

Reports of the Department of Geodetic Science

Report No. 175

ADVANCED LEAST-SQUARES METHODS

by

Helmut Moritz

The Ohio State University
Research Foundation
Columbus, Ohio 43212

Contract No. F19628-72-C-0120
Project No. 8607
Task No. 860701
Work Unit No. 86070101

Scientific Report No. 1

June, 1972

Contract Monitor: Bela Szabo
Terrestrial Sciences Laboratory

Approved for public release; distribution unlimited.

Prepared for

Air Force Cambridge Research Laboratories
Air Force Systems Command
United States Air Force
Bedford, Massachusetts 01730

FOREWORD

This report was prepared by Helmut Moritz, Professor, Technische Hochschule Graz, and Adjunct Professor, Department of Geodetic Science of The Ohio State University, under Air Force Contract No. F19628-72-C-0120, OSURF Project No. 3363 A1 , Project Supervisor, Urho A. Uotila, Professor, Department of Geodetic Science. The contract covering this research is administered by the Air Force Cambridge Research Laboratories, Air Force Systems Command, Laurence G. Hanscom Field, Bedford, Massachusetts, with Mr. Owen W. Williams and Mr. Bela Szabo, Project Scientists.

ABSTRACT

A general least-squares method (collocation) which encompasses, as special cases, least-squares adjustment and least-squares prediction, is presented in detail and applied to various problems occurring in geodesy and photogrammetry, such as interpolation and coordinate transformation.

In particular, this method permits an optimal simultaneous determination of geodetic positions and of the terrestrial gravity field by combining different data of any kind--terrestrial angle, distance and gravity measurements as well as data from advanced satellite techniques. To provide an adequate statistical background, an alternative statistical interpretation of the anomalous gravity field in terms of covariance analysis of individual functions is given, and its relation to the usual interpretation as a stochastic process on the sphere is discussed.

ADVANCED LEAST-SQUARES METHODS

CONTENTS

1. Introduction	1
2. Least-squares collocation	7
3. Accuracy	24
4. Application to interpolation	41
5. Application to transformation problems	51
6. Application to physical geodesy	67
7. Covariances	82
8. The statistical background	100
9. Measuring errors; stochastic process interpretation	114
10. The norm of the anomalous gravity field.	121
References	126

1. INTRODUCTION

Least-squares adjustment is familiar to every geodesist. Besides this classical technique, least-squares methods for interpolation and prediction of gravity have found their way into geodesy several years ago (Kaula, 1963; Moritz, 1962, 1963). Later on such least-squares prediction methods were applied to deflections of the vertical (Heitz, 1969; Grafarend, 1971 a,b).

After some preliminary investigations on the relation between least-squares adjustment and least-squares prediction (Kaula, 1963; Moritz, 1965; Wolf, 1969), it was Krarup (1968, 1969) who succeeded in completely clarifying the mathematical structure of this relation by exhibiting least-squares prediction as least-squares adjustment in a Hilbert space with a kernel function. Furthermore, he vastly generalized least-squares prediction of gravity to obtain a general least-squares theory (least-squares collocation) for estimating any element of the terrestrial gravity field. Subsequently this theory, in a somewhat simplified form, was applied to the solution of a number of problems in physical geodesy (Moritz, 1970 a,b).

A somewhat different line of development, already foreshadowed in (Kaula, 1963), leads to the incorporation of parameters representing systematic effects. For the case of gravity this was done in (Moritz, 1969); an improved deduction was given in (Moritz, 1970 c), where also applications to other fields of geodesy are outlined. This line of development finally leads to a joint least-squares determination of geometric position and of the gravity field (Krarup, 1971; Moritz, 1970 d).

Similar techniques are known in the theory of stochastic processes (Parzen, 1961).

The present report attempts a systematic and fairly comprehensive elementary presentation of the theory and its geodetic applications.

The following simple considerations are intended to give the reader some intuitive feeling for the problems involved.

The conventional model underlying least-squares adjustment by parameters is

$$x = AX + n , \quad (1-1)$$

where x , the "measurement", is the vector of the observations (usually denoted by ℓ), X is the vector of the parameters or unknowns, and n , the "noise", is the vector of the measuring errors (usually denoted by $-v$); A is a known rectangular matrix. It is evident that x is split up into a systematic part, AX , and a random part, n .

In the sequel, the number of observations will consistently be denoted by q , and the number of parameters, by m ; to get an overdetermined problem, we must have $m < q$. Then x and n are (column) vectors of q components, X is a (column) vector of m components, and A is a $q \times m$ matrix.

The present report considers the following generalization:

$$x = AX + s + n , \quad (1-2)$$

obtained by admitting, in addition to the noise n , a second random quantity s , which we shall call the "signal".

This is the model underlying least-squares collocation including systematic parameters.

Thus the measurement x consists of a systematic part, AX , and two random parts, s and n . Usually the systematic part will be nonlinear originally, and the linear form AX is obtained on linearization by Taylor's theorem.

The signal s may exist at points other than the measuring points; in particular, it may vary continuously even if x is measured at discrete points only. This is why we can use such a mathematical model for purposes of interpolation.

In fact, if the vector x consists of measurements of the same kind, then we may consider our problem as a generalized problem of interpolation, as Fig. 1-1 indicates.

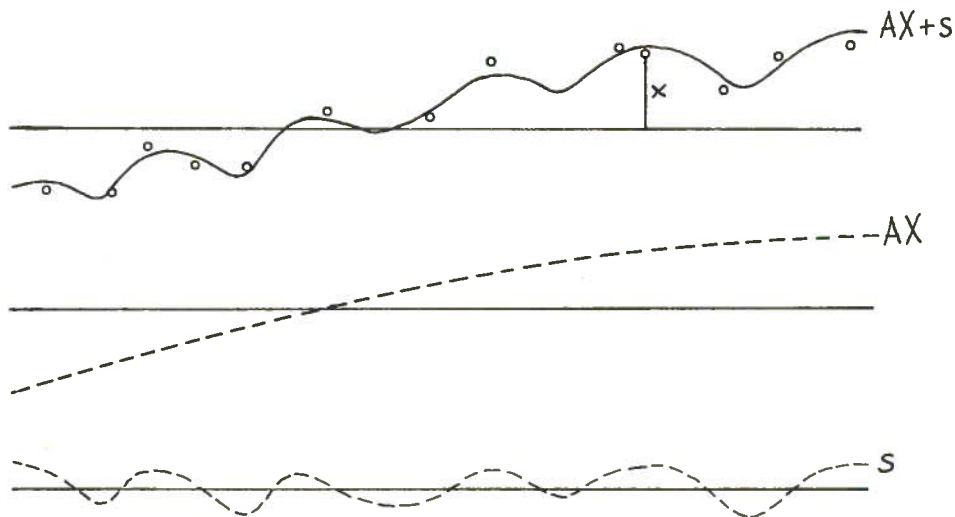


Figure 1-1

We have to determine the curve shown on top (full line) by means of discrete observations (small circles), which are furthermore affected by observational errors n . The curve to be interpolated consists of a systematic part, AX , and a random part, s , both of which are of importance.

If we consider the determination of the parameter X as adjustment, the removal of the noise as filtering, and the computation of s at points other than the measuring points as prediction, we may say that least-squares collocation combines adjustment, filtering and prediction.

The relevance of this model for geodetic problems is made evident by mentioning some conceivable applications from quite different fields of geodesy.

1. Gravity Measurements. - Here x is the gravimeter reading, s represents the gravity anomaly Δg , n is the random measuring error (including inertial noise in the case of marine and aerial gravimetry), and X represents systematic parameters of two different kinds: (a) the parameters of the normal gravity formula and (b) instrumental constants and other systematic effects on the measurement such as drift.

2. Satellite Observations. - Here x comprises optical or electronical measurements to artificial satellites, AX represents the normal orbit (after linearization with respect to the parameters), s represents gravitational perturbations of the orbit, and n comprises other random effects, in particular measuring errors.

3. Transformations in Geodesy and Photogrammetry. - Consider two overlapping local geodetic coordinate systems. If one system is transformed into the other, there may remain residual discrepancies or distortions which are irregular but strongly correlated. Thus AX represents the transformation formula, s comprises the residual distortions, and n is the effect of measuring errors on position. This is a combined transformation and interpolation problem, which is the two-dimensional analogue to the one-dimensional problem shown in Fig. 1-1.

Transformation problems of precisely the same nature are frequent in photogrammetry.

4. Graduation Errors of Theodolite Circles. - Here x is the circle reading, AX represents the "regular" graduation error, s represents the "irregular" graduation error, and n is the reading error.

In these generalized interpolation problems, the measurements making up the vector x are all of the same kind, so that we have the case of homogeneous measurements. This assumption is not necessary, however: we may also have heterogeneous measurements, as long as the corresponding signal quantities are functionally inter-related.

This is particularly important for applications to physical geodesy. Here we have one basic signal field, the earth's anomalous gravity field. The quantities referring to this field--gravity anomalies, geoidal heights, deflections of the vertical, anomalous second-order gradients, etc.--are all functionally inter-related through the field structure. Thus our present method permits the use of measurements of arbitrary field quantities and their combination. In (Moritz, 1970 a) we have considered many examples of this kind, where systematic effects AX have been assumed to be missing.

The admission of systematic effects greatly enlarges the scope of the theory. In fact, any geodetic measurement may be split up into three components:

1. a systematic part involving, on the one hand, the ellipsoidal reference system and, on the other hand, other parameters and systematic errors;
2. a random part expressing the effect of the anomalous gravity field; and
3. random measuring errors.

Consider, for example, a measured zenith distance z . Then we have (cf. Heiskanen and Moritz, 1967, p. 190), disregarding measuring errors,

$$z = z' - \epsilon. \quad (1-3)$$

Thus z' , the ellipsoidal zenith distance, represents the systematic part due to the ellipsoidal reference system; the quantity ϵ , the deflection of the vertical, represents the effect of the anomalous gravity field as the second component; and the measuring error of z (omitted in (1-3)) gives the third component.

Thus any geodetic measurement--from angular measurements in triangulation to satellite-to-satellite tracking--fits into the scheme (1-2).

The theory described and applied in the present report is developed in such a way as to be formally very similar to the well-known "general case of least-squares adjustment", that is, the adjustment of condition equations containing parameters. After the deduction from a minimum principle of the form $v^T P v = \text{minimum}$, it is shown that the solution possesses optimum accuracy, minimizing the mean square error of the result.

The presentation of the theory given here is itself an attempt to solve a minimum problem: to develop the subject to an extent adequate for geodetic applications, at the same time keeping the mathematical apparatus to a minimum. It has been possible to get along with ordinary linear algebra, bypassing Hilbert space techniques. The corresponding loss in generality and mathematical depth is probably acceptable to geodesists primarily interested in applications;

the more mathematically inclined reader will find the additional study of Krarup's (1969) monograph a fascinating experience. Still, the present elementary treatment may also have some theoretical interest from the point of view of logical economy.

This theory is then applied to various problems in geodesy and photogrammetry. In particular it permits an optimal simultaneous determination of geodetic positions and the terrestrial gravity field using heterogeneous data of any kind.

2. LEAST-SQUARES COLLOCATION

In sec. 1 we have introduced the basic equation of least-squares collocation:

$$x = AX + s' + n , \quad (2-1)$$

where x is the "measurement", s' is the "signal" (formerly denoted by s), and n is the "noise". The quantities x , s' and n are q -vectors (q is the number of observations); the m -vector X comprises the m parameters; and A is a prescribed $q \times m$ matrix. All vectors will be considered as column vectors.

It is convenient to substitute

$$z = s' + n , \quad (2-2)$$

so that (2-1) becomes

$$x = AX + z . \quad (2-3)$$

The vectors s' and n are purely random quantities whose expectation (average or mean value), M , is zero:

$$M\{s'\} = 0, \quad M\{n\} = 0. \quad (2-4)$$

Then, by (2-2),

$$M\{z\} = 0, \quad (2-5)$$

so that

$$z = x - AX \quad (2-6)$$

represents the observations "centered" by subtracting their mean value AX , or the purely random part of the observations.

Let us now assume that we wish to estimate (predict) the signal at an arbitrary number of "computation points" which may be different from the "data" points; cf. sec.1, in particular Fig. 1-1. Denote by p the number of computation points; then the signal vector to be computed will be the p -vector

$$s = \begin{bmatrix} s_1 & s_2 & \dots & s_p \end{bmatrix}^T, \quad (2-7)$$

T denoting the transpose (s is a column vector!). Since we want to denote by s the p -vector (2-7), we had to change the notation for the q -vector representing the signals that correspond to the observations x , denoting this vector by s' .

The vector (2-7) may be combined with the vector

$$z = \begin{bmatrix} z_1 & z_2 & \dots & z_q \end{bmatrix}^T \quad (2-8)$$

to give the vector

$$v = \begin{bmatrix} s_1 & s_2 & \dots & s_p & z_1 & z_2 & \dots & z_q \end{bmatrix}^T = \begin{bmatrix} s^T & z^T \end{bmatrix}^T, \quad (2-9)$$

comprising all $p + q$ random variables that enter into our problem.

The covariance matrix Q of this vector v may be written as a partitioned matrix:

$$Q = \begin{bmatrix} C_{ss} & C_{sx} \\ C_{xs} & C_{xx} \end{bmatrix}. \quad (2-10)$$

Here

$$C_{ss} = \text{cov}(s, s) = M\{ss^T\} \quad (2-11)$$

denotes the covariance matrix of the signal s ,

$$C_{xx} = \text{cov}(x, x) = M\{(x - AX)(x - AX)^T\} = M\{zz^T\} \quad (2-12)$$

denotes the covariance matrix of the measurement x
(note that

$$M\{x\} = AX \quad (2-13)$$

by (2-5) and (2-6)!), and

$$C_{sx} = \text{cov}(s, x) = M\{s(x-AX)^T\} = M\{sz^T\}, \quad (2-14)$$

$$C_{xs} = \text{cov}(x, s) = M\{(x-AX)s^T\} = M\{zs^T\}$$

denote the crosscovariance matrices between s and x ;
obviously

$$C_{xs} = C_{sx}^T, \quad (2-15)$$

T denoting the transpose as usual.

In this connection it is convenient also to
consider the covariance matrix of the noise,

$$C_{nn} = \text{cov}(n, n) = M\{nn^T\}, \quad (2-16)$$

and the covariance matrix of the vector s' entering
in (2-1),

$$C_{s's'} = \text{cov}(s', s') = M\{s's'^T\}, \quad (2-17)$$

and to introduce the abbreviations

$$C_{s's'} = C, \quad (2-18a)$$

$$C_{nn} = D, \quad (2-18b)$$

$$C_{xx} = \bar{C}. \quad (2-18c)$$

Thus \bar{C} is the covariance matrix of the observations,
 D is the covariance matrix of the measuring errors
and C is the covariance matrix of the "signal part"
of x .

Let us now assume that signal and noise are uncorrelated with each other. This assumption is justified if x is the result of a direct measurement, where the size of the signal s' has no influence on the size of its measuring error n . Then

$$\begin{aligned} C_{ns'} &= \text{cov}(n, s') = M\{ns'^T\} = 0, \\ C_{s'n} &= \text{cov}(s', n) = M\{s'n^T\} = 0. \end{aligned} \quad (2-19)$$

Now

$$z = s' + n,$$

so that

$$\begin{aligned} C_{xx} &= M\{zz^T\} = M\{(s' + n)(s'^T + n^T)\} = \\ &= M\{s's'^T + ns'^T + s'n^T + nn^T\} = \\ &= M\{s's'^T\} + M\{ns'^T\} + M\{s'n^T\} + M\{nn^T\} \\ &= C_{s's'} + C_{nn}, \end{aligned}$$

by (2-16), (2-17) and (2-19). With the notations (2-18a, b, c) we thus have

$$\bar{C} = C + D, \quad (2-20)$$

so that the covariance matrix of x is obtained by simply adding the covariance matrices of its signal and noise parts.

Under the same assumption (no correlation between signal and noise) we further have

$$\begin{aligned} C_{sx} &= M\{sz^T\} = M\{s(s'+n)^T\} = M\{ss'^T\} + M\{sn^T\} \\ &= M\{ss'^T\} , \end{aligned}$$

hence, together with (2-15)

$$\begin{aligned} C_{sx} &= \text{cov}(s, s') , \\ C_{xs} &= \text{cov}(s', s) , \end{aligned} \tag{2-21}$$

so that C_{sx} and C_{xs} are pure signal covariances.

All these covariances, and in particular the covariances that make up the matrix (2-10), will be assumed to be known in what follows.

In terms of the vector (2-9) we may write (2-3) as

$$AX + Bv - x = 0 \tag{2-22}$$

with

$$B = \begin{bmatrix} 0 & 0 & . & . & 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & . & . & 0 & 0 & 1 & . & . & 0 \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & . & . & 0 & 0 & 0 & . & . & 1 \end{bmatrix} = \begin{bmatrix} 0 & I \end{bmatrix} , \tag{2-23}$$

where O is the $q \times p$ zero matrix and I is the $q \times q$ unit matrix; the matrix A is the same as in (2-1).

Eq. (2-22) has the form of condition equations with parameters. It is therefore tempting to apply the algorithm of least-squares adjustment for this case, that is,

$$v^T P v = \text{minimum} \quad (2-24)$$

with the side condition (2-22), where

$$P = Q^{-1} \quad (2-25)$$

is known by (2-10). The vector v plays the role of "residuals", and the vector X represents the "unknowns".

The solution is well known; cf. (Gotthardt, 1968, pp. 238-241) or (Wolf, 1968, pp. 132-134). We minimize the function (k represents the vector of correlates or Lagrange multipliers)

$$\phi = \frac{1}{2} v^T P v - k^T (A X + B v - x) \quad (2-26)$$

by forming $\partial \phi / \partial v = 0$, $\partial \phi / \partial X = 0$ ¹⁾, which gives the equations

$$v^T P - k^T B = 0, \quad k^T A = 0. \quad (2-27)$$

¹⁾ These conditions are only necessary but not sufficient; that the solution obtained in this way really gives a minimum will be shown at the end of this section.

The result of the first equation,

$$v^T = k^T B P^{-1}$$

or

$$v = P^{-1} B^T k, \quad (2-28)$$

is substituted into (2-22); together with the transposed second equation this gives

$$B P^{-1} B^T k + A X = x, \quad A^T k = 0; \quad (2-29)$$

these are $q+m$ linear equations for the $q+m$ unknowns k and X .

To find X , we multiply the first equation of (2-29) by $A^T (B P^{-1} B^T)^{-1}$ and subtract the second to obtain

$$A^T (B P^{-1} B^T)^{-1} A X = A^T (B P^{-1} B^T)^{-1} x. \quad (2-30)$$

Eq. (2-30) determines X ; then v follows from (2-28) together with the first equation of (2-29) as

$$v = P^{-1} B^T (B P^{-1} B^T)^{-1} (x - A X). \quad (2-31)$$

Let us now specialize to our present case, where the matrix B is given by (2-23). We readily find

$$B P^{-1} B^T = B Q B^T = \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} C_{ss} & C_{sx} \\ C_{xs} & C_{xx} \end{bmatrix} \begin{bmatrix} 0 \\ I \end{bmatrix} = C_{xx} = \bar{C}, \quad (2-32)$$

by (2-18c). Thus (2-30) and (2-31) become

$$A^T \bar{C}^{-1} A X = A^T \bar{C}^{-1} x, \quad (2-33)$$

$$v = Q B^T \bar{C}^{-1} (x - A X). \quad (2-34)$$

The first equation determines the parameters:

$$X = (A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1} x; \quad (2-35)$$

the second equation may be partitioned using (2-9), (2-10) and (2-23):

$$\begin{bmatrix} s \\ z \end{bmatrix} = \begin{bmatrix} C_{ss} & C_{sx} \\ C_{xs} & C_{xx} \end{bmatrix} \begin{bmatrix} 0 \\ I \end{bmatrix} \bar{C}^{-1} (x - A X),$$

so that

$$s = C_{sx} \bar{C}^{-1} (x - A X). \quad (2-36)$$

If only one signal s_p is to be predicted, then the vector s will have only one component, s_p , and the matrix C_{sx} will reduce to the line vector

$$[C_{p1} \ C_{p2} \ \dots \ C_{pq}] = C_p^T. \quad (2-37)$$

Thus (2-36) will read

$$s_p = C_p^T \bar{C}^{-1} (x - A X). \quad (2-38)$$

These equations solve our problem: first the estimated values of the parameters X are computed from (2-35), then the estimated (predicted and/or filtered) values of the signal s are obtained from (2-36) or (2-38).

This is the method of least-squares collocation.

To illustrate these general formulas, let us write the occurring matrices explicitly for the case $p = 2$, $q = 3$, $m = 2$. Then

$$s = \begin{bmatrix} s_p \\ s_q \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix},$$

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}, \quad \bar{C} = \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \bar{C}_{13} \\ \bar{C}_{21} & \bar{C}_{22} & \bar{C}_{23} \\ \bar{C}_{31} & \bar{C}_{32} & \bar{C}_{33} \end{bmatrix}, \quad C_{sx} = \begin{bmatrix} C_{p1} & C_{p2} & C_{p3} \\ C_{q1} & C_{q2} & C_{q3} \end{bmatrix};$$

we have written $s_1 = s_p$, $s_2 = s_q$ for convenience (in this way it is, e.g., easy to distinguish the 11-component of the matrix (2-18a), C_{11} , from the corresponding component of the matrix C_{sx} by denoting the latter by C_{p1}).

