GEODÆTISK INSTITUT
MEDDELELSE NO. 44

A CONTRIBUTION TO THE MATHEMATICAL FOUNDATION OF PHYSICAL GEODESY

BY

TORBEN KRARUP

KØBENHAVN · DANMARK
1969

Acknowledgements

A night in 1965 in the lobby of a small hotel at Uppsala Dr. Erik Tengström opened my eyes to some difficulties in the mathematical foundation of physical geodesy and encouraged me to attempt to settle some of these questions. (At that time I must have underestimated the lot of mathematical literature I should have to study before I could really attack the problems – else I should not have accepted the task). During the past year in his circular letters to the members of the special study group on mathematical methods in geodesy Professor Helmut Moritz has given a sharp analysis of related problems. Without the inspirations from these two exponents of our science and the personal discussions I have had with Professor Moritz in Berlin last spring this publication would never have seen the light of day.

I am also glad to have the opportunity to express my gratitude to my colleagues at the Danish Geodetic Institute for their kind patience in listening to me when I could not help speaking about my ideas.

Last but not least my thanks will be addressed to the director of the Institute, Professor Einar Andersen, D.Sc., for the favourable conditions under which I could study and work and for the confidence he has shown me in publishing my paper.

January, 1969.

PRINTED BY BIANCO LUNO A-S
COPENHAGEN

Contents

Exposition of the Ideas

Traditionally geodesists look on the problem of determination of the potential of the earth as a boundary value problem: in principle we can measure some local property of the potential – the gravity – at the surface of the earth and until recently only there. Seeing that today the artificial satellites make it possible to measure effects of the potential also in the outer space, and that, on the other hand, we shall never be able to make gravity measurements at more than a finite number of points, it would perhaps be more natural to look on the main problem of physical geodesy as one of interpolation. The present paper was primarily meant to be an attempt to solve some of the theoretical and computational problems connected with this changed point of view, but gradually, as my ideas took their form and I became aware of the difficulties, my impression that the mathematical foundation of physical geodesy had reached a state of crisis grew stronger and stronger.

I think that in reality the cause of this crisis is a crisis of communication, not communication between geodesists but between geodesy and mathematics, and that physical geodesy shares this difficulty with other sciences which are interested in applied mathematics. Looking around on the mathematical methods used, e.g. in physical geodesy, you will find ideas which seem to have reached us by accident from the closed land of mathematics rather than methods naturally suited for our problems.

I shall give a few examples which, I hope, will make it clear what I mean.

The methods used for determining the potential from satellite measurements and from gravity measurements on the earth are completely different, and although they should be able to complete one another – the satellite data being best suited to give the more global information and the gravity data to give the more local information – it is nevertheless very difficult to combine these two sets of data. It is curious that geodesists who are accustomed to use and perhaps misuse adjustment methods in geometrical geodesy have not been able to find an adjustment method that can combine these two sorts of data.

Some geodesists have suspected that Molodenskiy's boundary value problem has some form of unstability. Would it not have been natural then to try to formulate the boundary value problem as an adjustment

problem, i.e. to try to obtain an improvement of the boundary values so as to make the boundary value problem uniquely solvable and so that the improvement in some least-squares sense would be as small as possible? We have, however, stuck to the traditional formulation of boundary value problems, used and perhaps relevant in other parts of physical mathematics, but – at least in my opinion – not relevant here.

This crisis of communication is well understandable. The mathematical foundation of physical geodesy has been the classical potential theory as it was, say in the late thirties, which theory is well described in a few well-known books. After that time mathematics has developed fast and has in the opinion of many people become an esoteric science. If you want to know something about modern potential theory you will first learn that nothing is written today about potential theory, but that this subject is generalized to the theory of linear elliptic partial differential equations, and that it is preferable also to read something about linear partial differential equations in general. But when you see the books you realize that in order to understand anything in them you have first to read one or two books on functional analysis (and to understand these you must first read something about topology). Well, after a few years of study you may begin to suspect what a modern potential theory would look like and to see that there really are new things of importance to physical geodesy in it. But a book on modern potential theory for geodesists could – if it existed – be read in a few weeks; only such a book does not exist. I understand why modern books on mathematics are written as they are, but as a user of mathematics I must deplore it.

