurv Rep. 5, 335 pp. (S 1)

6.* R.S. Mather, "The free air geoid for South Australia and its relation to the equipotential surfaces of the earth's gravitational field", Unisurv Rep. 6, 491 pp. (S 2)

7.* P.V. Angus-Leppan, (Editor), "Control for mapping" (Proceedings of Conference, May 1967), Unisurv Rep. 7, 329 pp. (G 5)

8.* G.G. Bennett & J.G. Freislich, "The teaching of field astronomy", Unisurv Rep. 8, 30 pp. (G 6)

9.* J.C. Trinder, "Photogrammetric pointing accuracy as a function of properties of the visual image", Unisurv Rep. 9, 64 pp. (G 7)

10.* P.V. Angus-Leppan, "An experimental determination of refraction over an icefield", Unisurv Rep. 10, 23 pp. (G 8)

11.* R.S. Mather, "The non-regularised geoid and its relation to the telluroid and regularised geoids", Unisurv Rep. 11, 49 pp. (G 9)

12.* G.G. Bennett, "The least squares adjustment of gyro-theodolite observations", Unisurv Rep. 12, 53pp. (G 10)

13.* R.S. Mather, "The free air geoid for Australia from gravity data available in 1968", Unisurv Rep. 13, 38 pp. (G 11)

* OUT OF PRINT

14.* R.S. Mather, "Verification of geoidal solutions by the adjustment of control networks using geocentric Cartesian co-ordinate systems", Unisurv Rep. 14, 42 pp. (G 12)

15.* G.G. Bennett, "New methods of observation with the Wild GAKI gyro-theodolite", Unisurv Rep. 15, 68 pp. (G 13)

16.* G.G. Bennett, "Theoretical and practical study of a gyroscopic attachment for a theodolite", Unisurv Rep. 16, 343 pp. (S 3)

17.* J.C. Trinder, "Accuracy of monocular pointing to blurred photogrammetric signals", Unisurv Rep. 17, 231 pp. (S 4)

18. A. Stolz, "The computation of three dimensional Cartesian co-ordinates of terrestrial networks by the use of local astronomic vector systems", Unisurv Rep. 18, 47 pp. (G 14)

19. R.S. Mather, "The Australian geodetic datum in earth space", Unisurv Rep. 19, 130 pp. (G 15)

20.* J.G. Fryer, "The effect of the geoid on the Australian geodetic network", Unisurv Rep. 20, 221 pp. (S 5)

21.* G.F. Toft, "The registration and cadastral survey of native- held rural land in the Territory of Papua and New Guinea", Unisurv Rep. 21, 441 pp. (S 6)

22. R.S. Mather et al, "Communications from Australia to Section V, International Association of Geodesy, XV General Assembly, International Union of Geodesy and Geophysics, Moscow 1971", Unisurv Rep. 22, 72 pp. (G 16)

23.* A.H. Campbell, "The dynamics of temperature in surveying steel and invar measuring bands", Unisurv Rep. S 7, 195 pp.

24. A. Stolz, "Three-D Cartesian co-ordinates of part of the Australian geodetic network by the use of local astronomic vector systems", Unisurv Rep. S 8, 182 pp.

25. Papers by R.S. Mather, H.L. Mitchell & A. Stolz on the following topics:- Four-dimensional geodesy, Network adjustment and Sea surface topography, Unisurv G 17, 73 pp.

26. Papers by L. Berlin, G.J.F. Holden, P.V. Angus-Leppan, H.L. Mitchell & A.H. Campbell on the following topics:- Photogrammetry co-ordinate systems for surveying integration, Geopotential networks and Linear measurement, Unisurv G 18, 80 pp.

27. R.S. Mather, P.V. Angus-Leppan, A. Stolz & I. Lloyd, "Aspects of four-dimensional geodesy", Unisurv G 19, 100 pp.

28. H.L. Mitchell, "Relations between MSL & geodetic levelling in Australia", Unisurv Rep. S 9, 264 pp.

29. A.J. Robinson, "Study of zero error & ground swing of the model MRA101 tellurometer", Unisurv Rep. S 10, 200 pp.

30. Papers by J.S. Allman, R.C. Lister, J.C. Trinder & R.S. Mather on the following topics:- Network adjustments, Photogrammetry, and 4-Dimensional geodesy, Unisurv G 20, 133 pp.

31. G.J.F. Holden, "An evaluation of orthophotography in an integrated mapping system", Unisurv Rep. S 12, 232 pp.

32. G.J. Hoar, "The analysis precision and optimization of control surveys", Unisurv Rep. S 13, 200 pp.

33. Papers by E. Grafarend, R.S. Mather & P.V. Angus-Leppan on the following topics:- Mathematical geodesy, Coastal geodesy and Refraction, Unisurv G 21, 100 pp.