Then the solution (2-35) may be written in the form

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}^{-1} \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix} \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \bar{C}_{13} \\ \bar{C}_{21} & \bar{C}_{22} & \bar{C}_{23} \\ \bar{C}_{31} & \bar{C}_{32} & \bar{C}_{33} \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (2-39)$$

with the abbreviation

$$\begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = R = A^T \bar{C}^{-1} A ,$$

and (2-36) becomes

$$\begin{bmatrix} s_P \\ s_Q \end{bmatrix} = \begin{bmatrix} C_{P1} & C_{P2} & C_{P3} \\ C_{Q1} & C_{Q2} & C_{Q3} \end{bmatrix} \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \bar{C}_{13} \\ \bar{C}_{21} & \bar{C}_{22} & \bar{C}_{23} \\ \bar{C}_{31} & \bar{C}_{32} & \bar{C}_{33} \end{bmatrix}^{-1} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} , \quad (2-40)$$

where

$$z = x - AX .$$

The alternative expression (2-38) gives

$$s_P = \begin{bmatrix} C_{P1} & C_{P2} & C_{P3} \end{bmatrix} \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \bar{C}_{13} \\ \bar{C}_{21} & \bar{C}_{22} & \bar{C}_{23} \\ \bar{C}_{31} & \bar{C}_{32} & \bar{C}_{33} \end{bmatrix}^{-1} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} , \quad (2-41)$$

which coincides with s_P as expressed by (2-40).

Already at this point we note the full formal analogy of this formula, for $A = 0$, with the well-known formula for least-squares prediction of gravity; cf. (Heiskanen and Moritz, 1967), p. 268, eq. (7-63). Much will have to be said on this analogy later on.

Properties of Least-Squares Collocation. - The solution expressed by (2-35) and (2-36) or (2-38) has the following properties:

- (a) The result is independent of the number p of signal quantities to be computed.
- (b) Both observed and computed quantities may be quite heterogeneous.
- (c) The observations may be errorless or affected by measuring errors (noise).
- (d) The method is invariant with respect to linear transformations of the data or of the results.
- (e) The solution is optimal in the sense that it gives the most accurate results obtainable on the basis of the available data.

As for (a), the vector s may consist of one component as in (2-38), or it may consist of a thousand components. The result for the same signal quantity will be the same in both cases, for the following reason. Eq. (2-35) depends on the observations only; the signal does not enter at all. In (2-36), each component of the vector s is obtained individually, without being influenced by the other components: each component of s is expressed by (2-36) in the form (2-38); compare (2-40) and (2-41).

This also shows that it is not necessary to incorporate in the vector s all signal quantities that correspond to the observations x , making up the vector s' in (2-1): the vector s needs only comprise those signal quantities we wish to compute.

As for (b) the (physical or mathematical) nature of the random quantities s and x is irrelevant. All we have to require is that (2-22) be satisfied, that s and z be purely random (i.e., of zero expectation), and that the covariance matrix (2-10) be known. In fact, all the quantities s and x may be quite heterogeneous and of a different nature. For instance, in applications to the gravity field to be discussed in sec. 6, x_1 may be a measured gravity anomaly, x_2 may be a component of the deflection of the vertical, and x_3 may be a tesseral harmonic obtained from satellite observations, whereas s_p may be a geoidal height to be computed.

The mathematical framework for the treatment of heterogeneous observations is thus formally the same as for homogeneous observations (e.g. problems of interpolation, prediction, filtering), so that in presenting this framework here we have been able to follow closely the presentation in (Moritz, 1970c), which unnecessarily restricted itself to the homogeneous case.

Property (c) is a straightforward consequence of the fact that the noise n does not directly enter into our minimum problem, as the vector v given by (2-9) does not contain it. We may thus have $n \neq 0$ or $n = 0$ without changing anything in our algorithm. This fact also accounts for the structural identity of (2-41) (in which there may be $n \neq 0$) with the formula for (errorless) gravity prediction, as we have pointed out above.

This fact strongly contributes to the theoretical and practical importance of this method because it has the adjustment of measuring errors, as it were, already built in.

It should also be mentioned that these formulas do not presuppose that signal and noise are uncorrelated. If they are correlated, then the matrix \bar{C} is no longer given by the simple expression (2-20), but (2-35) and (2-36) still hold.

Property (d) follows from the corresponding well-known invariance property of least-squares adjustment (cf. Tienstra, 1956, pp. 152 et seq.), since our minimum problem is formally identical to the adjustment of condition equations with parameters.

The proof of property (e) will be given in the following section.

The present method may thus be summarized as follows: take all observations x and all those signal quantities s which we should like to compute, and take existing linear relations (2-1) into account. In this way one obtains a random vector v which is subject to the condition (2-22). Then the application of the usual minimum principle (2-24), the underlying covariances being assumed to be known, gives the desired solution.

Some comments on the relationship between the present method and ordinary least-squares adjustment are in order. Formally, the present mathematical model is adjustment by condition equations with parameters. It differs, however, from an adjustment problem in the strict sense. In adjustment computations, the conditions connect all observations and, therefore, all residuals v :

the principle $v^T P v = \text{minimum}$ contains only those residuals which also occur explicitly in the condition equations.

In our present problem, however, the quantities that are most important here, namely the signal values s , do not enter at all into the condition equations (2-22), although they do enter into the minimum principle (2-24). Our problem thus contains additional random variables, which are related to the observations only indirectly through the joint covariances. Formally this makes no difference: it only means that the matrix A contains some all-zero columns, but conceptually it is quite significant.

To repeat, the essential relation between the observations x and the desired quantities s is through the covariances which are assumed to be known. This is the basic point that distinguishes our present least-squares estimation from ordinary least-squares adjustment; it also accounts, e.g., for the conceptual difference between least-squares gravity prediction and least-squares adjustment, cf. (Moritz, 1965).

Thus, rather than saying that we have reduced our problem of estimating the signal to the adjustment of condition equations with parameters, it is more appropriate to say that we have slightly generalized this adjustment model in such a way as to cover our problem.

Not only conceptually, but also computationally the covariances represent the crucial point in the present method, as we shall see; everything else is extremely simple--to be sure, apart from numerical problems in handling large matrices.

Proof that $v^T P v = \text{minimum}$. - The derivation just given should be supplemented by the proof that we really have a minimum of (2-24); cf. footnote on p.13.

Besides the least-squares values, denoted by v and X and given by (2-34) and (2-35), we consider arbitrary values for the same vectors, to be denoted by \hat{v} and \hat{X} , which also satisfy the condition (2-22):

$$A\hat{X} + B\hat{v} - x = 0 . \quad (2-42)$$

We put

$$\begin{aligned} \hat{v} &= v + v_1 , \\ \hat{X} &= X + X_1 . \end{aligned} \quad (2-43)$$

The vectors v_1 and X_1 satisfy the equation

$$AX_1 + Bv_1 = 0 . \quad (2-44)$$

We then have

$$\begin{aligned} \hat{v}^T P \hat{v} &= (v^T + v_1^T) P (v + v_1) \\ &= v^T P v + v_1^T P v + v^T P v_1 + v_1^T P v_1 . \end{aligned} \quad (2-45)$$

Now

$$v_1^T P v = v_1^T P P^{-1} B^T C^{-1} (x - AX)$$

by (2-34), with $Q = P^{-1}$. Using (2-35) this becomes

$$\begin{aligned}
v_1^T P v &= v_1^T B^T \bar{C}^{-1} \left[I - A (A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1} \right] x \\
&= -x_1^T A^T \bar{C}^{-1} \left[I - A (A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1} \right] x \\
&= -x_1^T A^T \bar{C}^{-1} x + x_1^T A^T \bar{C}^{-1} A (A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1} x \\
&= -x_1^T A^T \bar{C}^{-1} x + x_1^T A^T \bar{C}^{-1} x = 0 ;
\end{aligned}$$

here we have taken into account that

$$v_1^T B^T = -x_1^T A^T$$

by (2-44). Since the matrix P is symmetric, we have

$$v^T P v_1 = v_1^T P v = 0 .$$

Hence (2-45) reduces to

$$\hat{v}^T P \hat{v} = v^T P v + v_1^T P v_1 . \quad (2-46)$$

As the matrix P is positive-definite, we have

$$v_1^T P v_1 \geq 0 , \quad (2-47)$$

the equality corresponding to $v_1 = 0$. Therefore the minimum of (2-46) is

$$\hat{v}^T P \hat{v} = v^T P v$$

for $\hat{v} = v$, where v is the least-squares solution. This completes the proof.

3. ACCURACY

In this section we shall first derive expressions for the standard errors and error covariances of the quantities X and s obtained by an arbitrary linear unbiased estimation, then we shall specialize these expressions for least-squares collocation, and finally we shall show that collocation gives indeed optimum estimates in the sense that in this case the standard errors are the smallest that are possible for any linear estimation method.

Linear Estimates. - Consider any linear estimates for X and s , that is, expressions of the form

$$s = Lx + a, \quad (3-1a)$$

$$X = Gx + b, \quad (3-1b)$$

where L is a $p \times q$ matrix, G is a $m \times q$ matrix, a is a p -vector, and b is a m -vector. The quantities L , G , a , b are assumed to be independent of x , so that (3-1a,b) represent the estimated values of s and X as linear functions of the measurements x . This is the meaning of the term "linear estimate".

These estimates must reasonably satisfy the same relations as the original "true" values. In the sequel we shall always denote true values by an overbar.

Thus (2-3) and (2-2) are written for true values as

$$x = A\bar{X} + \bar{z} , \quad (3-2a)$$

$$\bar{z} = \bar{s}' + \bar{n} , \quad (3-2b)$$

and for their estimates as

$$x = AX + z , \quad (3-3a)$$

$$z = s' + n . \quad (3-3b)$$

Let us now postulate that the estimates (3-1a,b) be unbiased. That is, the average M of these estimates is to be the same as the average of the corresponding true values: we are to have

$$M\{s\} = 0 \quad (3-4)$$

because of (2-4), and

$$M\{X\} = \bar{X} , \quad (3-5)$$

where \bar{X} is the true value of the parameter vector; note that the estimate X is a random quantity because of (3-1b), whereas the true value \bar{X} is not.

Substitute (3-2a) into (3-1a) to obtain

$$s = LA\bar{X} + L\bar{z} + a ,$$

and form the average M . The result is

$$M\{s\} = LA\bar{X} + LM\{\bar{Z}\} + a ,$$

since \bar{X} and a are nonrandom quantities. By (3-4) and (2-5) this reduces to

$$0 = LA\bar{X} + a . \quad (3-6)$$

For the estimation formula (3-1a) to be meaningful, the vector a must be given beforehand and cannot depend on the true value \bar{X} , which is forever unknown and must be considered as arbitrary. Then (3-6) gives

$$LA = 0 , \quad a = 0 . \quad (3-7)$$

In fact, \bar{X} , being arbitrary, may be taken as small as we please, which is incompatible with a nonzero a in (3-6). Thus $a = 0$ and hence $LA\bar{X} = 0$. As \bar{X} is arbitrary, we must have $LA = 0$.

The conditions (3-7) are thus necessary for the estimate (3-1a) to be unbiased; it is obvious that these conditions are also sufficient.

Let us now investigate (3-1b) in a similar way. Substitute (3-2a) into (3-1b) to obtain

$$X = GA\bar{X} + G\bar{Z} + b ,$$

and form the average M , with the result

$$\bar{X} = GA\bar{X} + b$$

or

$$(I - GA)\bar{X} - b = 0 ,$$

where I denotes the unit matrix, Reasoning as before, we obtain the conditions

$$GA = I , \quad b = 0 , \quad (3-8)$$

which are likewise necessary and sufficient.

The unbiased linear estimates are thus

$$s = Lx \quad \text{with} \quad LA = 0 , \quad (3-9a)$$

$$X = Gx \quad \text{with} \quad GA = I . \quad (3-9b)$$

Let us now put

$$L = H(I - AG) , \quad (3-10)$$

for which the condition $LA = 0$ is automatically satisfied:

$$LA = H(I - AG)A = H(A - AGA) = H(A - A) = 0 ,$$

since $GA = I$ by (3-8); it may be shown that all matrices L satisfying $LA = 0$ can be represented in the form (3-10) (the matrix $I - AG$ is a projection matrix projecting onto the subspace $LA = 0$).

Then (3-9a) becomes

$$\begin{aligned} s &= Lx = H(I - AG)x \\ &= H(x - AGx) = H(x - AX) \end{aligned}$$

by (3-9b) .

In this way we finally obtain the general expression for unbiased linear estimates:

$$X = Gx , \quad (3-11)$$

$$s = H(x - AX) , \quad (3-12)$$

where the $p \times q$ matrix H is arbitrary and the $m \times q$ matrix G satisfies the condition

$$GA = I . \quad (3-13)$$

Errors of Estimation. - The individual error of the estimates is defined as the difference: true value minus estimated value. Thus the individual error of X is

$$\epsilon_X = \bar{X} - X = \bar{X} - Gx = \bar{X} - GA\bar{X} - G\bar{Z} = -G\bar{Z}$$

by (3-2a) and (3-13), and the individual error of s is

$$\epsilon_s = \bar{s} - s = \bar{s} - Lx = \bar{s} - LA\bar{X} - L\bar{Z} = \bar{s} - L\bar{Z}$$

because of (3-7). Thus we have

$$\epsilon_X = -G\bar{Z} , \quad (3-14)$$

$$\epsilon_s = \bar{s} - L\bar{Z} = \bar{s} - H(I - AG)\bar{Z} . \quad (3-15)$$

Let us now pass on to the corresponding standard errors and error covariances. We form (T denotes the transpose as usual)

$$\epsilon_X \epsilon_X^T = G \bar{Z} \bar{Z}^T G^T, \quad (3-16)$$

$$\begin{aligned} \epsilon_s \epsilon_s^T &= (\bar{S} - L \bar{Z}) (\bar{S}^T - \bar{Z}^T L^T) = \\ &= \bar{S} \bar{S}^T - L \bar{Z} \bar{S}^T - \bar{S} \bar{Z}^T L^T + L \bar{Z} \bar{Z}^T L^T. \end{aligned} \quad (3-17)$$

By the usual definition of the covariances, the matrices

$$E_{XX} = M\{\epsilon_X \epsilon_X^T\}, \quad E_{ss} = M\{\epsilon_s \epsilon_s^T\} \quad (3-18)$$

are the error covariance matrices of the vectors X and s . Forming the average M of (3-16) and (3-17) and taking into account the definitions (2-11), (2-12) and (2-14), we obtain

$$E_{XX} = G C_{xx} G^T, \quad (3-19)$$

$$E_{ss} = C_{ss} - L C_{xs} - C_{sx} L^T + L C_{xx} L^T. \quad (3-20)$$

These expressions give the error covariance matrices for any linear unbiased estimation. The diagonal terms represent the error variances (the squares of the standard errors) of the estimated quantities; the off-diagonal terms represent the error covariances.

We may also consider the cross-covariance matrices

$$E_{Xs} = M\{\epsilon_X \epsilon_s^T\}, \quad (3-21a)$$

$$E_{sX} = M\{\epsilon_s \epsilon_X^T\} = E_{Xs}^T. \quad (3-21b)$$

By (3-14) and (3-15) we have

$$\epsilon_X \epsilon_s^T = -G\bar{Z}\bar{S}^T + G\bar{Z}\bar{Z}^T L^T,$$

and hence

$$E_{Xs} = -GC_{xs} + GC_{xx}L^T = E_{sX}^T. \quad (3-22)$$

Least-Squares Collocation. - Let us now specialize these general expressions to the case of least-squares collocation. Comparing (3-11) and (3-12) to (2-35) and (2-36) we see that here

$$G = (A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1}, \quad (3-23)$$

$$H = C_{sx} \bar{C}^{-1}. \quad (3-24)$$

The condition (3-13) for an unbiased estimate is clearly satisfied.

By means of (3-23), the expression (3-19) becomes

$$\begin{aligned} E_{XX} &= G\bar{C}G^T = \\ &= (A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1} \bar{C} \bar{C}^{-1} A (A^T \bar{C}^{-1} A)^{-1} \\ &= (A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1} A (A^T \bar{C}^{-1} A)^{-1} \\ &= (A^T \bar{C}^{-1} A)^{-1}. \end{aligned} \quad (3-25)$$

By (3-10) with (3-23) and (3-24) we have

$$L = C_{sx} \bar{C}^{-1} [I - A(A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1}] , \quad (3-26)$$

so that

$$LC_{xs} = C_{sx} \bar{C}^{-1} [I - A(A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1}] C_{xs} . \quad (3-27)$$

In view of (2-15) it is readily seen that

$$C_{sx} L^T = LC_{xs} . \quad (3-28)$$

By (2-18c) and (3-26) we get

$$\begin{aligned} LC_{xx} L^T &= L \bar{C} L^T = \\ &= C_{sx} \bar{C}^{-1} [I - A(A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1}] \cdot \\ &\cdot \bar{C} [I - \bar{C}^{-1} A(A^T \bar{C}^{-1} A)^{-1} A^T] \bar{C}^{-1} C_{xs} \\ &= C_{sx} \bar{C}^{-1} [I - A(A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1}]^2 C_{xs} . \end{aligned}$$

By direct computation one immediately verifies that the matrix expression between brackets is idempotent, that is,

$$[I - A(A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1}]^2 = I - A(A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1} . \quad (3-29)$$

Thus

$$LC_{xx} L^T = C_{sx} \bar{C}^{-1} [I - A(A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1}] C_{xs} , \quad (3-30)$$

so that all three expressions (3-27), (3-28) and (3-30) entering into (3-20) are equal. Therefore, (3-20) reduces to

$$E_{ss} = C_{ss} - LC_{xs} . \quad (3-31)$$

Finally, (3-22) takes the form

$$\begin{aligned} E_{xs} &= -GC_{xs} + GC_{xx}L^T = \\ &= -(A^T\bar{C}^{-1}A)^{-1}A^T\bar{C}^{-1}C_{xs} + \\ &+ (A^T\bar{C}^{-1}A)^{-1}A^TL^T = \\ &= (A^T\bar{C}^{-1}A)^{-1}A^T[-\bar{C}^{-1}C_{xs} + \\ &+ \bar{C}^{-1}C_{xs} - \bar{C}^{-1}A(A^T\bar{C}^{-1}A)^{-1}A^T\bar{C}^{-1}C_{xs}] \\ &= -(A^T\bar{C}^{-1}A)^{-1}A^T\bar{C}^{-1}A(A^T\bar{C}^{-1}A)^{-1}A^T\bar{C}^{-1}C_{xs} \\ &= -(A^T\bar{C}^{-1}A)^{-1}A^T\bar{C}^{-1}C_{xs} . \end{aligned} \quad (3-32)$$

Thus we have

$$E_{xx} = (A^T\bar{C}^{-1}A)^{-1} , \quad (3-33)$$

$$E_{ss} = C_{ss} - C_{sx}\bar{C}^{-1}[I - A(A^T\bar{C}^{-1}A)^{-1}A^T\bar{C}^{-1}]C_{xs} , \quad (3-34)$$

$$E_{xs} = -(A^T\bar{C}^{-1}A)^{-1}A^T\bar{C}^{-1}C_{xs} ; \quad (3-35)$$

here we have used (3-26).

On writing (3-34) as

$$E_{ss} = C_{ss} - C_{sx} \bar{C}^{-1} C_{xs} + H A E_{xx} A^T H^T, \quad (3-36)$$

where H is given by (3-24) we see that the term $H A E_{xx} A^T H^T$ represents the effect of inaccurate estimation of the parameters X . If there are no parameters, this term is zero.

Similarly, (3-35) may be written more simply as

$$E_{xs} = -E_{xx} A^T H^T. \quad (3-37)$$

The error variance of any estimated quantity s_p , expressed by (2-38), is given by

$$m_p^2 = C_{pp} - C_p^T \bar{C}^{-1} C_p + h_p^T A E_{xx} A^T h_p \quad (3-38)$$

(m_p is the standard error of estimation), and the error covariance between two estimated quantities s_p and s_q is

$$\sigma_{pq} = C_{pq} - C_p^T \bar{C}^{-1} C_q + h_p^T A E_{xx} A^T h_q. \quad (3-39)$$

These expressions follow directly from (3-36); the vector C_p is given by (2-37), and the vector h_p , by

$$h_p = \bar{C}^{-1} C_p \quad \text{or} \quad h_p^T = C_p^T \bar{C}^{-1}, \quad (3-40)$$

in agreement with (3-24); analogously for C_Q and h_Q . The quantity C_{PP} is the diagonal element, referring to s_P , of the matrix C_{ss} , and C_{PQ} is a corresponding off-diagonal element.

The equations (2-35), (2-36), (3-33), (3-36) and (3-37) together constitute the basic computational formulas for least-squares collocation, giving the estimates together with their accuracy.

Alternative expressions for (2-36) and (3-36) are (2-38) and (3-38) together with (3-39). Note that (3-38) and (3-39) are directly comparable to equations (7-64) and (7-65) of (Heiskanen and Moritz, 1967, pp. 269-270), which are nothing but a special case of our present, more general, formulas.