As a type of problem that would be treated in a modern book on potential theory I can mention the problems concerning approximation of potentials and in this connection stability problems.

I shall here give a few examples:

We have learned in the classical potential theory that two potentials which are identical in an open set are identical in their whole domain of regularity. The new potential theory gives a valuable complement to this theorem:

Let us consider potentials regular in some domain Ω, and consider another domain $\Omega_0 \in \Omega$. For two potentials φ_1 and φ_2 regular in Ω we assume that

$$|\varphi_1(P) - \varphi_2(P)| < \varepsilon \text{ for } P \in \Omega_0, \tag{1}$$

where ε is some positive number.

We must now distinguish between two different cases:

1) The boundary of Ω is a part of the boundary of Ω_0.

Then there exists a constant $\delta > 0$ so that

$$|\varphi_1(P) - \varphi_2(P)| < \delta \text{ for } P \in \Omega. \tag{2}$$

2) The boundary of Ω is not a part of the boundary of Ω_0.

Then there does not exist a constant δ so that (2) is satisfied.

From this important theorem many conclusions can be drawn, most of which are more or less known by geodesists, e.g.:

The upward continuation of the potential (or the gravity) of the earth is stable and the downward continuation is unstable.

By going to the limit in (1) we see that the Dirichlet problem is stable but the Cauchy problem is unstable.

The potential at the surface of the earth (or the figure of the earth) cannot be found from dynamical satellite measurements alone.

These conclusions are just expressed as slogans; the exact meaning can be found only on the basis of the theorem. Thus the words stable and unstable have been given an exact content by (1) and (2).

In a parenthesis I should like to say a few words more on the stability problem in another connection. We know from the classical potential theory that both Dirichlet's and Neumann's problems have one and only one solution and that this is also the case as regards the third boundary value problem, where

$$\frac{\partial \varphi}{\partial n} + h\varphi \tag{3}$$

is given at the boundary, if $h \leq 0$. But for certain functions h for which $h \leq 0$ does not hold, the values of (3) at the boundary cannot be arbitrarily prescribed, but must satisfy certain linear condition(s) for the problem to have a solution, which is now not unique (Fredholm's alternative). If h is not given exactly, but for example found by measurements, it may happen that it has no meaning to say that h is or is not one of the exceptional coefficient functions. If this is the case, the boundary value problem will be unstable.

And now back to the approximation problems.

As far as I can see, the most important point of view introduced in physical geodesy since the appearance of Molodenskiy's famous articles is Bjerhammar's idea of calculating an approximation of the potential by collocation, at the points where gravity anomalies have been measured,

of potentials that are regular down to a certain sphere situated inside the surface of the earth.

From the classical theory this idea looks very venturesome, because we know that the actual potential of the earth is not regular down to the Bjerhammar sphere. Nor does the evidence that Bjerhammar produces in support of his idea seem convincing to me at all. As we shall see later on from the point of view of the new potential theory, very much can be said in favour of the determination of the potential by interpolation methods as well as of the use of potentials that are regular outside a Bjerhammar sphere.

It is curious that from the very beginning these two things (I mean the interpolation and the Bjerhammar sphere) have appeared together, exactly as I was forced to accept the Bjerhammar sphere as soon as I drew the conclusions from Moritz' interpolation formula for the gravity anomalies.

Moritz' interpolation formula rests upon the idea of an auto-correlation[1]) function for the gravity anomalies which is invariant under the group of rotations around the centre of the earth. In the first chapter I shall prove that from the existence of such an auto-correlation function on a sphere follows the existence of an auto-correlation function for the potential in the whole space outside the surface of the sphere and also that this auto-correlation function is invariant under the group of rotations. It appears from some af Kaula's articles that he too was aware of that.