34. Papers by R.S. Mather, J.R. Gilliland, F.K. Brunner, J.C. Trinder, K. Bretreger & G. Halsey on the following topics:- Gravity, Levelling, Refraction, ERTS imagery, Tidal effects on satellite orbits and Photogrammetry, Unisurv G 22, 96 pp.

35. Papers by R.S. Mather, E.G. Anderson, C. Rizos, K. Bretreger, K. Leppert, B.V. Hamon & P.V. Angus-Leppan on the following topics:- Earth tides, Sea surface topography, Atmospheric effects in physical geodesy, Mean sea level and Systematic errors in levelling, Unisurv G 23, 96 pp.

36. Papers by R.C. Patterson, R.S. Mather, R. Coleman, O.L. Colombo, J.C. Trinder, S.U. Nasca, T.L. Duyet & K. Bretreger on the following topics:- Adjustment theory, Sea surface topography determinations, Applications of LANDSAT imagery, Ocean loading of Earth tides, Physical geodesy, Photogrammetry and Oceanographic applications of satellites, Unisurv G 24.

37. E.G. Anderson, "The Effect of Topography on Solutions of Stokes' Problem", Unisurv Rep. S 14, 252 pp.

38. A.H.W. Kearsley, "The Computation of Deflections of the Vertical from Gravity Anomalies", Unisurv Rep. S 15, 181 pp.

39. Papers by S.M. Nakiboglu, B. Ducarme, P. Melchior, R.S. Mather, B.C. Barlow, C. Rizos, B. Hirsch, K. Bretreger, F.K. Brunner & P.V. Angus-Leppan on the following topics:- Hydrostatic equilibrium figures of the Earth, Earth tides, Gravity anomaly data banks for Australia, Recovery of tidal signals from satellite altimetry, Meteorological parameters for modelling terrestrial refraction and Crustal motion studies in Australia, Unisurv G 25.

40. Papers by R.S. Mather, E.G. Masters, R. Coleman, C. Rizos, B. Hirsch, C.S. Fraser, F.K. Brunner, P.V. Angus-Leppan, A.J. McCarthy & C. Wardrop on the following topics:- Four-dimensional geodesy, GEOS-3 altimetry data analysis, Analysis of meteorological measurements for microwave EDM and Meteorological data logging system for geodetic refraction research, Unisurv G 26, 113 pp.

41. Papers by F.K. Brunner, C.S. Fraser, S.U. Nasca, J.C. Trinder, L. Berlin, R.S. Mather, O.L. Colombo & P.V. Angus-Leppan on the following topics:- Micrometeorology in geodetic refraction, LANDSAT imagery in topographic mapping, Adjustment of large systems, GEOS-3 data analysis, Kernel functions and EDM reductions over sea, Unisurv G 27, 101 pp.

42. K. Bretreger, "Earth Tide Effects on Geodetic Observations", Unisurv S 16, 173 pp.

43. Papers by S.M. Nakiboglu, H.L. Mitchell, K. Bretreger, T.A. Herring, J.M. Rueger, K.R. Bullock, R.S. Mather, B.C. Forster, I.P. Williamson & T.S. Morrison on the following topics:- Variations in gravity, Oceanographic and geodetic levelling, Ocean loading effects on Earth tides, Deflections of the vertical, Frequencies of EDM instruments, Land information systems, Sea surface topography, Accuracy of Aerial Triangulation and Questionnaire to Surveyors, Unisurv G 28, 124 pp.

44. Papers by F.L. Clarke, R.S. Mather, D.R. Larden & J.R. Gilliland on the following topics:- Three dimensional network adjustment incorporating ξ, η and N, Geoid determinations with satellite altimetry, Geodynamic information from secular gravity changes and Height and free-air anomaly correlation, Unisurv G 29, 87 pp.

From June 1979 Unisurv G's name was changed to Australian Journal of Geodesy, Photogrammetry and Surveying (Aust.J.Geod.Photo.Surv.).

45. Aust.J.Geod.Photo.Surv. No. 30, (June,1979), 127 pp.:
Bretreger, "Ocean tide models from GEOS-3 altimetry";
Trinder & Smith, "Rectification of LANDSAT data";
Nakiboglu & Lim, "Numerical test of initial value method";
Herring, "The accuracy of deflections of vertical";
Angus-Leppan, "Radiation effects on temperature of metal tape";
Covell, "Errors of short range distance meters", and
Nasca, "Contour lines in engineering".