Least-Squares Collocation as the Most Accurate Estimation Method. - Let us now compare least-squares collocation to an arbitrary linear unbiased estimation method. Let the latter be characterized by the matrices \hat{G} and \hat{H} with

$$\hat{G}A = I \quad (3-41)$$

by (3-13), whereas G and H denote the corresponding least-squares matrices (3-23) and (3-24). In agreement with (3-10) we form the matrix

$$\hat{L} = \hat{H}(I - A\hat{G}), \quad (3-42)$$

whereas L is to be given by (3-26); obviously

$$\hat{L}A = 0. \quad (3-43)$$

Let us put

$$\hat{G} = G + g , \quad (3-44)$$

$$\hat{L} = L + \ell . \quad (3-45)$$

Then

$$gA = (\hat{G} - G)A = \hat{G}A - GA = I - I = O , \quad (3-46)$$

$$\ell A = (\hat{L} - L)A = \hat{L}A - LA = O . \quad (3-47)$$

By (3-19) and (3-20), the error covariance matrices for the arbitrary estimate are given by

$$\begin{aligned} \hat{E}_{XX} &= \hat{G}\bar{C}\hat{G}^T = (G+g)\bar{C}(G^T+g^T) \\ &= G\bar{C}G^T + g\bar{C}G^T + G\bar{C}g^T + g\bar{C}g^T , \\ \hat{E}_{SS} &= C_{SS} - \hat{L}C_{XS} - C_{SX}\hat{L}^T + \hat{L}C_{XX}\hat{L}^T \\ &= C_{SS} - (L+\ell)C_{XS} - C_{SX}(L^T+\ell^T) + (L+\ell)C_{XX}(L^T+\ell^T) \\ &= C_{SS} - LC_{XS} - C_{SX}L^T + LC_{XX}L^T - \\ &\quad - \ell C_{XS} - C_{SX}\ell^T + LC_{XX}\ell^T + \ell C_{XX}L^T + \ell C_{XX}\ell^T \end{aligned}$$

or

$$\hat{E}_{XX} = E_{XX} + g\bar{C}G^T + G\bar{C}g^T + g\bar{C}g^T , \quad (3-48)$$

$$\begin{aligned} \hat{E}_{SS} &= E_{SS} + (-\ell C_{XS} + \ell C_{XX}L^T) + (-C_{SX}\ell^T + LC_{XX}\ell^T) + \ell C_{XX}\ell^T , \\ &\quad (3-49) \end{aligned}$$

where E_{xx} and E_{ss} denote the error covariance matrices for least-squares collocation.

Now, by (3-23),

$$\begin{aligned} \bar{G}\bar{C}g^T &= (A^T\bar{C}^{-1}A)^{-1}A^T\bar{C}^{-1}\bar{C}g^T \\ &= (A^T\bar{C}^{-1}A)^{-1}A^Tg^T = 0 \end{aligned} \quad (3-50a)$$

because

$$A^Tg^T = 0$$

as the transpose of (3-46). Similarly

$$g\bar{C}G^T = 0 \quad (3-50b)$$

as the transpose of (3-50a).

By (3-26) we have

$$\begin{aligned} LC_{xx}l^T &= L\bar{C}l^T = C_{sx}\bar{C}^{-1}\left[I - A(A^T\bar{C}^{-1}A)^{-1}A^T\bar{C}^{-1}\right]\bar{C}l^T \\ &= C_{sx}l^T - C_{sx}\bar{C}^{-1}A(A^T\bar{C}^{-1}A)^{-1}A^Tl^T \\ &= C_{sx}l^T \end{aligned}$$

since

$$A^Tl^T = 0$$

as the transpose of (3-47). Thus

$$-C_{sx}l^T + LC_{xx}l^T = 0, \quad (3-51a)$$

and also for its transpose,

$$-lC_{xs} + lC_{xx}L^T = 0. \quad (3-51b)$$

On taking these relations into account, the expressions (3-48) and (3-49) reduce to

$$\hat{E}_{xx} = E_{xx} + g\bar{C}g^T, \quad (3-52a)$$

$$\hat{E}_{ss} = E_{ss} + l\bar{C}l^T. \quad (3-52b)$$

The error variances (squares of standard errors) of the estimated quantities are the diagonal terms of the matrices E_{xx} and E_{ss} . Let us consider the r -th component of X ; its error variance is given by

$$\hat{m}_r^2 = m_r^2 + \gamma_r \bar{C} \gamma_r^T, \quad (3-53)$$

where γ_r is the r -th row of the matrix g . Since \bar{C} is positive definite, as all covariance matrices are, we have

$$\gamma_r \bar{C} \gamma_r^T \geq 0; \quad (3-54)$$

thus

$$\hat{m}_r^2 \geq m_r^2, \quad (3-55)$$

so that least-squares collocation, to which m_r corresponds, gives indeed the smallest possible standard error of any component of the vector X .

If we apply the same reasoning, word by word, to any diagonal term of the matrix E_{ss} , we find that least-squares collocation gives also the smallest possible standard error of any component of the vector s .

Thus we have proved that least-squares collocation is optimal in the sense that it gives the most accurate results that are obtainable on the basis of the available data.

This property is well known from least-squares adjustment and least-squares prediction of gravity (cf. Heiskanen and Moritz, 1967, p. 268); we could, of course, also have used it as a basis for deriving the solution expressed by (2-35) and (2-36).

Error Covariances for the Complete Signal. - Consider the "complete signal" t consisting of the systematic part $A_1 X$ and the random signal s :

$$t = A_1 X + s. \quad (3-56)$$

The p -vector t comprises the "complete signals" at the p computation points; the term $A_1 X$ denotes the effect of the parameters X at the computation points, so that A_1 is a $p \times m$ matrix (note that the matrix A is a $q \times m$ matrix as (2-1) shows).

The derivation of the error covariance matrix of t , E_{tt} , is completely analogous to the derivation of E_{xx} and E_{ss} discussed previously in this section. The individual error of t is

$$\begin{aligned} \epsilon_t &= \bar{t} - t = A_1 \bar{X} + \bar{s} - (A_1 X + s) \\ &= A_1 \bar{X} + \bar{s} - (A_1 G + L)x \end{aligned}$$

$$\begin{aligned}
&= A_1 \bar{X} + \bar{S} - (A_1 G A + L A) \bar{X} - (A_1 G + L) \bar{Z} \\
&= \bar{S} - (A_1 G + L) \bar{Z} .
\end{aligned}$$

Here we have used (3-9a,b) and (3-2a). Thus

$$\epsilon_t = \bar{S} - K \bar{Z} \quad \text{with} \quad K = L + A_1 G , \quad (3-57)$$

whence

$$E_{tt} = C_{ss} - K C_{xs} - C_{sx} K^T + K C_{xx} K^T \quad (3-58)$$

follows in the same way as (3-20) follows from (3-15).

For least-squares prediction we have by (3-23)

$$K = L + A_1 G = L + A_1 E_{xx} A^T \bar{C}^{-1} ,$$

where L is given by (3-26) and E_{xx} by (3-33).

Hence,

$$\begin{aligned}
K C_{xs} &= L C_{xs} + A_1 G C_{xs} \\
&= L C_{xs} + A_1 E_{xx} A^T H^T
\end{aligned}$$

(by (3-23), (3-24) and (3-33)),

$$\begin{aligned}
C_{sx} K &= C_{sx} L + H A E_{xx} A_1^T , \\
K C_{xx} K^T &= L C_{xx} L^T + L \bar{C} G^T A_1^T \\
&\quad + A_1 G \bar{C} L^T + A_1 G \bar{C} G^T A_1^T .
\end{aligned}$$

By substituting (3-23) and (3-26) one verifies that

$$\bar{L}\bar{C}G^TA_1^T = 0, \quad A_1G\bar{C}L^T = 0,$$

and that

$$A_1G\bar{C}G^TA_1^T = A_1E_{XX}A_1^T.$$

Hence (3-58) becomes

$$\begin{aligned} E_{tt} = & C_{ss} - LC_{xs} - C_{sx}L^T + LC_{xx}L^T - \\ & - A_1E_{XX}A_1^TH^T - HAE_{XX}A_1^T + A_1E_{XX}A_1^T, \end{aligned}$$

and using (3-20) and (3-36) we finally get

$$E_{tt} = C_{ss} - C_{sx}\bar{C}^{-1}C_{xs} + (HA-A_1)E_{XX}(A_1^TH^T-A_1^T), \quad (3-59)$$

where H is given by (3-24) and E_{XX} by (3-33).

The use of E_{tt} is appropriate when the total signal t is of primary importance, such as in interpolation and transformation problems, whereas the use of E_{ss} is appropriate when the random signal s and the systematic part A_1X have different nature, such as in applications to physical geodesy.

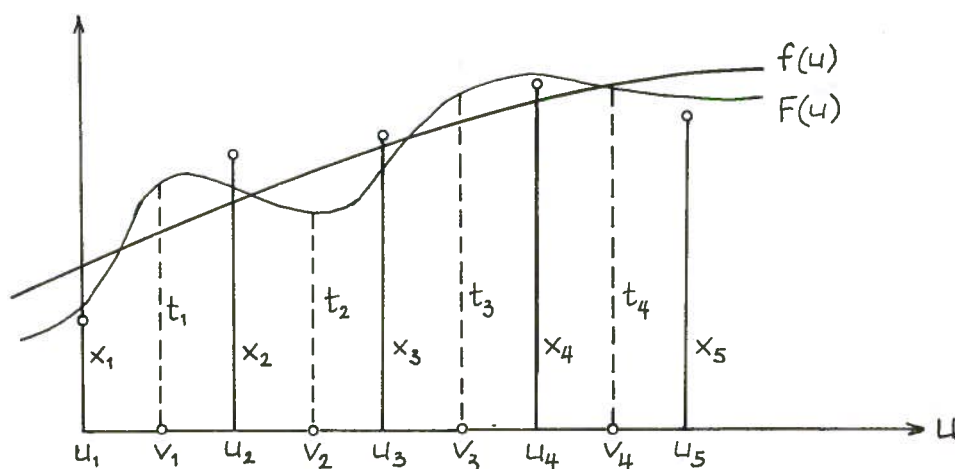
4, APPLICATION TO INTERPOLATION

The simplest application of least-squares collocation is to one-dimensional interpolation problems, as we have already outlined in sec. 1.

Let the function $F(u)$, which is to be interpolated, consist of a simple function $f(u)$ (e.g., a polynomial), and a random function (the "signal") $s(u)$:

$$F(u) = f(u) + s(u) ; \quad (4-1)$$

see Fig. 4-1.



u_1, u_2, \dots data points
 v_1, v_2, \dots interpolation points

Figure 4-1

The values of $F(u)$ at the "data points" u_1, u_2, \dots, u_q have been measured, but the measurements x_i are affected by random measuring errors n_i :

$$x_i = F(u_i) + n_i . \quad (4-2)$$

By (4-1), putting

$$s'_i = s(u_i) , \quad (4-3)$$

this becomes

$$x_i = f(u_i) + s'_i + n_i . \quad (4-4)$$

Let the function $f(u)$ depend on m parameters. For instance, if $f(u)$ is a polynomial

$$\begin{aligned} f(u) &= b_1 + b_2 u + b_3 u^2 + \dots + b_m u^{m-1} \\ &= \sum_{k=1}^m b_k u^{k-1} , \end{aligned} \quad (4-5)$$

then the parameters are the constants b_k ; they form a vector

$$X = [b_1 \ b_2 \ \dots \ b_m]^T . \quad (4-6)$$

Then

$$f(u_i) = \sum_{k=1}^m b_k u_i^{k-1} = \sum_{k=1}^m A_{ik} b_k \quad (4-7)$$

with

$$A_{ik} = u_i^{k-1} . \quad (4-8)$$

Thus (4-4) becomes

$$x_i = \sum_{k=1}^m A_{ik} b_k + s'_i + n_i , \quad (4-9)$$

or in matrix notation,

$$x = AX + s' + n , \quad (4-10)$$

with (4-6) and

$$x = [x_1 \ x_2 \ . \ . \ . \ x_q]^T = [x_i]^T , \quad (4-11a)$$

$$s' = [s'_1 \ s'_2 \ . \ . \ . \ s'_q]^T = [s'_i]^T , \quad (4-11b)$$

$$n = [n_1 \ n_2 \ . \ . \ . \ n_q]^T = [n_i]^T , \quad (4-11c)$$

$$A = \begin{bmatrix} 1 & u_1 & u_1^2 & . & . & . & u_1^{m-1} \\ 1 & u_2 & u_2^2 & . & . & . & u_2^{m-1} \\ . & . & . & & & & . \\ . & . & . & & & & . \\ . & . & . & & & & . \\ 1 & u_q & u_q^2 & . & . & . & u_q^{m-1} \end{bmatrix} = [u_i^{k-1}] . \quad (4-12)$$

But (4-10) is nothing else than the basic model (2-1).

The interpolation problem consists in finding the value of $F(u)$ at an arbitrary "interpolation point" on the u -axis; let these interpolation points be denoted by v_1, v_2, \dots, v_p (Fig. 4-1). We put

$$F(v_i) = t_i, \quad (4-13a)$$

$$s(v_i) = s_i, \quad (4-13b)$$

$$f(v_i) = \sum_{k=1}^m b_k v_i^{k-1}, \quad (4-14)$$

and introduce the p -vectors

$$t = [t_i]^T, \quad (4-15a)$$

$$s = [s_i]^T, \quad (4-15b)$$

and the $p \times m$ matrix

$$A_1 = [v_i^{k-1}], \quad (4-16)$$

which is analogous to the matrix (4-12) but has p instead of q rows.

Then the vector t , comprising the values of $F(u)$ to be interpolated, may be written as

$$t = A_1 X + s, \quad (4-17)$$

which is identical to the "complete signal" (3-56).

So far we have assumed the function $f(u)$ to be a polynomial (4-5), which is linear in the parameters b_k . We may, however, also use functions nonlinear in the parameters. If they are linearized in the usual way, we get again linear relations of the form (4-10) and (4-17).

The Covariance Function. - The signal $s(u)$ is a continuous function of u . Then all signal covariances can be derived from a single function, the covariance function $C(r)$, which is defined as

$$C(r) = M\{ss'\} , \quad (4-18)$$

where s refers to a point P and s' refers to a point P' such that $PP' = r$ (both points P and P' are on the u -axis); the average M is to be extended over all pairs of points P and P' that are at a constant distance $PP' = r$ apart. More precisely, we may form the function

$$C_1(r) = \frac{1}{2(a-r)} \int_{-a+r}^{a-r} s(u)s(u+r) du , \quad (4-19)$$

representing the average over a profile of length $2a$, and form the mean of as many different profiles as we can compute. As a matter of fact, the integral in (4-19) is to be approximated by a sum for numerical evaluation. The covariance computation program of Rapp (1966) may be applied also to the present problem.

It is in general not advisable to directly use the "raw" covariance values so obtained. It is preferable to fit an analytical expression to the raw values. This expression must, above all, represent a positive-definite function because all admissible covariance functions are

positive definite (Papoulis, 1965, p.349).

A simple positive definite function that may be appropriate in many cases is

$$C(r) = C_0 e^{-a^2 r^2} . \quad (4-20)$$

Signal and Error Covariances. - All signal covariances can be expressed in terms of the covariance function. We have

$$\text{cov}(s'_i, s'_j) = C(r_{ij}) \quad (4-21)$$

where

$$r_{ij} = |u_j - u_i| \quad (4-22)$$

is the distance between the points u_i and u_j to which the signals refer according to (4-3). In the same way,

$$\text{cov}(s_i, s'_j) = C(r'_{ij}) \quad (4-23)$$

where

$$r'_{ij} = |u_j - v_i| . \quad (4-24)$$

The quantities (4-21) form the matrix $C_{s's'} = C$ so that the covariance matrix of x is found by (2-20):

$$\bar{C} = C + D , \quad (4-25)$$

by adding to C the covariance matrix $D = C_{nn}$ of the measuring errors n .

On the other hand, the covariance matrix C_{sx} is composed only of signal covariances, so that its elements are directly given by (4-23).

Computational Formulas. - Now we are in a position to apply the formulas of sections 2 and 3. First, the parameter vector X is obtained by (2-35):

$$X = (A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1} x, \quad (4-26)$$

so that the "functional part" $f(u)$ of the interpolation curve $F(u)$ is determined.

The random part $s(u)$ at the interpolation points v_i , i. e., the vector (4-15b), is obtained by (2-36):

$$s = C_{sx} \bar{C}^{-1} (x - AX). \quad (4-27)$$

Then the interpolated values of $F(u)$, forming the "total signal", the vector t , are finally given by (4-17).

The accuracy of the estimated quantities X , s , and t is characterized by the matrices (3-33), (3-36), (3-37), and (3-59):

$$E_{XX} = (A^T \bar{C}^{-1} A)^{-1}, \quad (4-28a)$$

$$E_{ss} = C_{ss} - C_{sx} \bar{C}^{-1} C_{xs} + HA E_{XX} A^T H^T, \quad (4-28b)$$

$$E_{tt} = C_{ss} - C_{sx} \bar{C}^{-1} C_{xs} + (HA - A_1) E_{XX} (A^T H^T - A_1^T) \quad (4-28c)$$

are the error covariance matrices of X , s , and t , respectively; and

$$E_{xs} = -E_{xx} A^T H^T \quad (4-29)$$

is the error-crosscovariance matrix expressing the error correlation between the parameters X and the random signal s . The matrix H is an abbreviation for $C_{sx} \bar{C}^{-1}$.

Since t represents the curve $F(u)$ to be interpolated, the matrix E_{tt} is of particular importance for such interpolation problems.

A Numerical Example. - These formulas will be illustrated by a simple numerical example computed by Dr. K.P. Schwarz.

Here $f(u)$ is the straight line

$$f(u) = b_1 + b_2 u ,$$

so that

$$X = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

and $m = 2$. There are $q = 5$ equally spaced data points u_i and $p = 4$ interpolation points situated halfway between the data points, somewhat as in Fig. 4-1.

Data

u_i	x_i
0.000	0.6108
1.445	1.0863
2.890	2.9034
4.335	4.5925
5.780	6.2714

The covariance function is an expression (4-20) with

$$C_0 = 0.1260 , \quad a = 0.6 .$$

The covariance matrix of the measuring errors, C_{nn} , is assumed to be a multiple of the unit matrix I ,

$$C_{nn} = \sigma^2 I \quad \text{with} \quad \sigma^2 = 0.01 .$$

Covariance Matrices

$$C_{xx} = \begin{bmatrix} 0.1360 & 0.0594 & 0.0062 & 0.0002 & 0.0000 \\ 0.0594 & 0.1360 & 0.0594 & 0.0062 & 0.0002 \\ 0.0062 & 0.0594 & 0.1360 & 0.0594 & 0.0062 \\ 0.0002 & 0.0062 & 0.0594 & 0.1360 & 0.0594 \\ 0.0000 & 0.0002 & 0.0062 & 0.0594 & 0.1360 \end{bmatrix}$$

$$C_{ss} = \begin{bmatrix} 0.1260 & 0.0594 & 0.0062 & 0.0002 \\ 0.0594 & 0.1260 & 0.0594 & 0.0062 \\ 0.0062 & 0.0594 & 0.1260 & 0.0594 \\ 0.0002 & 0.0062 & 0.0594 & 0.1260 \end{bmatrix}$$

$$C_{xs} = \begin{bmatrix} 0.1044 & 0.1044 & 0.0232 & 0.0012 & 0.0000 \\ 0.0232 & 0.1044 & 0.1044 & 0.0232 & 0.0012 \\ 0.0012 & 0.0232 & 0.1044 & 0.1044 & 0.0232 \\ 0.0000 & 0.0012 & 0.0232 & 0.1044 & 0.1044 \end{bmatrix}$$

Results

Parameters

$$X = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.3252 \\ 0.9891 \end{bmatrix}$$

Interpolated Values

v_i	s_i	$t_i = F(v_i)$
0.7225	-0.2221	0.8177
2.1675	-0.5590	1.9101
3.6125	-0.1052	3.7932
5.0575	0.1082	5.4359

Error Covariance Matrices

$$E_{xx} = \begin{bmatrix} 0.1136 & -0.0234 \\ -0.0234 & 0.0081 \end{bmatrix}$$

$$E_{ss} = \begin{bmatrix} 0.1034 & 0.0532 & 0.0343 & 0.0025 \\ 0.0532 & 0.0564 & 0.0369 & 0.0343 \\ 0.0343 & 0.0369 & 0.0564 & 0.0532 \\ 0.0025 & 0.0343 & 0.0532 & 0.1034 \end{bmatrix}$$

$$E_{tt} = \begin{bmatrix} 0.0131 & -0.0024 & 0.0015 & -0.0009 \\ -0.0024 & 0.0121 & -0.0019 & 0.0015 \\ 0.0015 & -0.0019 & 0.0121 & -0.0024 \\ -0.0009 & 0.0015 & -0.0024 & 0.0131 \end{bmatrix}$$

$$E_{xs} = \begin{bmatrix} 0.1007 & 0.0578 & 0.0298 & -0.0080 \\ -0.0188 & -0.0048 & 0.0048 & 0.0188 \end{bmatrix}$$

5. APPLICATION TO TRANSFORMATION PROBLEMS

Least-squares collocation also lends itself to application to coordinate transformations occurring in geodesy and photogrammetry. In fact, such applications are the twodimensional (or threedimensional) analogue to one-dimensional interpolation as discussed in the preceding section.

Consider, for instance, the following simple transformation problem (Helmert transformation). Let the plane be referred to two coordinate systems ξ, η and $\bar{\xi}, \bar{\eta}$. These two systems are related by a translation and a rotation; furthermore the scale is different.

Thus the transformation equations are:

$$\bar{\xi} = a + k(\xi \cos \alpha + \eta \sin \alpha) , \quad (5-1)$$

$$\bar{\eta} = b + k(-\xi \sin \alpha + \eta \cos \alpha) .$$

The vector (a,b) represents the translation, α is the angle between the ξ and $\bar{\xi}$ axes, and k is a scale factor.