Now it is an obvious idea to generalize Moritz' interpolation formula so as to find the potential directly by collocation (or generalized interpolation) from the gravity anomalies and thus drop the complicated integration procedure using Stoke's formula or a related formula. It is not less obvious that this method would not be confined to the use of gravity anomaly data, but that for example data related to satellites or vertical deflections could be included as well. In fact, I have derived such a formula in the first chapter and, more than that, I have also generalized this formula to make it possible, not by exact collocation but by the aid of a given variance-covariance matrix for the measurements, to find a potential which corresponds to improved measurements and which is more likely according to the auto-correlation hypothesis. This is what I call a smoothing procedure.

As Moritz has pointed out in an article [11] which is as amusing as it is interesting, there is a close relation between his form of interpolation

[1]) The word "covariance" is used in this sense in the geodetic literature in the same way as it will be used in this paper from the first chapter and on.

and the classical least-squares adjustment. Therefore, I could rededuce the results described in the first chapter from the point of view of adjustment described in the second chapter. The classical reciprocity between variance-covariance and weight corresponds here to a reciprocity between auto-correlation and norm metrics. This recalls the beautiful geometric method of deducing the formula for classical adjustment where the weight defines an Euclidean norm in the vector space of the measurements.

The technical difficulty here is that the vector space in which the adjustment has to be made is not a finite dimensional one, but, when the metric has been defined, a Hilbert space.

The Hilbert space should be within the natural sphere of interest of geodesists, since it is the most general substratum for the least-squares adjustment and also an indispensable tool af modern potential theory.

But the Hilbert space is not enough. We must use a concept that exists in Hilbert spaces consisting of sufficiently well-behaved functions – and regular potentials are well-behaved in most respects. This is the concept of a reproducing kernel.

I have not found it possible in this paper to give an introduction to Hilbert spaces with a reproducing kernel, but good books exist on that subject. The book most relevant to the use I have made of Hilbert spaces with reproducing kernel is [9] which exists only in German, but the beautiful book [1], which contains little but enough about the subject and which contains many other interesting ideas not totally irrelevant to our problems, is also useful. On the other hand, I guess that most physical geodesists will be able to read this paper the first time with some profit without having ever seen the words "Hilbert spaces" and "reproducing kernel" together before.

The advantage we get by introducing the adjustment aspect is not only that we attain a closer connection to classical adjustment methods, which is only a formal gain, but also – and this is more essential – that we obtain a better insight in the set of potentials among which the solution is found. This set is exactly those potentials for which the norm (defining the metrics) is defined and finite, and – here it is again – all these potentials are regular outside a certain sphere in the interior of the earth: the Bjerhammar sphere. Again: the coupling between interpolation and the Bjerhammar sphere.

And now we can go back to Moritz' formula. (I mean, here as before, Moritz' formula without height correction). It presupposes the rotation invariance of the auto-correlation function, from which follows the

rotation invariance of the auto-correlation function for the potential and again the rotation invariance for the domain of definition of the potential itself, independent of the way it is calculated: from Moritz' formula and Stoke's formula, or directly from the generalized Moritz' formula. Again: the Bjerhammar sphere.

There is a close connection between the problem of the Bjerhammar sphere and the problem of the convergence of series in spherical harmonics or better the question of approximation of potentials by series in spherical harmonics, and here I believe we have reached the very core of the foundation of physical geodesy.

The consequence must be that not only the first two chapters of this paper but also very much of the work done by physical geodesists all over the world remain idle formalism until this approximation problem has been solved.

This is what I have tried to do in the third chapter. Here it appears that again the theory of Hilbert spaces with reproducing kernel is a valuable tool and that the central theorem – the Runge theorem – is a special case of a theorem, known as a theorem of the Runge type, which is well-known in the theory of linear partial differential equations and which concerns the approximation of solutions to a partial differential equation in some domain by solutions to the same equation in another domain containing the first one. The proof of Runge's theorem given in the appendix is a reduction and specialization of the theorem 3.4.3 in [5].