46.* Aust.J.Geod.Photo.Surv. No. 31, (December 1979) 177 pp.
(contributions to the XVII General Meeting of the IUGG, Canberra, 3-15 December, 1979):
Allman et al., "Readjustment of Australian geodetic survey";
Angus-Leppan, "Ratio method and meteorology" and
"Refraction in levelling",
Berlin, "Adjustment of continental networks",
Brunner, "Limiting factor to geodetic precision";
Coleman et al., "Sea surface slope along NE Australia",
Kahar and Kearsley, "Jawa geoid from altimetry and gravity",
Kearsley & van Gysen, "Outer zone effects on the Australian geoid",
Masters et al., "Tide models from GEOS-3 altimetry", and
Mather & Rizos, "Global mean sea level from altimetry".

47. C. Rizos, "The role of the gravity field in sea surface topography studies", Unisurv S 17 (February,1980), 299 pp.

48. Aust.J.Geod.Photo.Surv. No. 32, (June,1980) 121 pp.:
van Gysen, "Gravimetric deflections of the vertical",
Fraser, "Self calibration of non-metric camera",
Rizos, "Ocean circulation from SST studies",
Trinder, "Film granularity and visual performance".

49. Aust.J.Geod.Photo.Surv. No. 33, (December,1980) 85 pp.:
Burford, "Controlling geodetic networks";
Masters & Stolz, "Crustal motion from LAGEOS";
Fraser, "Variance analysis of adjustments", and
Brunner et al., "Incremental strain near Palmdale".

50. B.C. Forster, "Some measures of urban residual quality from LANDSAT multi-spectral data", Unisurv S 18, (January,1981) 223 pp.

51. Aust.J.Geod.Photo.Surv. No. 34, (June,1981) 94 pp.:
Welsch, "Variances of geodetic observations";
Gilliland, "Outer zones effects on geoid", and
Khalid, "Models of vertical refraction".

52. Richard Coleman, "A Geodetic Basis for recovering Ocean Dynamic Information from Satellite Altimetry", Unisurv S 19, (October,1981) 332 pp.

53. Aust.J.Geod.Photo.Surv. No.35 (December 1981):
Kahar, "Geoid in Indonesia",
Morgan, "Crustal Motion in Papua New Guinea",
Masters et al, "LAGEOS Range Filter",
Stolz et al, "Baseline Values from LAGEOS",
Bishop, "Digital Elevation Models".

54. Aust.J.Geod.Photo.Surv. No.36 (June 1982)
Nakiboglu and Torenberg, "Earth's Gravity and Surface Loading"
Nakiboglu, "Changes in Sea Level and Gravity"
Villanueva, "Geodetic Boundary Value Problem"
Stolz and Harvey, "Australian Geodetic VLBI Experiment"
Gilliland, "Free-air Geoid for South Australia"
Allman, "Geoid for S.E. Asia and S.W. Pacific"
Angus-Leppan, "Gravity Measurements in Levelling"

55. Douglas R. Larden, "Monitoring the Earth's Rotation by Lunar Laser Ranging", Unisurv Report S 20, 280 pp.

56. R. Patterson, "Approximation and Statistical Methods in Physical Geodesy", Unisurv Report S 21, 179 pp.

* Out of print

Proceedings

P.V. Angus-Leppan (Editor), "Proceedings of conference on refraction effects in geodesy & electronic distance measurement", 264 pp.
Price: $10.00

R.S. Mather & P.V. Angus-Leppan (Eds.), "Australian Academy of Science/International Association of Geodesy Symposium on Earth's Gravitational Field & Secular Variations in Position", 740 pp.
Price: $20.00

Monographs

1. R.S. Mather, "The theory and geodetic use of some common projections", (2nd edition), 125 pp. Price: $ 5.00

2. R.S. Mather, "The analysis of the earth's gravity field", 172 pp.
Price: $ 5.00

3. G.G. Bennett, "Tables for prediction of daylight stars", 24 pp.
Price: $ 2.50

4. G.G. Bennett, J.G. Freislich & M. Maughan, "Star prediction tables for the fixing of position", 200 pp. Price: $ 8.00

5. M. Maughan, "Survey computations", 98 pp.
Price: $ 5.00

6. M. Maughan, "Adjustment of Observations by Least Squares", 61 pp.
Price: $ 4.00

7. J.M. Rueger, "Introduction to Electronic Distance Measurement", (2nd Edition), 140 pp. Price: $ 10.00

Note: Prices do not include postage.

OTHER PRICES for 1982 (Surface Mail Postage Inclusive)

1. **Aust.J.Geod.Photo.Surv.** (formerly Unisurv G), 2 issues annually of approximately 100 pages each issue.

   To Libraries $18
   To Individuals $12

2. **Special Series** (Unisurv S)
   Research reports of 200 to 300 pages, published annually (on average).

   To Libraries $25 per copy
   To Individuals $18 per copy

To order, write to Publications Officer, School of Surveying, U.N.S.W.