Let the transformation parameters (a,b,α,k) be unknown, but let there be a number of "identical" points of which the coordinates ξ, η , and $\bar{\xi}, \bar{\eta}$ in both systems are given. The problem is to determine the transformation parameters and to transform a series of other points from the $\xi\eta$ -system to the $\bar{\xi}\bar{\eta}$ -system.

First we must linearize eqs. (5-1). The standard procedure would be to introduce approximate values $(a_o, b_o, \alpha_o, k_o)$ and to linearize by means of Taylor's theorem. In this particular case, however, one proceeds in a much simpler way. Instead of k and α we introduce new parameters c and d by

$$c = k \cos \alpha, \quad d = k \sin \alpha, \quad (5-2)$$

obtaining

$$\bar{\xi} = a + \xi c + \eta d, \quad (5-3)$$

$$\bar{\eta} = b - \xi d + \eta c.$$

Assume that ξ, η are affected by measuring errors v_ξ, v_η , and similarly $\bar{\xi}, \bar{\eta}$ by $\bar{v}_\xi, \bar{v}_\eta$. Then (5-3) becomes

$$\bar{\xi} + \bar{v}_\xi = a + (\xi + v_\xi)c + (\eta + v_\eta)d, \quad (5-4)$$

$$\bar{\eta} + \bar{v}_\eta = b - (\xi + v_\xi)d + (\eta + v_\eta)c,$$

or

$$\bar{\xi} = a + \xi c + \eta d + n_\xi, \quad (5-5)$$

$$\bar{\eta} = b - \xi d + \eta c + n_\eta,$$

where we have put

$$\begin{aligned} n_{\xi} &= -\bar{v}_{\xi} + v_{\xi}c + v_{\eta}d, \\ n_{\eta} &= -\bar{v}_{\eta} - v_{\xi}d + v_{\eta}c. \end{aligned} \quad (5-6)$$

If r is the number of "identical" points, we can form r pairs of equations (5-5). The $2r$ equations so obtained may be abbreviated as

$$x = AX + n \quad (5-7)$$

with

$$x = \begin{bmatrix} \bar{\xi}_1 \\ \bar{\eta}_1 \\ \bar{\xi}_2 \\ \bar{\eta}_2 \\ \vdots \\ \bar{\xi}_r \\ \bar{\eta}_r \end{bmatrix}, \quad n = \begin{bmatrix} n_{\xi 1} \\ n_{\eta 1} \\ n_{\xi 2} \\ n_{\eta 2} \\ \vdots \\ n_{\xi r} \\ n_{\eta r} \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & \xi_1 & \eta_1 \\ 0 & 1 & \eta_1 - \xi_1 & \\ 1 & 0 & \xi_2 & \eta_2 \\ 0 & 1 & \eta_2 - \xi_2 & \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \xi_r & \eta_r \\ 0 & 1 & \eta_r - \xi_r & \end{bmatrix}, \quad X = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}. \quad (5-8)$$

On writing (5-7) as

$$x - n = AX, \quad (5-9)$$

this is seen to be the usual observation equation for adjustment by parameters: x represents the observations, $-n$ the corrections, and X the parameters.

We might thus perform usual adjustment by parameters, but the behavior of the residuals n obtained in this way will frequently be different from the expected statistical behavior of measuring errors, corresponding to the assumed covariance matrix of n : the residuals will be larger and fairly strongly correlated for neighboring points, whereas n_{ξ} and n_{η} , though correlated when they refer to the same point by (5-6), are uncorrelated for different points.

If these residuals are plotted above the xy -plane, their contour lines might look somewhat like Fig. 5-1.

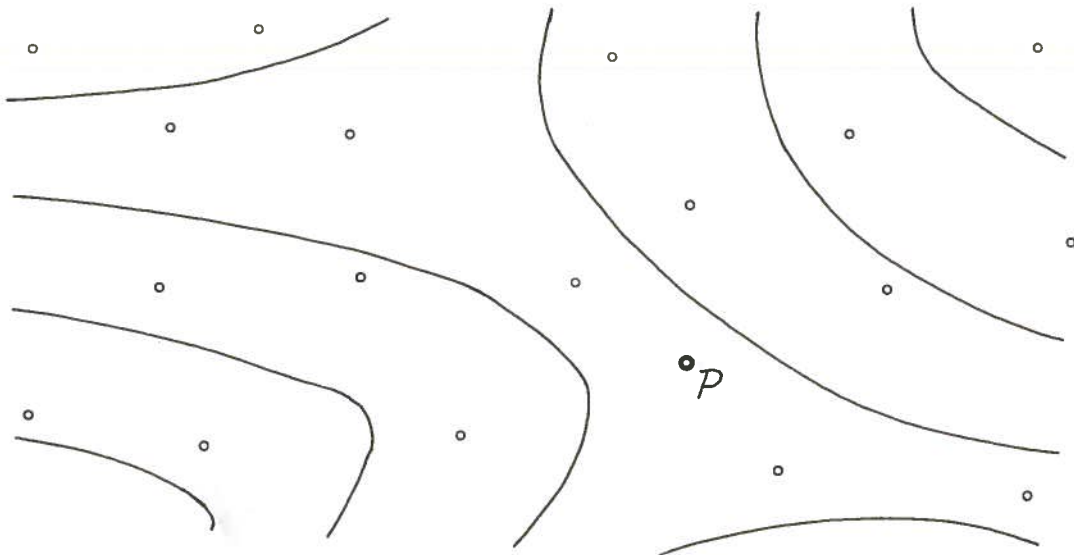


Figure 5-1

This indicates that, in this case, the chosen model (5-7) is not appropriate. We must take into account random distortions of the functional model such as shown in Fig. 5-1. We therefore add to (5-7) a term representing such random distortions, to arrive at the expression

$$x = AX + s + n , \quad (5-10)$$

which is nothing but the basic model (1-2).

This heuristic argument has been given here to justify the need for introducing this refined model. In fact, such a situation arises in many transformation problems of geodesy and, it seems, in most transformation problems of photogrammetry, where random distortions of the functional model occur almost inevitably.

Using the model (5-10), the solution to our problem is given by (2-35) and (2-38):

$$X = (A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1} x , \quad (5-11)$$

$$s_p = C_p^T \bar{C}^{-1} (x - AX) . \quad (5-12)$$

Eq. (5-11) determines the vector X , that is, the transformation parameters a, b, c, d . Then (5-12) determines the signal s_p , that is, the residual distortion in ξ or η , at any point P (Fig. 5-1).

A central role in this method is played by the covariances as represented by the matrices \bar{C} and C_p . These covariances will now be considered in some detail.

The Covariance Function. - The signal, the distortion in ξ or η at any point of the plane, may be represented as a continuous function of the coordinates ξ and η ,

$$s = s(\xi, \eta) ; \quad (5-13)$$

to fix our ideas, let s denote the distortion in the ξ -direction; for the moment, the distortion in the η -direction is disregarded.

As in the preceding section, all signal covariances can be expressed by a covariance function which is defined as follows. Consider the signal s at a point P and the signal s' at a point P' , and form their product ss' . Take the average M of the product ss' for all pairs of points P and P' , situated in the area under consideration, that are at a constant distance $PP' = r$ apart. This average product is the covariance of the signal s for the distance r ,

$$C(r) = M\{ss'\} \quad (PP' = r), \quad (5-14)$$

which, considered as a function of r , is the covariance function for our present problem.

The way of taking the average M may be illustrated by considering a rectangular area of sides $2a$ and $2b$ (Fig. 5-2). Let us form the average

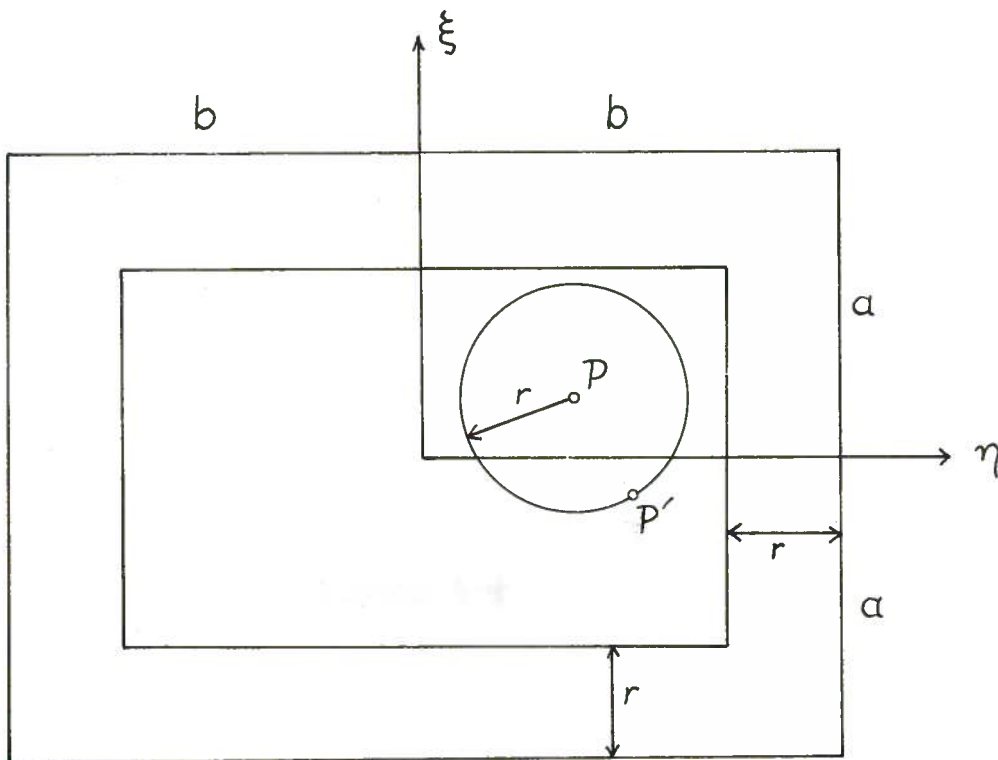


Figure 5-2

M over this area in two steps. First we average over a circle of radius r with center at P , letting P' move along the circle: this gives

$$\frac{1}{2\pi} \int_{\alpha=0}^{2\pi} s(\xi, \eta) s(\xi + r \cos \alpha, \eta + r \sin \alpha) d\alpha . \quad (5-15)$$

This quantity is then averaged by letting P move within the concentric rectangle of sides $2a-2r$ and $2b-2r$ with the result

$$C(r) = \frac{1}{8\pi(a-r)(b-r)} \int_{\xi=-a+r}^{a-r} \int_{\eta=-b+r}^{b-r} \int_{\alpha=0}^{2\pi} s(\xi, \eta) \cdot s(\xi + r \cos \alpha, \eta + r \sin \alpha) d\xi d\eta d\alpha . \quad (5-16)$$

This equation gives a value of $C(r)$ only for $r \leq c$ where c denotes a or b whichever is smaller (a formally satisfactory definition might be achieved by putting $b = a$ and then letting $a \rightarrow \infty$). Practically this is not very relevant since the covariance function can only be estimated by representative sampling and by fitting an analytical expression to the sampled values. This analytical expression then gives $C(r)$ also for larger values of r .

As in the preceding section, the function

$$C(r) = C_0 e^{-a^2 r^2} , \quad (5-17)$$

with suitable constants C_0 and a , provides a simple analytical expression that may be appropriate in many cases; it satisfies the necessary requirement of being positive definite.

The covariance function $C(r)$ readily gives the covariance between the signals at any two prescribed points, say P_i and P_j : we simply have

$$C_{ij} = C(r_{ij}) , \quad (5-18)$$

where r is the distance between P_i and P_j .

The fact that the covariance depends only on the relative position of P_i and P_j and not, e.g., on the coordinates ξ, η of P_i is called homogeneity. It is expressed by the average over ξ and η (the first two integrals) in (5-16). The fact that the covariance depends only on the distance r and not on the direction α is called isotropy; it is expressed by the average over α (the third integral) in (5-16).

The definition of the covariance function given here is a straightforward extension of the concept of the covariance function of the gravity anomaly Δg to an arbitrary signal s ; cf. (Heiskanen and Moritz, 1967, sec. 7-2).

Covariance Functions for the Distortions. - So far we have only considered one signal function (5-13); in reality, however, we must simultaneously consider both components of the distortion:

$$s_\xi = s_\xi(\xi, \eta) , \quad (5-19a)$$

the component in the ξ -direction, and

$$s_{\eta} = s_{\eta}(\xi, \eta) , \quad (5-19b)$$

the component in the η -direction.

Correspondingly we have four covariance functions: the expressions

$$C_{\xi\xi}(r) = M\{s_{\xi}s'_{\xi}\} , \quad (5-20a)$$

$$C_{\eta\eta}(r) = M\{s_{\eta}s'_{\eta}\} \quad (5-20b)$$

represent the autocovariance functions: the covariance function of s_{ξ} and that of s_{η} ; whereas the expressions

$$C_{\xi\eta}(r) = M\{s_{\xi}s'_{\eta}\} , \quad (5-21a)$$

$$C_{\eta\xi}(r) = M\{s_{\eta}s'_{\xi}\} \quad (5-21b)$$

represent the crosscovariance functions between s_{ξ} and s_{η} .

If all correlation functions are computed with the average M defined as in (5-16), then they are in fact functions of r only. Let us now assume complete isotropy, such that all directions are equivalent. Then there will be the same behavior in the ξ -direction as in the η -direction, so that $C_{\eta\eta} = C_{\xi\xi}$; let us put

$$C_{\xi\xi}(r) = C_{\eta\eta}(r) = C(r) . \quad (5-22)$$

Furthermore the crosscovariances will be zero:

$$C_{\xi\eta}(r) = C_{\eta\xi}(r) = 0 . \quad (5-23)$$

This is seen as follows. Consider a rotation of the coordinate system:

$$\bar{\xi} = \xi \cos \alpha + \eta \sin \alpha , \quad (5-24)$$

$$\bar{\eta} = -\xi \sin \alpha + \eta \cos \alpha .$$

Then the distortion components along the new axes $\bar{\xi}$ and $\bar{\eta}$ will be given by the same transformation:

$$\bar{s}_{\xi} = s_{\xi} \cos \alpha + s_{\eta} \sin \alpha , \quad (5-25)$$

$$\bar{s}_{\eta} = -s_{\xi} \sin \alpha + s_{\eta} \cos \alpha ,$$

and the new autocovariance function becomes

$$\begin{aligned} \bar{C}_{\xi\xi} &= M\{\bar{s}_{\xi}\bar{s}'_{\xi}\} = M\{(s_{\xi} \cos \alpha + s_{\eta} \sin \alpha)(s'_{\xi} \cos \alpha + s'_{\eta} \sin \alpha)\} \\ &= M\{s_{\xi}s'_{\xi} \cos^2 \alpha + (s_{\xi}s'_{\eta} + s_{\eta}s'_{\xi}) \cos \alpha \sin \alpha + s_{\eta}s'_{\eta} \sin^2 \alpha\} \\ &= C_{\xi\xi} \cos^2 \alpha + (C_{\xi\eta} + C_{\eta\xi}) \cos \alpha \sin \alpha + C_{\eta\eta} \sin^2 \alpha \quad (5-26) \end{aligned}$$

or

$$\bar{C}_{\xi\xi} = C_{\xi\xi} + (C_{\xi\eta} + C_{\eta\xi}) \cos \alpha \sin \alpha$$

by (5-22). Because of complete isotropy, there will be $\bar{C}_{\xi\xi} = C_{\xi\xi}$, so that

$$C_{\xi\eta} + C_{\eta\xi} = 0 . \quad (5-27)$$

Since $C_{\xi\eta}$ depends only on r and since $PP' = P'P$, we have

$$C_{\eta\xi} = C_{\xi\eta} \quad (5-28)$$

which, together with (5-27), gives (5-23).

A more general concept of isotropy, in which (5-23) need not be satisfied, will be discussed in sec. 8; cf. eqs. (8-18), but the special case expressed by (5-22) and (5-23) may be appropriate for many applications.

For instance, we may again assume

$$C_{\xi\xi}(r) = C_{\eta\eta}(r) = C_0 e^{-a^2 r^2}, \quad (5-29)$$

$$C_{\xi\eta}(r) = C_{\eta\xi}(r) = 0.$$

Signal and Error Covariances. - Now we are in a position to compute all covariances that enter into the collocation formulas (5-11) and (5-12).

If there are no observational errors, then all elements of the matrices $\bar{C} = C$ (by (2-20)) and C_p are signal covariances. The matrix C has the form

$$C = \begin{bmatrix} C_{11} & C_{12} & \cdot & \cdot & C_{1q} \\ C_{21} & C_{22} & \cdot & \cdot & C_{2q} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ C_{q1} & C_{q2} & & & C_{qq} \end{bmatrix} = [C_{ij}] , \quad (5-30)$$

where

$$C_{ij} = \text{cov}(s_i, s_j) , \quad (5-31)$$

s_i being the signal that corresponds to the measurement x_i .

We can distinguish four cases:

$$(a) \quad s_i = s_{\xi, A}, \quad s_j = s_{\xi, B},$$

that is, s_i is the s_{ξ} -component at some point A, and s_j is the s_{ξ} -component at some point B. Then

$$C_{ij} = C_{\xi\xi}(r_{AB}), \quad (5-32a)$$

where r_{AB} is the distance AB. Similarly, for the case

$$(b) \quad s_i = s_{\eta, A}, \quad s_j = s_{\eta, B}$$

we have

$$C_{ij} = C_{\eta\eta}(r_{AB}). \quad (5-32b)$$

If

$$(c) \quad s_i = s_{\xi, A}, \quad s_j = s_{\eta, B},$$

then

$$C_{ij} = C_{\xi\eta}(r_{AB}), \quad (5-32c)$$

and finally for

$$(d) \quad s_i = s_{\eta, A}, \quad s_j = s_{\xi, B}$$

we have

$$C_{ij} = C_{\eta\xi}(r_{AB}) . \quad (5-32d)$$

If (5-23) holds, then the last two covariances will be zero.

In this way all elements of the matrix (5-30) are obtained.

The matrix C_P has the form (2-37), briefly

$$C_P = \left[C_{Pi} \right]^T , \quad (5-33)$$

where

$$C_{Pi} = \text{cov}(s_P, s_i) . \quad (5-34)$$

Since also s_P is s_ξ or s_η at some point A , we again have one of the four cases just considered, so that we obtain C_{Pi} , too, by one of the formulas (5-32a-d).

So far we have assumed that there are no measuring errors. If there are measuring errors n_i , so that

$$x_i = s_i + n_i , \quad (5-35)$$

if these errors correspond to an error covariance matrix

$$D = \left[D_{ij} \right] \quad \text{with} \quad D_{ij} = \text{cov}(n_i, n_j) , \quad (5-36)$$

and if the n_i are uncorrelated to the signal s_i (which can reasonably be assumed here), then the matrix \bar{C} is given by

$$\bar{C} = C + D , \quad (5-37)$$

by adding the error covariance matrix D to the signal covariance matrix (5-30), whereas the matrix C_p remains the same signal covariance matrix as in the errorless case and is again given by (5-33) with (5-34). This follows at once from (2-20) and (2-21).

Concluding Remarks. - After determining the vector $X = (a, b, c, d)^T$ by (5-11), we can transform any point in the plane from the coordinate system $\xi\eta$ to the system $\bar{\xi}\bar{\eta}$ by means of the formulas

$$\begin{aligned}\bar{\xi} &= a + \xi c + \eta d + s_{\xi}(\xi, \eta), \\ \bar{\eta} &= b - \xi d + \eta c + s_{\eta}(\xi, \eta),\end{aligned}\tag{5-38}$$

which follow by adding to (5-3) the distortion components s_{ξ} and s_{η} in agreement with (5-10); these quantities s_{ξ} and s_{η} are to be predicted by (5-12).

These are the complete transformation formulas that also take distortion into account.

Obviously the quantities $\bar{\xi}$ and $\bar{\eta}$ are nothing else than the "complete signal", represented by the vector equation

$$t = AX + s\tag{5-39}$$

as the sum of the systematic part and the random signal. The accuracy of $\bar{\xi}$ and $\bar{\eta}$ is therefore expressed by the matrix E_{tt} (3-59).

It is instructive to compare (5-38) to the usual procedure. There the vector X is obtained from a least-squares adjustment, whereas the distortion s is either neglected or obtained by a graphical technique:

the residuals of the adjustment, both in ξ and in η , are considered as distortions s_ξ and s_η and plotted on the xy -plane; then lines of equal s_ξ and s_η are drawn (somewhat like Fig. 5-1), and s_ξ and s_η are obtained by graphical interpolation.

The present method replaces, as it were, this graphical interpolation by an objective numerical procedure with optimum properties.

It is obvious that the functional model AX of the transformation, instead of representing a Helmert transformation, may be any other transformation, e.g., an affine, projective, or conformal transformation.¹⁾

A nonlinear model may be said to incorporate systematic distortions. Hence a linear model (Helmert or affine transformation) may be supplemented by quadratic or higher-order terms, the parameters of which are included in the vector X , to take systematic distortions into account (Hubeny, 1953, pp. 110-115).

The decision to which extent distortions are "systematic", to be incorporated in the parameter vector X , and "random", to be represented by the signal s , is not always straightforward. As a general rule, smooth effects are to be represented in the form AX , and irregular effects are to be represented as a signal s , but there are borderline cases. In fact, irregular behavior might also be accounted for by a very high-order polynomial. To a certain extent, the separation of the distortion into a systematic and a random part will be a matter of computational convenience.

¹⁾ For a comprehensive treatment cf. (Schatz, 1970); he considers adjustment and prediction methods, but not yet the combination of both by collocation.