As far as I can see, all the problems of approximations of potentials relevant to physical geodesy are solved or can be solved by the methods used in the third chapter and the appendix. And it is curious that the investigations there *seem* to show that the adjustment method described in chapter II (or, equivalently, the interpolation method from chapter I) is the natural way to find the approximation, but the cause of this *may* be that I do not know about any other way to obtain an approximation to the potential of the earth which can be proved relevant.

The last chapter is dedicated to the applications of the method of approximation treated in the first chapters. I fear that most readers will feel disappointed by reading the last chapter as I myself felt disappointed when writing it. In the light of the many thoughts and feelings I have had as to the possibilities of application of the method it seems very poor to me. But in the less than 18 months elapsed from the first idea of the method origined until the paper was sent to the printers I have had to concentrate so much on the foundation and the studies of the relevant mathematical literature, that the physical or geodetic aspects could only be vaguely glanced at. However, I feel that now the time is ripe for a discussion of this matter and that especially the question relating to the geodetic applications is best furthered by discussion and collaboration, and so I hope that the few hints in chapter IV will be sufficient to start such a discussion.

Forse altri canterà con miglior plettro!

Chapter I

Covariance and Collocation

1. As starting point I shall take H. Moritz' least-squares prediction formula [4, (7–63) p. 268], and I assume that the reader is familiar with the statistical reasoning leading to this formula. To ensure a better understanding of what follows I shall make a few comments on the mathematical reasoning using Moritz' notation. The problem is to find the coefficients α_{Pi} in the linear prediction formula

$$\tilde{\Delta g}_P = \sum_{i=1}^{n} \alpha_{Pi} \Delta g_i, \tag{0}$$

so that the mean square error m_P of the predicted anomaly at P attains a minimum. We have

$$m_P^2 = C_{PP} - 2 \sum_{i=1}^{n} \alpha_{Pi} C_{Pi} + \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_{Pi} \alpha_{Pk} C_{ik}, \tag{1}$$

and the α_{Pi} are to be determined so that (1) attains a minimum.

With

$$\alpha_{Pi} = \beta_{Pi} + \gamma_{Pi} \tag{2}$$

(1) becomes

$$\left.\begin{aligned} m_P^2 = C_{PP} &- 2 \sum_{i=1}^{n} \beta_{Pi} C_{Pi} + \sum_{i=1}^{n} \sum_{k=1}^{n} \beta_{Pi} \beta_{Pk} C_{ik} \\ &+ 2 \sum_{i=1}^{n} \gamma_{Pi} \Big(\sum_{k=1}^{n} \beta_{Pk} C_{ik} - C_{Pi} \Big) \\ &+ \sum_{i=1}^{n} \sum_{k=1}^{n} \gamma_{Pi} \gamma_{Pk} C_{ik}. \end{aligned}\right\} \tag{3}$$

We may here choose the β_{Pi} as we like, and we shall choose them so that

$$\sum_{k=1}^{n} \beta_{Pk} C_{ik} = C_{Pi}, \tag{4}$$

which is possible if $\{C_{ik}\}$ is non-singular. Then the second line in formula (3) is zero, and the first line has a fixed value independent of γ_{Pi}.

If now $\{C_{ik}\}$ is a positive definite matrix, then the third line of (3) is ≥ 0 and is $= 0$ only if $\gamma_{Pi} = 0$, i. e. (1) attains its minimum for α_{Pi} satisfying the normal equations

$$\sum_{k=1}^{n} \alpha_{Pk} C_{ik} = C_{Pi}. \tag{5}$$

If, on the other hand, $\{C_{ik}\}$ is not definite, then the third line may attain any positive or negative value, and m_P^2 cannot have any minimum.

Thus we see that for the least-squares problem to have a meaning the covariance function $C(P, Q)$[1]) must be of positive type, i. e.

$$\sum_{i=0}^{n} \sum_{k=0}^{n} C(P_i, P_k) X_i X_k > 0, \tag{6}$$

for all sets of n points P_i and n corresponding numbers $X_i \neq 0$, for $i = 1, 2, \ldots n$, for all natural numbers n. It is evident, that only in this case has $C(P, Q)$ a meaning as a covariance function.

The α_{Pi} having been found by (5), the predicted value can be found by (0) or by

$$\tilde{\Delta g}_P = \sum_{i=1}^{n} \sum_{k=1}^{n} C_{ik}^{(-1)} C_{Pi} \Delta g_k. \tag{7}$$

If predicted values for several points are to be calculated, it is most economical first to solve the following set of "normal equations"

$$\sum_{i=1}^{n} \xi_i C_{ik} = \Delta g_k, \tag{8}$$

where ξ_i are independent of P. Then the set of solutions to (8) can be used to find the predicted values for all points by

$$\tilde{\Delta g}_P = \sum_{i=1}^{n} \xi_i C_{Pi}. \tag{9}$$

[1]) I regard t he covariance function $C(P, Q)$ as a symmetric function of the two points P and Q and not merely as a function of the distance PQ. This will be of importance in the following part of the paper. It should perhaps be noted that C_{ik} means $C(P_i, P_k)$ and C_{Pi} means $C(P, P_i)$.

It is evident that (7) follows again from (8) and (9), but (8) and (9) formulate a problem of interpolation and its solution:

(9): Find $\tilde{\Delta g}_P$ in the whole region as a linear combination of the functions $C(P, P_i)$ of P for $i = 1, 2, \ldots, n$, so that

(8): $\tilde{\Delta g}_P = \Delta g_k$ for $k = 1, 2, \ldots, n$.

On the assumption that $C(P, Q)$ is of positive type this problem has always one and only one solution.

Having regarded the method from this slightly changed point of view one gets the courage to try to generalize.

Let us jump into the new problem:

By first using Moritz' prediction formula with a given covariance function $C(P, Q)$ we can find $\tilde{\Delta g}_P$ for all points of some surface. Solution of the corresponding boundary value problem will then give us the disturbing potential T. But can we, by using another covariance function $K(P, Q)$ defined in a domain including the outer space of the earth and describing the covariance between values of T at the points P and Q, find the potential T directly from the measured values of Δg_i?

Or to put the question in another way: Can we find the disturbing potential T by collocation (i. e. so that the corresponding gravity anomalies attain a given set of values at given points) so that T corresponds to the interpolated anomalies that could be found by Moritz' formula?

I shall prove that this is not only possible but also relatively simple.

Evidently we cannot find the covariance function directly, for to do so we should have a population of earths similar to our own and have the opportunity to measure any T for all these earths at all points. But let us for a moment assume that by a miracle we were given such a function; how then could we use it and what can we say a priori about its properties?

First of all we see that the whole theory about Moritz' prediction formula may be used unchanged. Given the values of the potential T at n points P_i, T can be predicted for all points P in the domain Ω where K is valid. The form of (9) will here be

$$\tilde{T}_P = \sum_{i=1}^{n} \xi_i K(P, P_i) \tag{10}$$

where ξ_i is found from the normal equations

$$\sum_{k=1}^{n} K(P_k,P_i)\,\xi_k = T_P. \tag{11}$$

Using the Laplace operator Δ on both sides of (10) for $n = 1$ we find that

a) for all P_i $K(P,P_i)$ satisfies the Laplace equation

$$\Delta_P K(P,P_i) = 0$$

at all points $P \in \Omega$ and is a regular potential at infinity as a function of P.

The last remark follows from T having the corresponding property. (By regularity at infinity for a potential φ I mean in this paper that 1) $\lim \varphi = 0$ and 2) $\lim r \cdot \varphi$ exists. It is evident that T has at least these two properties.)