Investigations into this problem of separation have been performed by Lauer (1971) for the case of gravity prediction which is quite analogous, being covered by the same model (Moritz, 1969, sec.10). Lauer has tried to split off a systematic part from the gravity anomaly field before applying least-squares prediction.

Finally we wish to point out the difference between the collocation formula (5-11) for the parameters and the adjustment formula: for least-squares collocation we have

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

whereas conventional least-squares adjustment yields

$$X = (A^T D A)^{-1} A^T D^{-1} x , \quad (5-40)$$

as one readily verifies.

Thus the only difference, as far as the parameters are concerned, is that the error covariance matrix D is replaced by the matrix \bar{C} that incorporates both signal and error covariances by (5-37). This fact will also be of importance for the determination of geodetic parameters to be studied in the following section.

6. APPLICATION TO PHYSICAL GEODESY

Historically the first and still of foremost importance are the applications of advanced least-squares methods to physical geodesy. The first geodetic application of such methods--beyond classical least-squares adjustment--was in fact to a problem of physical geodesy, namely, to interpolation and prediction of gravity (Kaula, 1963; Moritz, 1962, 1963); see also (Heiskanen and Moritz, 1967, sec. 7-6). Also the decisive generalization of Krarup (1968, 1969), to some extent foreshadowed by Kaula (1963, 1967) was motivated by the intention to solve a principal problem of physical geodesy, the determination of the terrestrial gravity field.

Even without introduction of systematic parameters, a very general and powerful method is obtained, which permits an optimal determination of the anomalous gravity field from data of different type, systematic effects being absent. Problems of this kind have been studied, in some detail, in a previous report (Moritz, 1970a). Although the present report can be read independently, some knowledge of the previous report may be useful for a complete understanding of secs. 6 and 7.

We shall here extend the scope of this method by admitting systematic parameters X . In this way we are able to take into account, not only systematic errors, but also parameters of the reference ellipsoid, station coordinates, etc., obtaining a unified treatment of both geometrical and physical quantities.

In the application to physical geodesy, the anomalous gravity field is, so to speak, the signal field: all signals s are quantities of this field, such as gravity anomalies, deflections of the vertical, geoidal heights, anomalous

gradients, spherical-harmonic coefficients, etc. All these quantities may be derived from one basic function, e.g., from the anomalous potential T , by linear operations such as differentiations or integral formulas.

In the same way, all signal covariance functions (say, the covariance functions for the deflection of the vertical) may be derived from one basic covariance function, e.g., from the covariance function for the anomalous potential, by the corresponding linear operations.

The problem of covariances has already been treated in sec. 4 of (Moritz, 1970a). In order to lead the reader familiar with that report directly to the new aspects to which the introduction of systematic parameters gave rise, we shall postpone the consideration of covariances to the following three sections, where the statistical background of least squares collocation as applied to the gravity field will be rather broadly discussed.

The basic model (1-2),

$$x = AX + s + n, \quad (6-1)$$

applies, in fact, to our situation. As we have already remarked in sec. 1, any geodetic measurement may be split up into three components:

- (a) a systematic part comprising effects of the ellipsoidal reference system, station coordinates and other geometric parameters, as well as systematic measuring errors;
- (b) a random signal part s expressing the effect of the anomalous gravity field;
- (c) a random noise part n expressing the observational errors.

We shall now demonstrate this fact by analyzing various measurable quantities x in this way. It is obviously sufficient to effect a decomposition

$$x = AX + s, \quad (6-2)$$

disregarding the measuring error n , because n can always be added afterwards.

Consider first the classical measurements of physical geodesy: magnitude and direction of the gravity vector, the first being gravity g , the second being the direction of the plumb line as defined by astronomical latitude, ϕ , and astronomical longitude, λ .

We have

$$g = \gamma + \Delta g, \quad (6-3)$$

where γ is normal gravity and Δg is the gravity anomaly. Normal gravity depends on the four parameters (denoted by p_1, p_2, p_3, p_4) defining the reference system used:

$$\gamma = \gamma(p_1, p_2, p_3, p_4); \quad (6-4)$$

to get a linear expression, introduce approximate values p_i^0 , set $p_i = p_i^0 + \delta p_i$ and linearize; then X is the vector

$$X = [\delta p_1 \ \delta p_2 \ \delta p_3 \ \delta p_4]^T. \quad (6-5)$$

(In the Geodetic Reference System 1967 (Levallois, 1971) we have $p_1 = a$, the semimajor axis, $p_2 = GM$, the product of gravitational constant and mass of the earth,

$p_3 = J_2$, the zonal harmonic coefficient of degree 2, and
 $p_4 = \omega$, the rotational velocity of the earth.)

Thus in (6-3), γ represents AX , and Δg represents s .

In a similar way we may decompose the astronomical coordinates Φ and Λ :

$$\Phi = \phi + \xi, \quad (6-6)$$

$$\Lambda = \lambda + \eta \sec \phi,$$

where ϕ and λ are the corresponding geodetic coordinates which depend on the reference ellipsoid, such that

$$\phi = \phi(p_1, p_2, p_3, p_4), \quad \lambda = \lambda(p_1, p_2, p_3, p_4), \quad (6-7)$$

analogous to (6-4), and where the deflections of the vertical express the effect of the anomalous gravity field.

That is, ϕ and λ constitute the systematic part AX , and ξ and $\eta \sec \phi$ represent the signal part s .

In (Moritz, 1970a) we have considered the parameters of the reference ellipsoid as known; then the vector (6-5) is zero, so that we have been able to consider Δg , ξ , η directly as observations.

As we shall see in the following section, the signal field should not contain spherical harmonics of degrees 0, 1, and 2. This implies that the reference ellipsoid is in an absolute (i.e., geocentric) position. Cf. (Heiskanen and Moritz, 1967), pp. 99-100; henceforth this book will be referred to briefly as "HM".

Astrogeodetic deflections of the vertical correspond to a non-geocentric reference ellipsoid; they must therefore be transformed by shifting the ellipsoid into an absolute position (HM, p.209), before applying them in combination procedures such as the one described in (Moritz, 1970a, sec. 9).

It is, however, also possible to determine the shift parameters simultaneously by collocation: we write the astrogeodetic deflections ξ^a , η^a in the form (HM, p. 213):

$$\begin{aligned}\xi^a &= \frac{1}{R}(\delta x_o \sin\phi \cos\lambda + \delta y_o \sin\phi \sin\lambda - \delta z_o \cos\phi) + \xi, \\ \eta^a &= \frac{1}{R}(\delta x_o \sin\lambda - \delta y_o \cos\lambda) + \eta,\end{aligned}\tag{6-8}$$

where δx_o , δy_o , δz_o are the components of the shift of the reference ellipsoid and R is a mean radius of the earth.

Obviously this again fits into the model (6-2): ξ^a and η^a are observations representing x , the first terms on the right-hand side represent AX with

$$X = [\delta x_o, \delta y_o, \delta z_o]^T, \tag{6-9}$$

and the geocentric deflections ξ and η form the signal s .

The present method thus makes it possible to obtain at the same time:

- (a) a combined gravimetric-astrogeodetic geoid;
- (b) the shift of the astrogeodetic reference ellipsoid to its absolute position.

It may be regarded as a combination of the method described in (Moritz, 1970a, sec. 9)--with respect to (a)-- and the determination of the shift parameters by combining astrogeodetic and gravimetric data (HM, sec. 5-10)-- with respect to (b).

But also any other observational quantities, which at first sight seem to be purely geometric, fit into the general scheme (6-2), for instance, measured azimuths, horizontal angles, and zenith distances.

By eqs. (5-10) and (5-13) of HM, p. 186 we have, after a slight change of notation,

$$A = A' + \eta \tan \phi + (\xi \sin A - \eta \cos A) \cot z, \quad (6-10)$$

where A denotes the measured ("astronomical") azimuth and A' denotes the ellipsoidal ("geodetic") azimuth.

A measured horizontal angle may be considered as the difference between two azimuths:

$$\omega = A_2 - A_1,$$

so that we have

$$\begin{aligned} \omega = \omega' + \xi (\sin A_2 \cot z_2 - \sin A_1 \cot z_1) \\ + \eta (-\cos A_2 \cot z_2 + \cos A_1 \cot z_1), \end{aligned} \quad (6-11)$$

where ω' is the ellipsoidal horizontal angle, that is, the horizontal angle reduced to the reference ellipsoid, A_1 and z_1 are azimuth and zenith distance to target 1, and A_2 and z_2 are the same quantities for target 2.

Similarly we have for a measured zenith distance z (HM, p. 190):

$$z = z' - \xi \cos A - \eta \sin A, \quad (6-12)$$

where z' is the zenith distance reduced to the ellipsoid. (This is identical to (1-3) since $\epsilon = \xi \cos A + \eta \sin A$.)

In these expressions, the ellipsoidal quantities A' , ω' , z' represent the "systematic" part of the observations, pertaining to the reference ellipsoid. If we vary the ellipsoidal parameters p_i by δp_i , these quantities take indeed the form AX with (6-5). The terms containing as factors the deflections of the vertical, ξ and η , represent the signal s , that is, the effect of the anomalous gravity field on the quantities under consideration.

This might be symbolized as

$$A = A' + s_A, \quad (6-13)$$

with the "signal part of A " given by

$$s_A = \eta \tan \phi + (\xi \sin A - \eta \cos A) \cot z; \quad (6-14)$$

the quantities s_ω and s_z , the signal parts of ω and z , are to be understood accordingly.

These signal parts are thus nothing else than the quantities representing the reduction to the reference ellipsoid in the familiar sense (HM, secs. 5-4 and 5-5).

It is evident that the expressions (6-10), (6-11), and (6-12) presuppose a geocentric reference ellipsoid; otherwise ξ and η are to be replaced by ξ^a and η^a as given by (6-8), which introduces additional parameters δx_0 , δy_0 , δz_0 .

Thus, in principle, all measurements of A , ω and z give information, not only on the geometry through their ellipsoidal parts A' , ω' , and z' , but also on the gravity field through their signal parts s_A , s_ω , and s_z . Since s_ω is readily seen to be very small, the contribution of ω to the determination of the gravity field will in general be negligible, corresponding to the well-known fact that the reduction of ω to the ellipsoid can usually be neglected (HM, p. 189). On the other hand, s_z is significant, which is in agreement with the possibility of using zenith distances for determining deflections of the vertical (HM, p. 176).

Let us finally consider satellite observations. Take, for instance, photographic observations of right ascension α and declination δ , and electronic measurements of distances ℓ to the satellite. According to HM, p. 355 we have relations of the form

$$\begin{aligned}\alpha &= \alpha(x_p, y_p, z_p; t; a_0, e_0, i_0, \Omega_0, \omega_0, T_0; J_{nm}, K_{nm}) , \\ \delta &= \delta(x_p, y_p, z_p; t; a_0, e_0, i_0, \Omega_0, \omega_0, T_0; J_{nm}, K_{nm}) , \\ \ell &= \ell(x_p, y_p, z_p; t; a_0, e_0, i_0, \Omega_0, \omega_0, T_0; J_{nm}, K_{nm}) .\end{aligned}\tag{6-15}$$

The parameter vector X consists of corrections to the station coordinates x_p , y_p , z_p , to the time t , and to the orbital elements a_0 , e_0 , i_0 , Ω_0 , ω_0 , T_0 ; the influence of the reference gravity field is also implicitly contained and may be taken into account by suitable parameters.

The signal parts are represented by the effect of J_{nm} and K_{nm} , the coefficients of the expansion of the gravitational potential into spherical harmonics:

$$\begin{aligned}
 s_{\alpha} &= \sum_{m,n} \left(\frac{\partial \alpha}{\partial J_{nm}} \delta J_{nm} + \frac{\partial \alpha}{\partial K_{nm}} \delta K_{nm} \right), \\
 s_{\delta} &= \sum_{m,n} \left(\frac{\partial \delta}{\partial J_{nm}} \delta J_{nm} + \frac{\partial \delta}{\partial K_{nm}} \delta K_{nm} \right), \\
 s_{\ell} &= \sum_{m,n} \left(\frac{\partial \ell}{\partial J_{nm}} \delta J_{nm} + \frac{\partial \ell}{\partial K_{nm}} \delta K_{nm} \right),
 \end{aligned} \tag{6-16}$$

where

$$\delta J_{nm} = J_{nm} - J'_{nm}, \quad \delta K_{nm} = K_{nm} - K'_{nm} = K_{nm} \tag{6-17}$$

are the differences between the actual coefficients J_{nm} and K_{nm} and their normal values J'_{nm} and K'_{nm} referring to the reference gravity field; we have $J'_{nm} = 0$ for $m \neq 0$ and $K'_{nm} = 0$ throughout because of the rotational symmetry of the reference field.

On linearizing (6-15) and taking (6-16) into account we obtain expressions of the form

$$\begin{aligned}
 \alpha &= \alpha' + s_{\alpha}, \\
 \delta &= \delta' + s_{\delta}, \\
 \ell &= \ell' + s_{\ell},
 \end{aligned} \tag{6-18}$$

again falling into our usual model.

Doppler measurements, but also recently proposed observational schemes such as satellite altimetry, satellite-to-satellite tracking, or gradiometry (Moritz, 1971) may be treated in precisely the same way. The same holds, e.g., for the determination of zonal harmonics from variations of the orbital parameters (HM, sec. 9-6).

Systematic errors are taken into account by including them in the vector X ; and to provide for random errors, we add a term n to arrive again at the general model (2-1).

We are thus in a position to apply the basic equations for least-squares collocation. First, the parameters X are obtained from (2-35):

$$X = (A^T \bar{C}^{-1} A)^{-1} A^T \bar{C}^{-1} x ; \quad (6-19)$$

then any signal s_p will be given by (2-38):

$$s_p = C_p^T \bar{C}^{-1} (x - AX) . \quad (6-20)$$

Let us recall the meaning of these equations. The vector x comprises all observations of various types as we have just considered; the matrix \bar{C} is the covariance matrix of x . The signal s_p is an arbitrary quantity of the anomalous gravitational field, say a geoidal height at a certain point, a deflection of the vertical at 10 km elevation, or the spherical-harmonic coefficient K_{93} . Any field quantity may be obtained in this way by taking the appropriate covariance vector C_p .

As important limiting cases we have:

$$\text{Case 1: } X = 0 ,$$

$$\text{Case 2: } s = 0 .$$

The limiting case 1, that of no systematic parameters occurring or all such parameters being known, has been the subject of (Moritz, 1970a); here (6-20) reduces to

$$s_p = C_p^T \bar{C}^{-1} x . \quad (6-21)$$

The limiting case 2 corresponds to the absence of the anomalous gravity field. The earth is then considered as an equipotential ellipsoid. Since with $s = 0$ also the signal covariances C are zero, the matrix

$$\bar{C} = C + D \quad (6-22)$$

will consist only of the covariance matrix D of the measuring errors, so that in (6-19) \bar{C} is to be replaced by D :

$$x = (A^T D^{-1} A)^{-1} A^T D^{-1} x . \quad (6-23)$$

This corresponds to the result of a pure geometrical adjustment in the usual sense, the earth being identified with an ellipsoid.

In a way, our general method splits up the problem into two steps which are very similar to these limiting cases. The first step, the determination of the parameters by (6-19), corresponds to Case 2, with the error covariance matrix D replaced by the general covariance matrix \bar{C} , which incorporates also the signal covariances. The replacement of D by \bar{C} is the only effect of the anomalous gravity field on the determination of the parameters x (which describe, e.g., the geometry).

Then, using the X so obtained, the observations x may be centered by subtracting AX to obtain

$$z = x - AX .$$

As the second step, s_p is obtained by (6-20), which amounts to using the centered observations z instead of x in (6-21). In this way our problem reduces to Case 1, so that all methods described in (Moritz, 1970a) are now applicable to the observations centered by subtracting AX .

The accuracy of the estimated quantities X and s is expressed by (3-33), (3-36) and (3-37).

Concerning the properties of the solution the reader is referred to sec. 2. We only remark, in addition, that all calculated quantities s , whatever they are, refer to one and the same anomalous gravity field, so that our method is indeed consistent. This is a consequence of the invariance of the method with respect to linear transformations of the result; cf. also (Moritz, 1970a, sec. 3).

The signal field so obtained is optimal in the sense that the accuracy of all computed quantities is the highest obtainable on the basis of the given data.

This field is also the smoothest gravity field that is consistent with the given data; cf. sec. 10, where an appropriate definition of smoothness will be given. This fact is important because spurious irregularities (caused, e.g., by the instability of downward continuation) are avoided in this way.

Bjerhammar (1964) has formulated a basic problem of physical geodesy as the determination of a gravity field (or more generally the set of possible gravity fields) compatible with the given discrete measurements. If we generalize this "problem of Bjerhammar" to the determination of an optimal gravity field that is compatible with all given measurements (or different kinds), we may say that least-squares collocation solves this problem.

Here compatibility with the given measurements means that the resulting signal field is such that exact measurements are reproduced exactly (as, e.g., in the usual least-squares gravity prediction), whereas inaccurate data are adjusted in such a way that the effect of the measuring errors is removed as much as possible.

As we have mentioned above and shall consider in detail in the following section, all signal covariances can be derived from one basic covariance function $K(P,Q)$. Each choice of this function $K(P,Q)$ corresponds to a possible gravity field compatible with the given data; if we vary the assumed basic function $K(P,Q)$, we get a set of possible gravity fields.

This is reminiscent of finding the set of unbiased linear estimates in adjustment theory, which is most elegantly expressed in terms of generalized inverses of rectangular matrices (Bjerhammar, 1958, 1971a,b). In fact, there is a close conceptual analogy between these two problems. To further explore this analogy for our present problem would lead us beyond our present elementary approach.

Finally some remarks on the relation of least-squares collocation to conventional methods using formulas such as Stokes' and Vening Meinesz' integrals are in order. We shall illustrate the problems involved by two examples.

Let first all x_i be errorless measurements of gravity anomalies Δg at various points at sea level, and use (6-21) to compute Δg at every point at sea level; this is, then, a pure case of least-squares prediction of gravity in the usual sense. From the continuous global Δg -field obtained in this way, compute the anomalous potential T at some point at sea level by Stokes' formula. Alternatively, compute T directly from the measured values x_i , using again (6-21) with the appropriate covariances C_p . The resulting value for T will be the same in both cases because of the invariance with respect to linear transformations of the result (Stokes' formula expresses T as a linear (integral) transformation of Δg .)

Let now the coverage of the geoid by gravity observations x_i become denser and denser. Then the interpolated Δg -field tends to the true Δg -field, and the value of T ^{to that} obtained by applying Stokes' formula to the true Δg -field. Thus the classical approach may be considered as an appropriately defined limiting case of least-squares collocation.

As a second example, consider the problem of Bjerhammar just mentioned: the gravity anomalies are given at discrete points of the topographic earth's surface. As a limiting case, for continuous coverage of the whole earth's surface by gravity measurements, this problem reduces to the "problem of Molodensky", the well-known boundary-value problem of physical geodesy (HM, p. 291). The Bjerhammar problem may again be solved by (6-21) (Moritz, 1970a, sec. 5); if the gravity coverage becomes denser and denser, this solution tends to a solution of Molodensky's problem. As, under certain assumptions, the solution of Molodensky's problem is unique, this limiting solution will coincide with the usual solution of Molodensky's problem by means of integral formulas.

These and similar considerations show that the usual methods of physical geodesy can, in fact, be considered as limiting cases of least-squares collocation for idealized data distributions.

At first sight it seems to be difficult to believe that the simple matrix formula (6-20) is equivalent to complicated procedures such as the solution of Molodensky's problem, equivalent in the sense just outlined. The reason is that all covariances forming the vector C_p are derived, in a relatively simple way, from a covariance function $K(P,Q)$ which may be selected to have a relatively simple expression. In fact, starting from the covariance function of the potential, the covariances of all relevant quantities such as gravity anomalies, deflections of the vertical, or higher gradients are derived analytically by differentiations, which are much simpler to perform than the (often iterative) numerical integral operations occurring in the usual procedures of physical geodesy.

By taking for the covariance function a function that can be analytically continued down to sea level, difficulties in the analytical downward continuation will not arise; such difficulties beset conventional reduction procedures.

These considerations will make it clear why least-squares collocation is at the same ^{time} a generalization of classical procedures, so to speak with built-in interpolation and vertical reduction, and a conceptual and computational simplification--the latter at least as long as the occurring matrices can be handled by the computer.

7. COVARIANCES

In this section we shall consider the basic statistical prerequisites of the method considered in the preceding section, particularly the covariances which have been seen to play a central role in this method.

First we must define the average M . For local applications, it is an average over the area under consideration. Take, for instance, a plane rectangular area of sides $2a$ and $2b$. Referring this area to axes ξ and η , as we did in sec. 5 (Fig. 5-2), we may write for the average of the gravity anomaly Δg :

$$M\{\Delta g\} = \frac{1}{4ab} \int_{\xi=-a}^a \int_{\eta=-b}^b \Delta g(\xi, \eta) d\xi d\eta. \quad (7-1)$$

For global applications, with which we are mainly concerned, the average must be extended over the whole earth, that is, over the unit sphere σ :

$$\begin{aligned} M\{\Delta g\} &= \frac{1}{4\pi} \iint_{\sigma} \Delta g d\sigma \\ &= \frac{1}{4\pi} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \Delta g(\theta, \lambda) \sin\theta d\theta d\lambda, \end{aligned} \quad (7-2)$$

where θ (polar distance) and λ (longitude) are spherical coordinates.

The basic condition

$$M\{s\} = 0 \quad (7-3)$$

(cf. (2-4)) is thus equivalent to the absence of a zero-degree term in the spherical-harmonic expansion of the anomalous gravity field (HM, p.252).

The Covariance Function of the Gravity Anomaly. - The mathematical properties of the covariance function are discussed at length in HM, secs. 7-2 and 7-3. We shall first recall, without derivation, some of these properties which are particularly relevant for our present purpose.

The covariance of the gravity anomalies at two points P and Q ,

$$C(P,Q) = \text{cov}(\Delta g_P, \Delta g_Q) = M\{\Delta g_P \Delta g_Q\} , \quad (7-4)$$

is a function of the spherical distance ψ between P and Q :

$$C(P,Q) = C(\psi) . \quad (7-5)$$

This function $C(\psi)$ is the covariance function of the gravity anomaly. The average M is now defined as follows:

$$M\{\Delta g_P \Delta g_Q\} = \frac{1}{8\pi^2} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\alpha=0}^{2\pi} \Delta g(\theta, \lambda) \Delta g(\theta', \lambda') \cdot \sin \theta d\theta d\lambda d\alpha . \quad (7-6)$$

Here (θ, λ) denote the spherical coordinates of the point P , and (θ', λ') denote those of Q ; α represents the azimuth from P to Q (cf. Figure 7-2 in HM, p. 257); the coordinates (θ', λ') are related to ψ and α through

$$\cos \psi = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos (\lambda' - \lambda) , \quad (7-7)$$

$$\tan \alpha = \frac{\sin \theta' \sin (\lambda' - \lambda)}{\sin \theta \cos \theta' - \cos \theta \sin \theta' \cos (\lambda' - \lambda)} ;$$

and ψ is constant with respect to the integration.

The average M in (7-6) differs from the average in (7-2) by the fact that in (7-6) one also averages over the azimuth α to obtain dependence only on the spherical distance ψ . However, M as in (7-6) is in fact the general definition which includes the average in (7-2) as a special case: we have

$$\begin{aligned} \frac{1}{8\pi^2} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\alpha=0}^{2\pi} \Delta g(\theta, \lambda) d\theta d\lambda d\alpha &= \\ &= \frac{1}{2\pi} \int_{\alpha=0}^{2\pi} d\alpha \frac{1}{4\pi} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \Delta g(\theta, \lambda) d\theta d\lambda = \\ &= \frac{1}{4\pi} \iint_{\sigma} \Delta g d\sigma , \end{aligned} \quad (7-8)$$

identical to (7-2), since $\Delta g(\theta, \lambda)$ does not depend on α .

In the same way we see that the average M in (7-1) is a special case of the general definition as in (5-16) (we now have $r = 0$ as P and P' coincide).

It is obvious that the average over the sphere,

$$\frac{1}{4\pi} \iint_{\sigma} d\sigma = \frac{1}{4\pi} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \sin\theta d\theta d\lambda, \quad (7-9a)$$

expresses homogeneity, whereas the average with respect to α ,

$$\frac{1}{2\pi} \int_{\alpha=0}^{2\pi} d\alpha, \quad (7-9b)$$

expresses isotropy.

The function $C(\psi)$ may be expanded into a series of Legendre polynomials (zonal spherical harmonics):

$$C(\psi) = \sum_{n=0}^{\infty} c_n P_n(\cos\psi). \quad (7-10)$$

The coefficients c_n are given by

$$c_n = M\{\Delta g_n^2\} = \sum_{m=0}^n (\bar{a}_{nm}^2 + \bar{b}_{nm}^2), \quad (7-11)$$

where Δg_n is the Laplace harmonic of degree n in the expansion of Δg in spherical harmonics, and \bar{a}_{nm} and \bar{b}_{nm} are the fully normalized harmonic coefficients in this expansion:

$$\Delta g(\theta, \lambda) = \sum_{n=0}^{\infty} \Delta g_n(\theta, \lambda) \quad (7-12)$$

with

$$\Delta g_n(\theta, \lambda) = \sum_{m=0}^n (\bar{a}_{nm} \cos m\lambda + \bar{b}_{nm} \sin m\lambda) \bar{P}_{nm}(\cos \theta), \quad (7-13)$$

where $\bar{P}_{nm}(\cos \theta)$ is a fully normalized Legendre function (cf. HM, sec. 1-14).

From (7-11) it follows that all c_n must be positive. This ensures that the covariance function is positive definite as it should be; cf. (Krarup, 1969, p. 23).

The Covariance Function of the Potential. - The covariance function of the anomalous potential T will be denoted by $K(P, Q)$. It is also a function only of s or ψ if $C(P, Q)$ is, because Δg and T have the same properties of homogeneity and isotropy. Thus there are completely analogous formulas:

$$K(P, Q) = \text{cov}(T_P, T_Q) = M\{T_P T_Q\}, \quad (7-14)$$

$$K(P, Q) = K(\psi), \quad (7-15)$$

$$K(\psi) = \sum_{n=0}^{\infty} k_n P_n(\cos \psi), \quad (7-16)$$

$$k_n = M\{T_n^2\} = \sum_{m=0}^n (\bar{A}_{nm}^2 + \bar{B}_{nm}^2), \quad (7-17)$$

$$T(\theta, \lambda) = \sum_{n=0}^{\infty} T_n(\theta, \lambda), \quad (7-18)$$

$$T_n(\theta, \lambda) = \sum_{m=0}^n (\bar{A}_{nm} \cos m\lambda + \bar{B}_{nm} \sin m\lambda) \bar{P}_{nm}(\cos \theta). \quad (7-19)$$

The relation between the coefficients k_n of $K(\psi)$ and c_n of $C(\psi)$ is extremely simple. According to HM, p. 97 we have

$$T_n = \frac{R}{n-1} \Delta g_n . \quad (7-20)$$

Thus by (7-17),

$$k_n = M\{T_n^2\} = \frac{R^2}{(n-1)^2} M\{\Delta g_n^2\} , \quad (7-21)$$

and by (7-11)

$$k_n = \frac{R^2}{(n-1)^2} c_n , \quad (7-22)$$

which is the desired relation.

The covariance function of the anomalous potential is conceptually more fundamental and more convenient for derivative analytical computations than the covariance function of the gravity anomaly. The latter covariance function, however, can be determined more directly since the gravity anomaly is accessible to observation, rather than the potential. Thus it is convenient to determine $C(P,Q)$ first and to derive $K(P,Q)$ from it by expressions such as (7-22) or by the general formulas for covariance propagation to be considered below.

Extension into Space. - So far we have considered points P, Q on the sphere only. To extend the definition of $K(P,Q)$ to points outside the sphere, recall that the anomalous potential T is harmonic outside the sphere, that is, that it satisfies Laplace's equation:

$$\Delta T = 0 . \quad (7-23)$$

We now go back to the definition (7-14), where now P and Q are points in space:

$$K(P,Q) = M\{T_P T_Q\} . \quad (7-24)$$

Consider Q as fixed, so that $K(P,Q)$ is a function of P , and apply the Laplace operator (7-23):

$$\Delta_P K(P,Q) = M\{\Delta_P T_P T_Q\} = 0 ; \quad (7-25a)$$

we have written Δ_P to indicate that Δ is applied to the point P . In the same way,

$$\Delta_Q K(P,Q) = M\{T_P \Delta_Q T_Q\} = 0 . \quad (7-25b)$$

Thus the covariance function $K(P,Q)$ in space is a harmonic function both with respect to P and with respect to Q .

Now it is well known that the n th-degree spherical harmonic of a function harmonic outside a sphere depends on the radius vector r through $r^{-(n+1)}$. Thus $K(P,Q)$ in space must have the form

$$K(P,Q) = \sum_{n=0}^{\infty} \frac{\text{constants}}{r_P^{n+1} r_Q^{n+1}} P_n(\cos\psi) .$$

For $r_P = r_Q = R$, on the surface on the sphere, this expression must reduce to (7-16). This determines the constants. The result is

$$K(P,Q) = \sum_{n=0}^{\infty} k_n \left(\frac{R^2}{r_P r_Q} \right)^{n+1} P_n(\cos\psi) , \quad (7-26)$$

which expresses the spatial covariance function of the anomalous potential.

The extension of the covariance function $C(P, Q)$ of the gravity anomalies follows readily from the fact that $r\Delta g$ is a harmonic function in space (HM, p. 90). Thus the function

$$r_P r_Q C(P, Q) = M\{r_P \Delta g_P \cdot r_Q \Delta g_Q\}$$

is the covariance function of a harmonic function and must, therefore, again be harmonic both with respect to P and with respect to Q . This gives

$$r_P r_Q C(P, Q) = \sum_{n=\max(r_P, r_Q)}^{\infty} \frac{\text{constants}}{r_P^{n+1} r_Q^{n+1}} P_n(\cos\psi),$$

so that, on determining the constants by means of (7-10),

$$C(P, Q) = \sum_{n=0}^{\infty} c_n \left(\frac{R^2}{r_P r_Q} \right)^{n+2} P_n(\cos\psi). \quad (7-27)$$

This is the spatial covariance function of the gravity anomaly.

Practical Choice of the Covariance Function. - The covariance functions $C(P, Q)$ and $K(P, Q)$ being related through (7-22), one function is determined by the other.

The most obvious way of getting values for the coefficients c_n or k_n is using (7-17) with spherical harmonic coefficients determined from satellite observations or from a combination of satellite and gravimetric data. This is possible for, say, $n \leq 20$; the higher c_n

or k_n may be set equal to zero. The resulting covariance function will, however, be too smooth for most practical purposes as it contains only a limited number of spherical harmonics.

A better way of determining $C(\psi)$ is from observed gravity anomalies, using (7-6). This has first been done by Kaula (1959); see also HM, p. 254. To get in this way a representative covariance function, it should be determined from samples of gravity anomalies that are distributed as evenly as possible over the whole earth and over regions of different topography.

To make such an empirical function sufficiently well-defined for our present purposes, it is necessary to fit an analytical expression to the empirically obtained values.

In fact, the covariance matrices \bar{C} to be inverted in formulas such as (6-20) may be rather large and unstable, so that their elements must be computed with high numerical precision. Generally speaking it is extremely important that all computations be internally consistent, based on the chosen covariance function.

On the other hand, the choice of the covariance function itself is less critical (provided it is positive definite) because the results of least-squares collocation are not very sensitive with respect to the covariance function chosen, in the same way as the results of ordinary least-squares adjustment do not depend strongly on the weights.

Therefore, even simple analytical expressions may be of use. Krarup (1969, p. 62) suggests the function

$$K(P, Q) = B \left[\left(\frac{r_P r_Q}{r_o^2} \right)^2 - 2 \frac{r_P r_Q}{r_o^2} \cos \psi + 1 \right]^{-\frac{1}{2}} \quad (7-28)$$

with suitably chosen constants B and r_o . If r_o is slightly smaller than R , then such a function approximates, say, Kaula's covariance function locally well.

As Krarup has also remarked, one might combine satellite harmonics and empirical gravity anomaly covariances to fit them to an expression of the form

$$K(P, Q) = \sum_{n=0}^N A_n \left(\frac{R^2}{r_P r_Q} \right)^{n+1} P_n(\cos \psi) + B \left[\left(\frac{r_P r_Q}{r_o^2} \right)^2 - 2 \frac{r_P r_Q}{r_o^2} \cos \psi + 1 \right]^{-\frac{1}{2}} \quad (7-29)$$

with suitable coefficients A_n , B , and r_o .

Another possibility is to start from the spherical-harmonic expansion (7-26) and look for a function with simple coefficients k_n .

For example, we may take $k_o = k_1 = k_2 = 0$ and

$$k_n = \frac{A}{(n-1)(n-2)} \quad \text{for } n \geq 3. \quad (7-30)$$

With these coefficients, the series (7-26) may be summed so that a closed expression is obtained:

$$\begin{aligned} K(P, Q) = & A \left(\frac{r_o^2}{r_P r_Q} \right)^3 \left[P_2(\cos \psi) \left(1 + \ln \frac{2}{M} \right) + \frac{1}{4} \sin^2 \psi \right] - \\ & - A \left(\frac{r_o^2}{r_P r_Q} \right)^2 \cos \psi \ln \frac{2}{M} + \\ & + A \left(\frac{r_o^2}{r_P r_Q} \right) \frac{N}{2} \left[3 \left(\frac{r_o^2}{r_P r_Q} \right) \cos \psi - 1 \right], \end{aligned} \quad (7-31)$$

where

$$L = \left[1 - 2 \left(\frac{r_o^2}{r_P r_Q} \right) \cos \psi + \left(\frac{r_o^2}{r_P r_Q} \right)^2 \right]^{\frac{1}{2}},$$

$$M = 1 - L - \left(\frac{r_o^2}{r_P r_Q} \right) \cos \psi, \quad (7-32)$$

$$N = 1 + L - \left(\frac{r_o^2}{r_P r_Q} \right) \cos \psi,$$

and A and r_o are suitable constants. According to Lauritzen (1971), to whom this function is due, it fits excellently global gravity and satellite data, with

$$r_o = 0.9945 R, \quad (7-33)$$

$$A = 7.8488,$$

R denoting again a mean radius of the earth.

A remark on the gravity anomalies to be used in these computations is in order. As we are always concerned with the external gravity field, we are to use unreduced gravity anomalies, deflections of the vertical, etc., referring to the surface of the earth in the sense of Molodensky; cf. HM, chapter 8. This means that the covariance function $C(P, Q)$ is to be computed from free-air anomalies.

In principle, one could also use appropriately reduced gravity anomalies (in the sense of HM, secs. 8-2 or 8-11). This has even the advantage that some effects

of topography, which may have a quasisystematic character, are eliminated, especially when using isostatic anomalies (for global applications); as a matter of fact, the indirect effect has to be taken into account afterwards. In practical applications it must be examined whether the gain in accuracy justifies the enormous increase of computational work caused by gravity reduction.

Whereas this may appear doubtful in global applications, the use of Bouguer anomalies for local interpolation of gravity is well established. In this case, as in all local applications, it is particularly important that the gravity anomalies be appropriately centered by subtracting the mean anomaly for the area under consideration; this is necessary in order to have $M\{\Delta g\} = 0$ locally; then the covariance function is to be computed from these locally centered Bouguer anomalies.

Related to the use of Bouguer anomalies is the consideration of correlation with elevation; cf. HM, sec. 7-10. It would also be possible to include the (average local) rock density ρ as a parameter to be estimated by (6-19); other possibilities of using (6-19) and (6-20) for gravity interpolation, by splitting off "quasisystematic" effects, are discussed by Lauer (1971).

For global problems, one must use the same covariance function for the whole earth. In order to exclude in this case, as much as possible, systematic effects from the signal field, the spherical harmonics, not only of degree zero because of (7-3), but also of degrees one and two should be missing. This is possible by properly selecting the parameters of the reference ellipsoid (GM and a for degree zero and J_2 for degree two) and its position (geocentric, for degree one); cf. HM, sec. 5-11. Thus, in the above formulas, we should have

$$k_0 = k_1 = k_2 = c_0 = c_1 = c_2 = 0 ; \quad (7-34)$$

this condition is satisfied by the function (7-31).

The spherical theory underlying the definitions (7-2) and (7-6) and the developments based on them, correspond to the spherical approximation in which the flattening of the reference ellipsoid is neglected; cf. HM, pp. 87-88. This approximation seems to be sufficient in practically all cases.

For local problems, the sphere may even be formally replaced by a plane, and corresponding plane analytical expressions may be used for the covariance function; cf. (Moritz, 1970a, secs. 4 and 7; 1971, sec. 5).

Covariances of Derived Quantities. - All quantities of the anomalous gravity field can be derived from the anomalous potential T by linear operations such as differentiation or multiplication by a factor. For instance we have for the geoidal height N (HM, p. 85):

$$N = \frac{T}{G} , \quad (7-35)$$

for the gravity anomaly Δg (HM, p. 89):

$$\Delta g = - \frac{\partial T}{\partial r} - \frac{2}{R} T , \quad (7-36)$$

and for the components ξ and η of the deflection of the vertical (HM, p. 112):

$$\xi = \frac{1}{GR} \frac{\partial T}{\partial \theta} , \quad \eta = - \frac{1}{GR \sin \theta} \frac{\partial T}{\partial \lambda} , \quad (7-37)$$

where G is mean gravity and R is a mean earth radius.

Instead of using T as the basic function from which all other quantities are to be derived, we might also use Δg for the same purpose. For instance, N is then derived from Δg by Stokes' formula:

$$N = \frac{R}{4\pi G} \iint_{\sigma} \Delta g S(\psi) d\sigma \quad (7-38)$$

(HM, p. 94).

This formula may be written in the form

$$N_P = \frac{R}{4\pi G} \iint_{\sigma} \Delta g_A S(\psi_{PA}) d\sigma_A, \quad (7-38')$$

which makes explicit the computation point P and the variable point A carrying the surface element $d\sigma$. Eq. (7-38') may be abbreviated as

$$N_P = L_{PA} \Delta g_A, \quad (7-39)$$

where L_{PA} is the "linear operator" acting on Δg to transform it into N ; in the present case, L_{PA} simply denotes the application of Stokes' integral.

After these preliminaries, let us now consider the computation of the covariances of these derived quantities, starting from a basic covariance function. To fix our ideas, let T be the basic field function, to be denoted by f , and let its covariance $K(P, Q)$ be the basic covariance function.

Let now u and v be two quantities derived from f by

$$u_P = U_{PA} f_A, \quad v_Q = V_{QB} f_B. \quad (7-40)$$

Here U_{PA} and V_{QB} symbolize the linear operations by which u and v are derived from f ; for example, in (7-39) we have $U_{PA} = L_{PA}$, $f = \Delta g$.

By definition, the covariance between u and v is its average product

$$\text{cov}(u_P, v_Q) = M\{u_P v_Q\}, \quad (7-41)$$

since, as all signal quantities, u and v are supposed centered: $M\{u\} = M\{v\} = 0$. Substituting (7-40) we have

$$\begin{aligned} M\{u_P v_Q\} &= M\{U_{PA} f_A V_{QB} f_B\} \\ &= U_{PA} V_{QB} M\{f_A f_B\} \end{aligned} \quad (7-42)$$

(as to the interchangeability of M and the operators U and V cf. sec. 8), so that by (7-14) and (7-41) we obtain

$$\text{cov}(u_P, v_Q) = U_{PA} V_{QB} K(A, B) \quad (7-43)$$

as the desired result.

This equation may be called law of propagation of covariances because it is in principle nothing but an extension of the familiar law of error propagation, as generalized in (Moritz, 1961).

For applications, the law of covariance propagation is most conveniently expressed verbally as follows:

To the covariance function $K(A,B)$, considered as a function of B , apply the operation that determines the quantity v from T . To the result, considered as a function of A , apply the operation that determines the quantity u from T . The latter result is the covariance between u and v .

The meaning of this procedure will be illustrated by an example. Starting from $K(A,B)$, compute the covariance between Δg at a point P and N at a point Q ,

$$\text{cov}(\Delta g_P, N_Q) .$$

By (7-35) and (7-36) we have

$$\Delta g_P = U_{PA} T_A = - \left(\frac{\partial T}{\partial r} \right)_P - \frac{2}{R} T_P , \quad (7-44a)$$

$$N_Q = V_{QB} T_B = \frac{1}{G} T_Q , \quad (7-44b)$$

so that now $A = P$, $B = Q$.

Let us now apply our verbal rule for covariance propagation: To the covariance function $K(A,B) = K(P,Q)$, considered as a function of $B = Q$, apply the operation that determines N from T . By (7-44b) this is simply multiplication by $1/G$, so that the result is

$$\frac{1}{G} K(P,Q) .$$

To this quantity, considered as a function of $A = P$, apply the operation that determines Δg from T . By (7-44a) we find

$$- \frac{\partial}{\partial r_P} \left(\frac{1}{G} K(P, Q) \right) - \frac{2}{R} \left(\frac{1}{G} K(P, Q) \right),$$

so that the result is

$$\text{cov}(\Delta g_P, N_Q) = - \frac{1}{G} \frac{\partial K(P, Q)}{\partial r_P} - \frac{2}{GR} K(P, Q). \quad (7-45)$$

In this way we can derive all signal covariances from $K(P, Q)$. It is evident that, using Δg as the basic field function f , one can also derive all signal covariances from $C(P, Q)$, the covariance function of the gravity anomaly. Many other examples will be found in (Moritz, 1970a, 1971).

This gives all elements of the matrices C and C_P . If there are no measuring errors, then $\bar{C} = C$, so that we are ready to apply the collocation formulas such as (6-19) and (6-20) and the corresponding accuracy evaluations.

If there are measuring errors uncorrelated with the signal field, then we have again

$$\bar{C} = C + D,$$

so that the total covariance matrix \bar{C} is found by adding the signal covariance matrix C and the error covariance matrix D . The matrix C_P remains unchanged as a pure signal covariance matrix.

The required lack of correlation between noise and signal may usually be assumed to hold for direct measurements. It will no longer hold if the measurements have already been subjected to a preliminary least-squares collocation such as least-squares filtering; in this case the needed covariances may be obtained computationally; cf. (Moritz, 1969, sections 3, 4, and 9).

8. THE STATISTICAL BACKGROUND

The final three sections of this report deal with a closer study of statistical questions related to least-squares estimation of the anomalous gravity field. They are thus largely theoretical and may be omitted by readers interested only in practical applications.