From the definition of $K(P, Q)$ as a covariance function follows

b) $K(P, Q) = K(Q,P)$.

Then it is trivial that, for all P_i, $K(P_i,P)$ as a function of P is a regular potential in Ω.

c) $K(P, Q)$ is a function of positive type.

This follows as in the theory explaining Moritz' formula.

As a harmonic function $K(P, Q)$ must be arbitrarily often differentiable with respect to the coordinates of the points P and Q in the domain Ω. Now we shall investigate what can be deduced from a given $K(P, Q)$ about the covariance between derivatives of T.

From the definition of $K(P, Q)$ as a covariance function we have

$$M\{T_P, T_Q\} = K(P, Q). \tag{12}$$

For P, Q, R being three points in Ω follows

$$M\{T_P, T_R - T_Q\} = K(P,R) - K(P, Q). \tag{13}$$

Let the distance from Q to R be l, so that for some unity vector $\bar{e}$

$$R = Q + l\bar{e}; \tag{14}$$

then (13) gives

$$M\left\{T_P, \frac{T_Q + l\bar{e} - T_Q}{l}\right\} = \frac{K(P, Q + l\bar{e}) - K(P, Q)}{l}, \tag{15}$$

and for $1 \to 0$

$$M\left\{T_P, \left(\frac{\partial T}{\partial l}\right)_Q\right\} = \left(\frac{\partial}{\partial l} K(P, Q)\right)_Q, \tag{16}$$

where $\left(\frac{\partial f}{\partial l}\right)_Q$ means the derivative of the function f in the direction $\bar{e}$ taken at the point Q.

Evidently (16) can be generalized to arbitrarily high derivatives in arbitrary directions and to all linear differential operators. Specially, we get

$$0 = M\{T_P, \Delta T_Q\} = \Delta_Q K(P, Q), \tag{17}$$

i. e. again the result that $K(P, Q)$ is harmonic as a function of Q.

The normal potential U being known, Δg can be found from T by a differential operator, let it be called $\mathscr{L}$, so that we get

$$M\{\Delta g_P, \Delta g_Q\} = \mathscr{L}_P \mathscr{L}_Q K(P, Q). \tag{18}$$

That is to say that if on a surface ω bounding Ω we know the covariance function $C(P, Q)$ for the gravity anomaly then we must have

d) $\mathscr{L}_P \mathscr{L}_Q K(P, Q) = C(P, Q)$ for $P, Q \in \omega$,

and now it follows from a), b) and d) that $K(P, Q)$ can be found from $C(P, Q)$ under certain conditions.

2. I hope that the reader now looks upon the possibility of finding $K(P, Q)$ with so much optimism that he is interested in learning how to use the covariance function.

I shall try to show it, using the same argumentation as that used in [4, p. 266ff.].

We assume that we have found the anomalies Δg_i at n points

$$P_i = Q_i \qquad i = 1, 2, \ldots, n,$$

and that we use a linear prediction for the potential T at the point P.

$$\tilde{T}_P = \sum_{i=1}^{n} \alpha_{Pi} \Delta g_i. \tag{19}$$

If the correct value for T at P is T_P, the error of prediction ε_P is

$$\varepsilon_P = T_P - \tilde{T}_P = T_P - \sum_{i=1}^{n} \alpha_{Pi} \Delta g_i. \tag{20}$$