The anomalous gravity field (the Δg -field or also the T-field , etc.) is often considered as a stochastic process on a sphere. The topic of stochastic processes is highly developed and widely applied and has a strong intuitive appeal (cf. Papoulis, 1965). On the other hand, their rigorous mathematical theory represents the most advanced and difficult branch of probability theory (cf. Doob, 1953).

Thus, while stochastic processes provide a convenient terminology and suggest the application of standard techniques to our present problem, they are less satisfactory as a logical basis for least-squares collocation: why should we put this simple theory on such a complex mathematical foundation?

This is not really necessary, however. In fact, more adequate, as a logical foundation for our present problem, than the theory of probability is the covariance analysis of individual functions, or "generalized harmonic analysis" in the terminology of Norbert Wiener (1930), who was the first to use this concept in a systematic and rigorous way; see also (Doob, 1949, sec. 1).

Primarily, probability does not enter at all into the consideration of the gravity field. In probability theory, we always have an ensemble of many possible realizations of the same quantity, but here we have only one anomalous gravity field. From the point of view of logical economy (the well-known "Occam's razor") it is thus preferable to develop the theory consistently in terms of this individual field, avoiding the introduction of a fictitious ensemble. This is achieved by using the covariance analysis of individual functions mentioned above, and this is what we did in the preceding section (and also, e.g., in sec. 5).

In fact, the averages M in (7-2) and (7-6) are not ensemble averages but spherical averages formed with one individual field function. (In time series terminology, they are not "phase averages" but "time averages".)

Thus all signal covariances are defined in a completely non-probabilistic way. Disregarding measuring errors, the same is true for the standard error of the predicted quantities. This is best seen in the well-known case of gravity prediction. Following the presentation in (Heiskanen and Moritz, 1967, sec. 7-6), we see that M in the expression (7-55), loc. cit., is the spherical average of one function ε_p^2 which is defined on the sphere and represents the square of the individual prediction error at any point. The same holds for the error covariance σ_{pq} .

Now, this definition of the average M is precisely what we need: we have only one gravity field but we are interested in the average accuracy over a certain area, or over the whole earth.

Derived Covariances. - If our definition of M as a spherical average is to be meaningful, it must be consistent with the law of covariance propagation as considered in the preceding section.

The issue at stake is the interchangeability of the average M with the operators U and V in (7-42). If M is considered as an ensemble average, then there is obviously no problem since M acts on the ensemble, whereas U and V act on the individual function. With the present definition of M as an individual spherical average, this interchangeability is not immediately obvious because now M , U and V act on the individual function.

The subsequent investigation will show that our present definition of M is indeed consistent with covariance propagation, but also that the problem is not entirely elementary.

Radial Derivatives. - As we have seen, all quantities of interest to physical geodesy are derived from the anomalous potential T essentially by partial differentiations.

Consider, first, the radial derivative $\partial T / \partial r$. Let us write again f for the basic function (in the present case, T) and $K(P, Q)$ for its covariance function.

$$K(P, Q) = M\{f_P f_Q\} . \quad (8-1)$$

The validity of the covariance propagation law is equivalent to the possibility of interchanging radial differentiation $\partial / \partial r_P$ (or $\partial / \partial r_Q$) with averaging M . Thus there should be

$$\frac{\partial K(P, Q)}{\partial r_P} = \frac{\partial}{\partial r_P} M\{f_P f_Q\} = M\left\{\frac{\partial f_P}{\partial r_P} f_Q\right\} , \quad (8-2)$$

and similarly for $\partial / \partial r_Q$.

The function f , being nothing else than the anomalous potential T , is a harmonic function in space. Let the point P have the spherical coordinates (θ, λ, r_p) , and similarly (θ', λ', r_Q) for Q . Then, in agreement with (7-6) we have

$$K(P, Q) = \frac{1}{8\pi^2} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\alpha=0}^{2\pi} f(\theta, \lambda, r_p) f(\theta', \lambda', r_Q) \cdot \sin\theta d\theta d\lambda d\alpha. \quad (8-3)$$

(In fact, (7-26) follows from (8-3) in the same way as (7-10) follows from (7-6); cf. (Heiskanen and Moritz, 1967, sec. 7-3).)

From the representation (8-3), however, it follows directly that (8-2) holds, since r_p enters into the definite integral (8-3) as a parameter, so that the rule for differentiating definite integrals with respect to a parameter can be applied. The same result obviously holds for $\partial/\partial r_Q$.

Now also the interchangeability for linear combinations of f and $\partial f/\partial r$ with constant coefficients follows; for instance, for the linear operation (7-44a) that transforms T into Δg .

Horizontal Derivatives. - This case is already less trivial. It is, however, significant indeed since, as (7-37) shows, the components of the deflection of the vertical are basically horizontal derivatives of T (derivatives along the surface of a sphere).

As a first step it will be convenient to introduce longitudinal and transversal horizontal derivatives. The longitudinal derivative, $\partial/\partial \ell$, is taken along the great circle connecting P and Q; the transversal derivative, $\partial/\partial m$, is taken at right angles to this line; both points, P and Q, are now supposed to lie on the same sphere. In

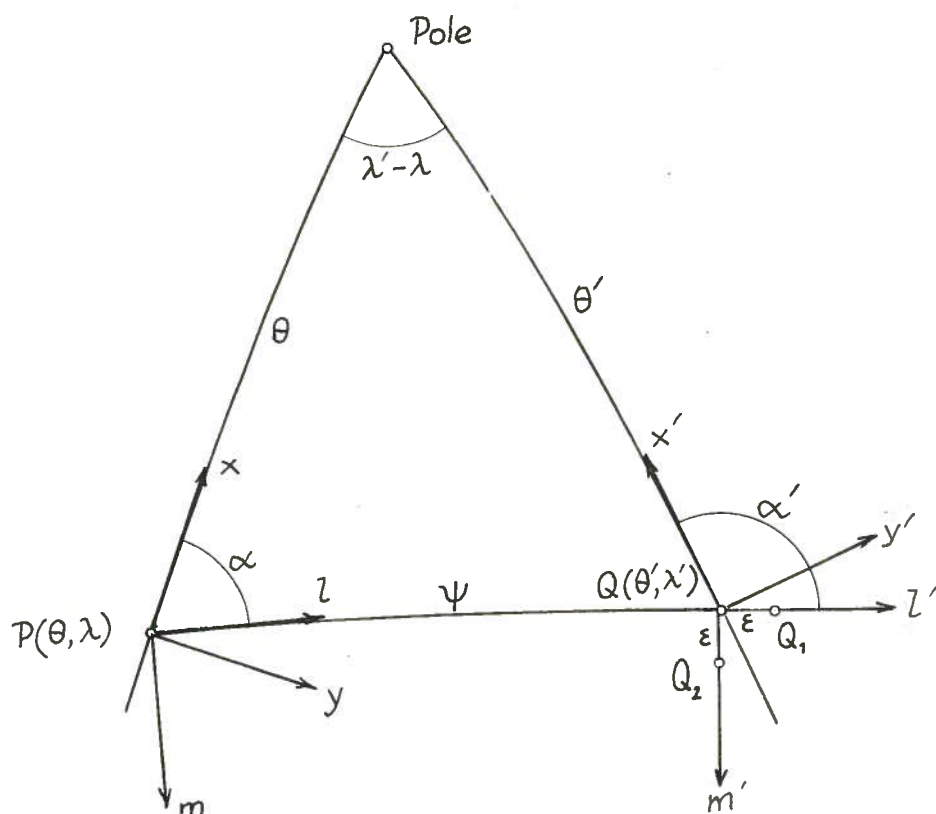


Figure 8-1

Fig. 8-1, the longitudinal and transversal directions at P are denoted by ℓ and m ; at Q, by ℓ' and m' . For simplicity, we put the radius of the sphere $R = 1$.

Again we have

$$K(P, Q) = M\{f_P f_Q\}, \quad (8-4)$$

but now, since P and Q refer to the same sphere $r_P = r_Q = 1$, we may simplify (8-3) as

$$K(P, Q) = \frac{1}{8\pi^2} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\alpha=0}^{2\pi} f(\theta, \lambda) f(\theta', \lambda') \cdot \sin\theta d\theta d\lambda d\alpha . \quad (8-5)$$

As usual,

$$\cos\psi = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\lambda' - \lambda), \quad (8-6)$$

and ψ is constant with respect to the integration.

For the validity of the covariance propagation law we should have, corresponding to (8-2),

$$\frac{\partial K}{\partial \lambda} = \frac{\partial}{\partial \lambda} M\{f_P f_Q\} = M\left\{\frac{\partial f_P}{\partial \lambda} f_Q\right\} , \quad (8-7a)$$

$$\frac{\partial K}{\partial \lambda'} = \frac{\partial}{\partial \lambda'} M\{f_P f_Q\} = M\left\{f_P \frac{\partial f_Q}{\partial \lambda'}\right\} , \quad (8-7b)$$

$$\frac{\partial K}{\partial m} = \frac{\partial}{\partial m} M\{f_P f_Q\} = M\left\{\frac{\partial f_P}{\partial m} f_Q\right\} , \quad (8-7c)$$

$$\frac{\partial K}{\partial m'} = \frac{\partial}{\partial m'} M\{f_P f_Q\} = M\left\{f_P \frac{\partial f_Q}{\partial m'}\right\} . \quad (8-7d)$$

To prove (8-7b), shift Q along the great circle PQ to the point Q_1 such that $QQ_1 = \epsilon$ (Fig. 8-1). Then

$$K(P, Q_1) = \frac{1}{8\pi^2} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\alpha=0}^{2\pi} f(\theta, \lambda) \left[f(\theta', \lambda') + \frac{\partial f}{\partial \lambda'} \epsilon \right] \cdot \sin \theta d\theta d\lambda d\alpha$$

to first order in ϵ . Subtracting (8-5), dividing by ϵ and letting $\epsilon \rightarrow 0$ we find

$$\begin{aligned} \frac{\partial K}{\partial \lambda'} &= \lim_{\epsilon \rightarrow 0} \frac{K(P, Q_1) - K(P, Q)}{\epsilon} \\ &= \frac{1}{8\pi^2} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\alpha=0}^{2\pi} f(\theta, \lambda) \frac{\partial f}{\partial \lambda'}(\theta', \lambda') \sin \theta d\theta d\lambda d\alpha, \end{aligned}$$

which is identical to (8-7b).

To prove (8-7d), shift Q in a transversal direction to the point Q_2 such that $QQ_2 = \epsilon$ (Fig. 8-1). Then

$$K(P, Q_2) = \frac{1}{8\pi^2} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\alpha=0}^{2\pi} f(\theta, \lambda) \left[f(\theta', \lambda') + \frac{\partial f}{\partial m'} \epsilon \right] \sin \theta d\theta d\lambda d\alpha,$$

again to first order in ϵ . Subtracting (8-5), dividing by ϵ and letting $\epsilon \rightarrow 0$ we find

$$\begin{aligned} \frac{\partial K}{\partial m'} &= \lim_{\epsilon \rightarrow 0} \frac{K(P, Q_2) - K(P, Q)}{\epsilon} \\ &= \frac{1}{8\pi^2} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\alpha=0}^{2\pi} f(\theta, \lambda) \frac{\partial f}{\partial m'}(\theta', \lambda') \sin \theta d\theta d\lambda d\alpha, \end{aligned}$$

identical to (8-7d).

To prove (8-7a) and (8-7c), it is sufficient to remark that we may interchange the role of P and Q as long as (8-6) is satisfied. Thus (8-5) may also be written as

$$K(P, Q) = \frac{1}{8\pi^2} \int_{\lambda'=0}^{2\pi} \int_{\theta'=0}^{\pi} \int_{\alpha'=0}^{2\pi} f(\theta, \lambda) f(\theta', \lambda') \cdot \\ \cdot \sin\theta' d\theta' d\lambda' d\alpha' . \quad (8-8)$$

Now we can apply to the point P the same reasoning as we just did to Q , to obtain (8-7a) and (8-7c).

In this way we see that horizontal derivatives in a longitudinal and a transversal direction can be interchanged with the averaging symbolized by M^1 ; above we have seen that this is also true for the radial derivatives. In the same way we can prove interchangeability also for the higher derivatives along these three orthogonal directions ℓ, m, r .

Since any (first or higher) derivative is a linear combination of the derivatives along ℓ, m, r (cf. eq. (8-12) below), we see that M can be interchanged with any operations by which geodetically significant quantities are derived from the anomalous potential T , so that our definition of M as a spherical average is indeed consistent with the law of covariance propagation. This is what we were to show.

Longitudinal and Transversal Covariances. - The covariances for derivatives along longitudinal and transversal directions have a particularly simple form. We have

$$C_{\ell\ell}(P, Q) = \text{cov}\left(\frac{\partial f_P}{\partial \ell}, \frac{\partial f_Q}{\partial \ell'}\right) = \frac{\partial^2 K}{\partial \ell \partial \ell'} ,$$

1)

The foregoing considerations are readily seen to be valid even if $r_Q \neq r_P$.

$$\begin{aligned}
C_{mm}(P,Q) &= \text{cov}\left(\frac{\partial f_P}{\partial m}, \frac{\partial f_Q}{\partial m'}\right) = \frac{\partial^2 K}{\partial m \partial m'} , \\
C_{\ell m}(P,Q) &= \text{cov}\left(\frac{\partial f_P}{\partial \ell}, \frac{\partial f_Q}{\partial m'}\right) = \frac{\partial^2 K}{\partial \ell \partial m'} , \\
C_{m\ell}(P,Q) &= \text{cov}\left(\frac{\partial f_P}{\partial m}, \frac{\partial f_Q}{\partial \ell'}\right) = \frac{\partial^2 K}{\partial m \partial \ell'} .
\end{aligned} \tag{8-9}$$

First we calculate, using (8-6),

$$\frac{\partial K}{\partial \theta} = \frac{dK}{d\psi} \frac{\partial \psi}{\partial \theta} = \frac{K'(\psi)}{\sin \psi} (\sin \theta \cos \theta' - \cos \theta \sin \theta' \cos(\lambda' - \lambda)) ,$$

$$\frac{\partial K}{\partial \lambda} = \frac{dK}{d\psi} \frac{\partial \psi}{\partial \lambda} = \frac{K'(\psi)}{\sin \psi} (-\sin \theta \sin \theta' \sin(\lambda' - \lambda)) ,$$

since, for $r_P = r_Q = 1$, K is a function only of ψ , $K(P,Q) = K(\psi)$; $K'(\psi)$ denotes $dK/d\psi$. Continuing in the same way we find $\partial^2 K / \partial \theta \partial \theta'$, $\partial^2 K / \partial \lambda \partial \lambda'$, $\partial^2 K / \partial \theta \partial \lambda'$, and $\partial^2 K / \partial \lambda \partial \theta'$, which is easy but leads to lengthy formulas which will not be reproduced here. Finally we assign to P and Q the following spherical

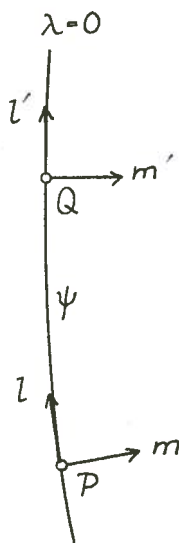


Figure 8-2

coordinates:

$$P(\theta=\theta, \lambda=0) , \quad Q(\theta'=\theta-\psi, \lambda'=0) ,$$

corresponding to Fig. 8-2, so that the points P and Q lie on the same meridian $\lambda = 0$. For this special position of P and Q we obtain simply

$$\frac{\partial^2 K}{\partial \theta \partial \theta'} = -K''(\psi) ,$$

$$\frac{\partial^2 K}{\partial \lambda \partial \lambda'} = -\frac{K'(\psi)}{\sin \psi} \sin \theta \sin \theta' ,$$

$$\frac{\partial^2 K}{\partial \theta \partial \lambda'} = \frac{\partial^2 K}{\partial \lambda \partial \theta} = 0 ,$$

with $K''(\psi) = d^2 K / d\psi^2$.

For this position of P and Q we have, by Fig. 8-2,

$$\frac{\partial^2 K}{\partial \ell \partial \ell'} = \frac{\partial^2 K}{\partial \theta \partial \theta'} ,$$

$$\frac{\partial^2 K}{\partial m \partial m'} = \frac{1}{\sin \theta \sin \theta'} \frac{\partial^2 K}{\partial \lambda \partial \lambda'} ,$$

since $d\ell = -d\theta$, $dm = \sin \theta d\lambda$; similar expressions hold for $\partial^2 K / \partial \ell \partial m'$ and $\partial^2 K / \partial m \partial \ell'$. We thus find for the expressions (8-9)

$$K_{\ell \ell}(\psi) = -K''(\psi) ,$$

$$K_{mm}(\psi) = -\frac{K'(\psi)}{\sin \psi} , \quad (8-10)$$

$$K_{\ell m}(\psi) = K_{m \ell}(\psi) = 0 ,$$

so that longitudinal and transversal derivatives are uncorrelated.

Because of the rotational symmetry, this result is valid for an arbitrary position of P and Q , independent of the special position used for the purpose of deriving it.

Derivatives along the Meridian and Parallel. - We may write (7-37) as

$$\xi = -\frac{1}{G} \frac{\partial T}{\partial x}, \quad \eta = -\frac{1}{G} \frac{\partial T}{\partial y}, \quad (8-11)$$

$\partial/\partial x$ denoting the derivative along the meridian, and $\partial/\partial y$ denoting the derivative along the parallel, cf. Fig. 8-1. The deflection components ξ and η are thus essentially derivatives of T along the meridian and the parallel.

With the notations of Fig. 8-1 we have

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{\partial f}{\partial \ell} \cos \alpha - \frac{\partial f}{\partial m} \sin \alpha, \\ \frac{\partial f}{\partial y} &= \frac{\partial f}{\partial \ell} \sin \alpha + \frac{\partial f}{\partial m} \cos \alpha, \end{aligned} \quad (8-12)$$

or briefly,

$$\begin{aligned} f_x &= f_\ell \cos \alpha - f_m \sin \alpha, \\ f_y &= f_\ell \sin \alpha + f_m \cos \alpha, \end{aligned} \quad (8-13)$$

and similarly at Q ,

$$\begin{aligned} f'_x &= f'_\ell \cos \alpha' - f'_m \sin \alpha', \\ f'_y &= f'_\ell \sin \alpha' + f'_m \cos \alpha'. \end{aligned} \quad (8-14)$$

Thus we compute

$$\begin{aligned}
 f_x f'_x &= (f_\ell \cos \alpha - f_m \sin \alpha) (f'_\ell \cos \alpha' - f'_m \sin \alpha') \\
 &= f_\ell f'_\ell \cos \alpha \cos \alpha' + f_m f'_m \sin \alpha \sin \alpha' - \\
 &\quad - f_\ell f'_m \cos \alpha \sin \alpha' - f_m f'_\ell \sin \alpha \cos \alpha' .
 \end{aligned}$$

Forming the average M we obtain the covariance

$$K_{xx} = M\{f_x f'_x\} = \text{cov} \left[\left(\frac{\partial f}{\partial x} \right)_P, \left(\frac{\partial f}{\partial x'} \right)_Q \right], \quad (8-15)$$

and similarly for the other covariances. The result is

$$\begin{aligned}
 K_{xx}(\psi, \alpha) &= K_{\ell\ell}(\psi) \cos \alpha \cos \alpha' + K_{mm}(\psi) \sin \alpha \sin \alpha', \\
 K_{xy}(\psi, \alpha) &= K_{\ell\ell}(\psi) \cos \alpha \sin \alpha' - K_{mm}(\psi) \sin \alpha \cos \alpha', \\
 K_{yx}(\psi, \alpha) &= K_{\ell\ell}(\psi) \sin \alpha \cos \alpha' - K_{mm}(\psi) \cos \alpha \sin \alpha', \\
 K_{yy}(\psi, \alpha) &= K_{\ell\ell}(\psi) \sin \alpha \sin \alpha' + K_{mm}(\psi) \cos \alpha \cos \alpha'.
 \end{aligned} \quad (8-16)$$

These covariances depend on ψ and α (α' is a function of α and ψ), whereas the covariances $K_{\ell\ell}$ and K_{mm} depend only on ψ and $K_{\ell m}$ and $K_{m\ell}$ are even zero. This shows again the significance of the longitudinal and transversal covariances.

One might wonder how a dependence on α is at all possible if the average is defined as in (7-6),

$$M\{\cdot\} = \frac{1}{8\pi^2} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\alpha=0}^{2\pi} (\cdot) \sin\theta d\theta d\lambda d\alpha, \quad (8-17)$$

since there is averaging over α . The reason is that the average is to be interpreted such that if Q moves along a circle of radius ψ with P as its center, then the xy -system is understood to rotate in such a way that the angle α between the x axis and the arc PQ remains always constant. This is then the angle α in (8-16), which is different from the integration variable α in (8-17).

Case of a Plane. - The plane may be considered as a limiting case of a sphere whose radius becomes infinite. For the plane we obviously have $\alpha' = \alpha$, and instead of the spherical distance ψ we take the plane distance r , so that (8-16) reduces to

$$\begin{aligned} K_{xx}(r, \alpha) &= K_{\ell\ell}(r) \cos^2 \alpha + K_{mm}(r) \sin^2 \alpha, \\ K_{xy}(r, \alpha) &= [K_{\ell\ell}(r) - K_{mm}(r)] \sin \alpha \cos \alpha = K_{yx}(r, \alpha), \\ K_{yy}(r, \alpha) &= K_{\ell\ell}(r) \sin^2 \alpha + K_{mm}(r) \cos^2 \alpha, \end{aligned} \quad (8-18)$$

which provides the covariance for the components of the plane vector

$$\vec{v} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = \text{grad } f, \quad (8-19)$$

which is nothing else than the gradient vector of the function f .