Then we can find the error covariance

$$\sigma_{PQ} = M\{\varepsilon_P \varepsilon_Q\}. \tag{21}$$

$$\left.\begin{aligned}\sigma_{PQ} &= M\left\{(T_P - \sum_i \alpha_{Pi}\Delta g_i)(T_Q - \sum_k \alpha_{Qk}\Delta g_k)\right\} \\ &= \{M(T_P T_Q - \sum_i \alpha_{Pi} T_Q \Delta g_i - \sum_k \alpha_{Qk} T_P \Delta g_k \\ &\quad + \sum_i \sum_k \alpha_{Pi}\alpha_{Qk}\Delta g_i \Delta g_k \\ &= K(P,Q) - \sum_{i=1}^n \alpha_{Pi}\mathscr{L}_{P_i}K(Q,P) - \sum_{k=1}^n \alpha_{Qk}\mathscr{L}_{Q_k}K(P,Q) \\ &\quad + \sum_{i=1}^n \sum_{k=1}^n \alpha_{Pi}\alpha_{Qk}\mathscr{L}_{P_i}\mathscr{L}_{Q_k}K(P,Q).\end{aligned}\right\} \tag{22}$$

For $P = Q$ this becomes

$$\varepsilon_P^2 = K(P,P) - 2\sum_{i=1}^n \alpha_{Pi}\mathscr{L}_{Q_i}K(P,Q) + \sum_{i=1}^n \sum_{k=1}^n \alpha_{Pi}\alpha_{Pk}\mathscr{L}_{P_i}\mathscr{L}_{Q_k}K(P,Q). \tag{23}$$

If ε_P^2 is to attain a minimum, α_{Pi} must satisfy the normal equations

$$\sum_{k=1}^n \alpha_{Pk}\mathscr{L}_{P_k}\mathscr{L}_{Q_i}K(P,Q) = \mathscr{L}_{Q_i}K(P,Q). \tag{24}$$

If the points $P_i \in \omega$, then the coefficients of the normal equations are $C(P_i,P_k)$, and the matrix is positive definite so that we can find the inverse matrix N_{ik} defined by

$$\sum_{j=1}^n N_{ij}\mathscr{L}_{P_j}\mathscr{L}_{Q_k}K(P,Q) = \delta_{ik}. \tag{25}$$

By the aid of N_{ik} the potential can be predicted for any point P in Ω by the formula

$$\tilde{T}_P = \sum_{i=1}^n \sum_{k=1}^n N_{ik}\mathscr{L}_{Q_i}K(P,Q)\Delta g_k. \tag{26}$$

The error covariance for the points P and Q is found by

$$\sigma_{PQ} = K(P,Q) - \sum_{i=1}^n \sum_{k=1}^n N_{ik}\mathscr{L}_P K(P,Q_i)\mathscr{L}_Q K(Q,P_k). \tag{27}$$

Also here we can turn the question in another way as in the case of formula (7):

We want to represent $\tilde{T}_P$ as follows

$$\tilde{T}_P = \sum_{i=1}^n \xi_i \mathscr{L}_{Q_i}K(P,Q_i) \tag{28}$$

for all points in Ω so that for $i = 1, 2, \ldots, n$ (ξ_i are constants)

$$\mathscr{L}_{P_k}\tilde{T}_P = \Delta g_k. \tag{29}$$

This is precisely what is meant by collocation.

Now we get the normal equations

$$\sum_{i=1}^n \xi_i \mathscr{L}_{P_k}\mathscr{L}_{Q_i}K(P,Q) = \Delta g_k, \tag{30}$$

and once they are solved, T_P can be found by (28). Substitution of ξ_i from (30) in (28) gives again (26).

It follows from (28) and a) that T_P is a potential, and from (29) and (30) that it gives the wanted values for the gravity anomalies at the points P_i. So far, we have in fact eliminated the statistical reasoning. The formula (27), on the other hand, follows only from the statistical hypothesis.

3. My treatment of the statistical aspect has so far been very loose. I shall now try to give it a better foundation.

Let us here, as in most of this paper, regard Ω as the domain outside a sphere with its centre O in the centre of gravity of the earth and the potential T, regular in Ω, as a stochastic process on Ω and invariant with respect to rotations around the centre O. All functions of two points $P, Q \in \Omega$ that are invariant with respect to such rotations are functions of the distances r_P and r_Q of the points P and Q from O and of the angle w between OP and OQ. From this follows that $K(P,Q)$ is a function of not six but only three independent parameters, and it can be found from T as the mean of

$$T_{P'} \cdot T_{Q'} \tag{31}$$

over all pairs of points P' and Q' satisfying

$$\left.\begin{aligned} OP' &= OP \\ OQ' &= OQ \\ < POQ' &= < POQ. \end{aligned}\right\} \tag{32}$$