Even for a general vector field in the plane (which is not necessarily a gradient field), the covariances have the form (8-18), provided the vector field is homogeneous and isotropic (Grafarend, 1971a,b,c).

If the longitudinal and the transversal covariances are equal, then (8-18) reduces to

$$\begin{aligned} K_{xx}(r) &= K_{yy}(r) , \\ K_{xy}(r) &= K_{yx}(r) = 0 . \end{aligned} \tag{8-20}$$

This simple case has been used in sec. 5, cf. eqs. (5-22) and (5-23); here the average M was understood in such a way that the xy -system did not rotate during the process of averaging.

9. MEASURING ERRORS; STOCHASTIC

PROCESS INTERPRETATION

The statistical interpretation as put forward in the preceding section is particularly simple when no measuring errors are present. In this case, the average M was considered, not as an ensemble average, but as an average of individual functions over the sphere, in view of the fact that there is only one individual gravity field.

Incorporation of Measuring Errors. - It is not difficult also to incorporate measuring errors into this model. As far as measuring errors (the noise n) are concerned, the average M can only be interpreted as an ensemble average, a mathematical expectation in the usual probabilistic sense, since repeated measurements of the same quantity s (which as such remains the same) will be affected by different measuring errors.

Thus, as far as the signal is concerned, the proper interpretation is the average over the sphere (8-17), now to be denoted by M_1 :

$$M_1\{\cdot\} = \frac{1}{8\pi^2} \int_{\lambda=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\alpha=0}^{2\pi} (\cdot) \sin\theta d\theta d\lambda d\alpha, \quad (9-1)$$

whereas as far as the noise is concerned, the proper interpretation is the mathematical expectation (the ensemble average), to be denoted by M_2 .

Now, the natural way to define the average M in the general case (for both signal and noise) is a combination of M_1 and M_2 , symbolically

$$M = M_1 M_2 . \quad (9-2)$$

This means that the average of a quantity q is formed as

$$M\{q\} = M_1\{M_2\{q\}\} , \quad (9-3)$$

by first forming the ensemble average, M_2 , and then the average over the sphere, M_1 .

In fact, if q is either a signal s or a measuring error n , the average M reduces to its proper meaning, M_1 or M_2 , respectively. We have

$$M_2\{s\} = s ,$$

since s remains the same throughout repeated measurements, so that by (9-3),

$$M\{s\} = M_1\{M_2\{s\}\} = M_1\{s\} ;$$

and

$$M\{n\} = M_1\{M_2\{n\}\} = M_2\{n\} ,$$

since $M_2\{n\}$, not depending on the gravity field, is unaffected by the averaging M_1 if we assume homogeneity and isotropy (invariance with respect to rotations of the sphere). Thus

$$M\{s\} = M_1\{s\} , \quad M\{n\} = M_2\{n\} , \quad (9-4)$$

as it should be.

If the quantity q to be averaged is the product of a function $q_1(s)$ that depends only on the signal s and a function $q_2(n)$ that depends only on the noise n ,

$$q = q_1(s)q_2(n) , \quad (9-5)$$

then the definition (9-3) gives in the same way

$$\begin{aligned} M\{q\} &= M_1\{M_2\{q_1(s)q_2(n)\}\} = \\ &= M_1\{q_1(s)M_2\{q_2(n)\}\} = \\ &= M_2\{q_2(n)\}M_1\{q_1(s)\} . \end{aligned}$$

Thus

$$M\{q\} = M_1\{q_1(s)\}M_2\{q_2(n)\} \quad (9-6)$$

is simply the product of the two averages $M_1\{q_1(s)\}$ and $M_2\{q_2(n)\}$.

It is also evident that M is linear:

$$M\{c_1q_1 + c_2q_2\} = c_1M\{q_1\} + c_2M\{q_2\} \quad (9-7)$$

(with constant coefficients c_1 and c_2), in view of the linearity of both M_1 and M_2 .

Let us now verify that the present definition of M is consistent with the development in secs. 2 and 3. First, (2-5) is satisfied if (2-4) are:

$$\begin{aligned} M\{z\} &= M\{s' + n\} = M\{s'\} + M\{n\} \\ &= M_1\{s'\} + M_2\{n\} = 0 , \end{aligned}$$

in view of (2-4) and (9-4).

Consider now the covariances. We have

$$C_{s's'} = M\{s's'^T\} = M_1\{s's'^T\} = C ,$$

$$C_{nn} = M\{nn^T\} = M_2\{nn^T\} = D ,$$

$$C_{ns} = M\{ns^T\} = M_2\{n\}M_1\{s^T\} = 0 ,$$

$$C_{sn} = M\{sn^T\} = M_1\{s\}M_2\{n^T\} = 0 ,$$

as usual, and also

$$\begin{aligned} C_{xx} &= M\{zz^T\} = M\{(s'+n)(s'^T+n^T)\} = \\ &= M\{s's'^T + ns'^T + s'n^T + nn^T\} = C + D \end{aligned}$$

$$C_{sx} = M\{sz^T\} = M\{ss'^T + sn^T\} = C_{ss'} = \text{cov}(s, s') ,$$

$$C_{xs} = M\{zs^T\} = M\{s's^T + ns^T\} = C_{s's} = \text{cov}(s', s) ,$$

everything as in sec. 2.

Therefore we can also apply the average M , as defined by (9-2) or (9-3), to (3-16) and (3-17) to obtain (3-19) and (3-20), and everything proceeds as usual.

This shows that our present definition of M is indeed consistent with the developments in secs. 2 and 3.

Interpretation as a Stochastic Process. - Although we believe that the interpretation of the statistical background as given above and in the preceding section is the most natural and logically the simplest, it is often useful and convenient to use stochastic process terminology and results. It is thus desirable to find an interpretation of the anomalous gravity field as a stochastic process.

Recently Lauritzen (1971) proved an important negative result in this direction: it is impossible to find a stochastic process, harmonic outside the sphere, that is both Gaussian and ergodic.

Thus we have to give up either ergodicity or the Gaussian character of the process.

Ergodicity means that the individual average over the sphere (i.e., what we have denoted by M_1 in the preceding section) should be equal to the phase average (the average over the underlying probability space). Ergodicity is very important in our case since only the individual average is physically defined and can be determined empirically, since there is only one gravity field. Therefore, any appropriate stochastic process model should be such that the phase average, which is the probabilistically relevant definition of an average, should be equal to the individual average, which is the empirically relevant definition.

On the other hand, the Gaussian character of the process is less important, since it is well-known that least-squares estimation, as the linear unbiased estimation of least variance, is meaningful also for non-Gaussian (not normally distributed) random variables.

Therefore, we should look for an ergodic non-Gaussian stochastic process model for the anomalous gravity field.

Fortunately there exists such a model which is so simple as to be almost trivial. This model consists in identifying probability space Ω with "rotation group space", which is the space over which the integration in (9-1) is extended: $0 \leq \lambda < 2\pi$, $0 \leq \theta \leq \pi$, $0 \leq \alpha < 2\pi$; the name is justified since we may consider λ, θ, α as the three Eulerian angles by which a spatial rotation is described. The set of sample functions (the probabilistic ensemble) arises from one basic function f by the set of rotations of the sphere.

For instance, let f be the anomalous potential T ; then any sample function is again identical with the same function T , but referred to a different spherical coordinate system (different origin and different orientation). To every point of probability space Ω , i.e., to every triple of Eulerian angles, there corresponds a different sample function (because of the rotation of the coordinate system); all these different functions, however, are in reality nothing else but simple transformations of one and the same basic function, which again takes account of the uniqueness of the terrestrial gravity field.

It is clear that, in this case, the phase average is identical to the spherical average (9-1), so that our simple model is in fact "trivially ergodic" (Moritz, 1967).

This almost trivial stochastic process interpretation is obviously nothing else but a formulation, in different terms, of the interpretation in terms of a single function as considered in the preceding section. Therefore, the investigations in that section may also be interpreted as showing the consistency of the present stochastic process model.

If measuring errors are present, then the probability space for the whole system, comprising signal and noise, may be taken as the product space (Cartesian product) of rotation group space (the product space for the signal) and the probability space of the noise. This model provides the statistical independence of signal and noise as expressed, in our new interpretation, by (9-6).

As an alternative model, we might identify, by a measure-preserving mapping, the probability space of the noise with rotation group space, so that the latter provides a simple probabilistic background for the covariance analysis of both signal and noise.

10. THE NORM OF THE ANOMALOUS GRAVITY FIELD

The minimum principle underlying least-squares adjustment,

$$v^T P v = \text{minimum} , \quad (10-1)$$

may be interpreted as

$$\|v\| = \text{minimum} , \quad (10-2)$$

where $\|v\|$ is the "norm" of the vector v defined by

$$\|v\|^2 = v^T P v ; \quad (10-3)$$

$\|v\|$ may be considered as the measure of the "size" of the vector v .

Thus least-squares adjustment may be regarded as seeking a vector v of minimum norm compatible with the given data.

Let us try to generalize this simple idea to least-squares collocation of the terrestrial gravity field.

In sec. 2 we have used the minimum principle, analogous to (10-1),

$$v^T Q^{-1} v = \text{minimum} , \quad (10-4)$$

where the vector v , as given by (2-9), comprises the $p + q$ random variables s and z that enter into our problem, and where Q is the covariance matrix of the vector v .

Instead of v , we might use the vector

$$u = \begin{bmatrix} s^T & n^T \end{bmatrix}^T = \begin{bmatrix} s_1 & s_2 & \dots & s_p & z_1 & z_2 & \dots & z_q \end{bmatrix}^T. \quad (10-5)$$

This vector has again $p + q$ components; we shall assume that the number of signals to be computed is $p \geq q$ and that the first q signals to be computed coincide with the q components of the vector s' that corresponds to the observations x according to (2-1).

Then u is obviously a regular linear transformation of the vector v . Since least-squares collocation is invariant with respect to linear transformations, we may replace the condition (10-4) by

$$u^T K^{-1} u = \text{minimum}, \quad (10-6)$$

where K is the covariance matrix of the vector u :

$$K = \begin{bmatrix} C_{ss} & C_{sn} \\ C_{ns} & C_{nn} \end{bmatrix} = \begin{bmatrix} C & O \\ O & D \end{bmatrix}, \quad (10-7)$$

assuming that signal and noise are uncorrelated and using the notation (2-18a,b).

Then

$$K^{-1} = \begin{bmatrix} C^{-1} & O \\ O & D^{-1} \end{bmatrix}, \quad (10-8)$$

so that (10-6) becomes

$$s^T C^{-1} s + n^T D^{-1} n = \text{minimum}. \quad (10-9)$$

Since the minimum condition (10-9) is equivalent to (10-4), it will lead to the same solution (2-35) and (2-38). This may also be verified by direct computation, cf. (Moritz, 1970b).

Let us assume that the signal to be computed is the anomalous potential T at a number p of points distributed over the sphere. Let the number p of computation points increase indefinitely until these points cover densely the surface of the sphere. Then the quantity $s^T C^{-1} s$ may be shown to tend to a limit

$$\|s\|^2 = \lim_{p \rightarrow \infty} s^T C^{-1} s = \|T\|^2 ; \quad (10-10)$$

the quantity $\|s\|$ is called the norm of the function $s = T$, or the norm of the anomalous gravity field. In a sense, $\|s\|$ is a measure of the average size, or of the average smoothness, of the function s over the sphere.

Let us assume, for the moment, that there are no measuring errors, $n = 0$. Then, since the number p is arbitrary, the solution (2-35) and (2-38) will remain the same if $p \rightarrow \infty$; it thus satisfies the minimum principle, following from (10-9),

$$\|s\| = \text{minimum} , \quad (10-11)$$

that is, the resulting function $s = T$ has minimum norm.

If measuring errors are present, then we have instead

$$\|s\|^2 + \|n\|^2 = \text{minimum} , \quad (10-12)$$

where we have put

$$\|n\|^2 = n^T D^{-1} n . \quad (10-13)$$

The concept of norm, introduced in the present context by Krarup (1969), leads to interesting cross-connections to the theory of Hilbert spaces with kernel functions (Meschkowski, 1962).

If the estimate \tilde{T} of T has the spherical-harmonic expansion

$$\tilde{T}(\theta, \lambda) = \sum_{n=3}^{\infty} \sum_{m=0}^n (\bar{\alpha}_{nm} \cos m\lambda + \bar{\beta}_{nm} \sin m\lambda) \bar{P}_{nm}(\cos \theta) , \quad (10-14)$$

analogous to (7-18 and 19), then the norm $\|T\|$ is given according to (Krarup, 1969, p. 32) by

$$\|T\|^2 = \sum_{n=3}^{\infty} \frac{1}{k_n} \sum_{m=0}^n (\bar{\alpha}_{nm}^2 + \bar{\beta}_{nm}^2) , \quad (10-15)$$

with k_n as in (7-16).

Consider now the norm $\|\Delta g\|$ of the estimate $\tilde{\Delta g}$ of the gravity anomaly Δg . The transition from \tilde{T} to $\tilde{\Delta g}$ means multiplication of the coefficients $\bar{\alpha}_{nm}$ and $\bar{\beta}_{nm}$ by $(n-1)/R$, according to (7-20). In the same way, k_n is to be replaced by

$$c_n = \frac{(n-1)^2}{R^2} k_n .$$

Now in the expression for $\|\Delta g\|^2$ analogous to (10-15), the factor $(n-1)^2/R^2$ cancels, and there remains (10-15), or

$$\|\Delta g\| = \|T\| . \quad (10-16)$$

Thus the norm is the same whether we compute it for T or Δg .

In the same way we can show invariance of the norm under transition to other quantities of the gravity field. Thus $\|s\|$ really expresses an intrinsic property of the signal field, which justifies the name, norm of the anomalous gravity field.

Thus we have been justified in saying, in sec. 6, that least-squares collocation gives the smoothest gravity field that is consistent with the given data, in line with the maxim "when in doubt, smooth" (from a letter of Sir Harold Jeffreys).

Acknowledgment. - The author is indebted to Mr. T. Krarup and Mr. S.L. Lauritzen for clarifying discussions, and to Dr. K.P. Schwarz for programming and computing the numerical example in Section 4.

REFERENCES

- Bjerhammar, A. (1958) A generalized matrix algebra. Transactions, Royal Institute of Technology, Stockholm, No. 124.
- Bjerhammar, A. (1964) A new theory of geodetic gravity. Transactions, Royal Institute of Technology, Stockholm, No. 243.
- Bjerhammar, A. (1969) Linear prediction and filtering. Royal Institute of Technology, Division: Geodesy, Stockholm.
- Bjerhammar, A. (1971a) Linear unbiased estimation theory. Presented at XV General Assembly, IUGG, Moscow.
- Bjerhammar, A. (1971b) Estimation with singular inverses. Tellus, XXIII, No. 6.
- Doob, J.L. (1949) Time series and harmonic analysis. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman (ed.). Univ. of California Press, p. 303.
- Doob, J.L. (1953) Stochastic Processes. Wiley, New York.
- Gotthardt, E. (1968) Einführung in die Ausgleichungsrechnung. Wichmann, Karlsruhe.
- Grafarend, E. (1971a) Statistische Modelle zur Prädiktion von Lotabweichungen. Vermessungstechnik, 19, pp. 66-68.
- Grafarend, E. (1971b) A combined gravimetric-astrogeodetic method for telluroid and vertical deflection analysis. In: Beiträge aus der BRD zur Vorlage bei der XV. Generalversammlung der IUGG, Veröffentlichungen der Deutschen Geodätischen Kommission, B, No. 188, pp. 23-36.
- Grafarend, E. (1971c) Lotabweichungsverteilungen Westdeutschlands und ihre gruppentheoretische Struktur. Mitteilungen aus dem Institut für Theoretische Geodäsie der Universität Bonn, No. 1.
- Heiskanen, W.A., and H. Moritz (1967) Physical Geodesy. Freeman, San Francisco.
- Heitz, S. (1969) Eine astronomisch-geodätische Geoidbestimmung für Westdeutschland. Veröffentlichungen der Deutschen Geodätischen Kommission, B, No. 167.

- Hubeny, K. (1953) Isotherme Koordinatensysteme und konforme Abbildungen des Rotationsellipsoids. Sonderheft 13, Österreichische Zeitschrift für Vermessungswesen, Wien.
- Kaula, W.M. (1959) Statistical and harmonic analysis of gravity. J. Geophys. Research, 64(12), pp. 2401-2421.
- Kaula, W.M. (1963) Determination of the earth's gravitational field. Reviews of Geophysics, 1(4), pp. 507-551.
- Kaula, W.M. (1966) Global harmonic and statistical analysis of gravimetry. Gravity Anomalies: Unsurveyed Areas, Geophysical Monograph No. 9, Am. Geophys. Union, Washington, D.C., pp. 58-67.
- Kaula, W.M. (1967) Theory of statistical analysis of data distributed over a sphere. Reviews of Geophysics, 5(1), pp. 83-107.
- Krarpup, T. (1968) A framework for least-squares determination of the potential of the earth. Report to the members of IAG Special Study Group 5:31.
- Krarpup, T. (1969) A contribution to the mathematical foundation of physical geodesy. Publ. No. 44, Danish Geodetic Institute, Copenhagen.
- Krarpup, T. (1971) Introducing integrated geodesy. Lecture read at the Technical University of Berlin, May 1971.
- Kraus, K. (1972) Interpolation nach kleinsten Quadraten in der Photogrammetrie. Bildmessung und Luftbildwesen, 40(1), pp. 7-12.
- Lauer, S. (1971) Kovarianzen, Power-Spektren und Prädiktionen im lokalen Schwerfeld. In: Beiträge aus der BRD zur Vorlage bei der XV. Generalversammlung der IUGG, Veröffentlichungen der Deutschen Geodätischen Kommission, B, No. 188, pp. 37-49.
- Lauritzen, S. (1971) The probabilistic background of some statistical methods in physical geodesy. Danish Geodetic Institute, Copenhagen.
- Levallois, J.-J., ed. (1971) Geodetic Reference System 1967. Publication Spéciale du Bulletin Géodésique, Paris.
- Meschkowski, H. (1962) Hilbertsche Räume mit Kernfunktion. Springer, Berlin.
- Monget, J.M. (1969) Une nouvelle méthode d'analyse statistique des données gravimétriques. Bulletin Géodésique, No. 94, pp. 427-441.
- Monget, J.M., and M. Albuissou (1971) A new statistical treatment of gravity data. Bulletin Géodésique, No. 102, pp. 451-466.

- Moritz, H. (1961) Fehlertheorie der graphisch-mechanischen Integration; Grundzüge einer allgemeinen Fehlertheorie im Funktionenraum. Sonderheft 22, Österreichische Zeitschrift für Vermessungswesen, Wien.
- Moritz, H. (1962) Interpolation and prediction of gravity and their accuracy. Report No. 24, Inst. Geod. Phot. Cart., Ohio State Univ.
- Moritz, H. (1963) Interpolation and prediction of point gravity anomalies. Ann. Acad. Sci. Fenn., A, III, 69.
- Moritz, H. (1965) Schwerevorhersage und Ausgleichungsrechnung. Zeitschrift für Vermessungswesen, 90, pp. 181-184.
- Moritz, H. (1967) Statistical analysis of gravity anomalies. In: International Dictionary of Geophysics, Pergamon Press, Oxford, pp. 1437-1441.
- Moritz, H. (1969) A general theory of gravity processing. Report No. 122, Dept. of Geodetic Science, Ohio State Univ.
- Moritz, H. (1970a) Least-squares estimation in physical geodesy. Ibid., Report No. 130.
- Moritz, H. (1970b) Combination of satellite harmonics and gravimetry. Ibid., Report No. 146.
- Moritz, H. (1970c) A generalized least-squares model. Studia geophysica et geodaetica, 14, pp. 353-362.
- Moritz, H. (1970d) Die Methode der kleinsten Quadrate in der physikalischen Geodäsie. To appear in: Festschrift Karl Ledersteger.
- Moritz, H. (1971) Kinematical Geodesy II. Report No. 165, Dept. of Geodetic Science, Ohio State Univ.
- Papoulis, A. (1965) Probability, Random Variables, and Stochastic Processes. McGraw-Hill, New York.
- Parzen, E. (1961) An approach to time series analysis. Annals of Mathematical Statistics, 32, pp. 951-989.
- Rapp, R.H. (1966) A Fortran program for the computation of localized gravity anomaly statistics. Report No. 74, Dept. of Geodetic Science, Ohio State Univ.
- Rapp, R.H. (1971) A procedure for combining satellite and gravimetric data for position and gravity field determinations. J. Geophys. Research, 76(20), pp. 4986-4990.
- Schatz, U. (1970) Das Problem der optimalen Stützpunktdichte und der optimalen Maschengröße bei Transformationen ungleichartiger Koordinaten. Dissertation, Bonn University.

- Tienstra, J.M. (1956) Theory of the Adjustment of Normally Distributed Observations. Argus, Amsterdam.
- Tscherning, C.C. (1970) Determination of the geoid by collocation. Danish Geodetic Institute, Copenhagen.
- Wiener, N. (1930) Generalized harmonic analysis. Acta Math., 55, p. 117. (Reprint: MIT Paperback Series, No. 51, 1966.)
- Wolf, H. (1968) Ausgleichungsrechnung nach der Methode der kleinsten Quadrate. Dümmler, Bonn.
- Wolf, H. (1969) Prädiktion und Punktausgleichung. Zeitschrift für Vermessungswesen, 94, pp. 165-169.