In this way the means can be interpreted for simple interpolations (i. e. when no linear operators $\mathscr{L}$ are involved), and this is analogous to the interpretation of the means by Moritz. But in the case of operators not invariant with respect to rotation it does not work. Here a set of potentials Σ must be given so that the mean can be taken over Σ point by point, and the set Σ must have the ergodity property so that at each point the mean of $\varphi \in \Sigma$ equals the mean of an individual φ over the sphere (with centre at O) passing through the point in question.

Let $\varphi(P)$ be a potential regular in Ω. If R is a rotation (we regard only rotations about O) then $\varphi(RP)$ is also a potential regular in Ω. We shall then define Σ_φ for a given potential φ as the set of potentials $\varphi(RP)$ for all rotations R and interpret the mean at a point as the mean over Σ_φ using as a measure μ the invariant measure for the group of rotations G normalized so that $\mu(G) = 1$. When we use this interpretation of $M\{\ \}$, the reasoning leading to formula (26) runs without difficulty.

Now one could ask whether it is reasonable to represent the disturbing potential for a non-spherical planet as a rotation invariant stochastic process. My answer is that it may give a reasonable result if it is done in a reasonable way: The interpolation, i. e. the mathematical procedure that gives a potential T satisfying a finite set of conditions (from measurements), has an infinity of solutions. It must be theoretically exact, i. e. it must give one of these solutions, but for economical reasons we want to use a statistical method that can give us a solution that seems to us to use the observations as well as possible. Therefore, a rather rough approximation in the statistical hypothesis is not disastrous, but merely an economic question. The situation is exactly parallel to that in adjustment of geodetic networks, where it is essential that we use a high precision of the physical (or geometrical) constants, but much less important to use "exact" weight (or covariance) coefficients.

But the acceptance of a disturbing potential with some rotation invariant properties confronts us with a much severer problem: The T that we calculate must be defined and harmonic not only in the space outside to the surface of the earth but also down to a Bjerhammar sphere, and we know that the physically existing T of the earth cannot be extended to a potential with this property.

Now back to the formulae.

A rotation invariant $K(P, Q)$ restricted to a spherical surface in Ω with centre at O and radius R is a function of w only and can be expanded into a series in Legendre polynomials:

$$K(P,Q) = \sum_{n=0}^{\infty} A_n (2n+1) P_n(\cos w) \text{ for } r_P = r_Q = R. \tag{33}$$

Since $K(P, Q)$ is a potential as a function of P and as a function of Q, it can be expanded as follows:

$$K(P,Q) = \sum_{n=0}^{\infty} (2n+1) \frac{R^{2n+2} A_n}{r_P^{n+1} r_Q^{n+1}} P_n(\cos w) \text{ for all } P, Q \in \Omega. \tag{34}$$

We see that $K(P, Q)$ is a function of r_P, r_Q and w only, and that it is symmetric and harmonic in Ω and regular at infinity. We shall just find the conditions for its being of positive type, and therefore we must investigate the expression

$$\sum_{i=1}^{N} \sum_{k=1}^{N} K(P_i, Q_k) x_i x_k = \sum_i \sum_k K(i,k) x_i x_k \tag{35}$$

$$\begin{aligned}
\sum_i \sum_k K(i,k) x_i x_k &= \sum_{n=0}^{\infty} \left\{ (2n+1) A_n R^{2n+2} \sum_i \sum_k \frac{P_n(\cos w_{ik})}{r_i^{n+1} r_k^{n+1}} x_i x_k \right\} \\
&= \sum_{n=0}^{\infty} A_n R^{2n+2} \left\{ \sum_{m=0}^{n} \sum_i \sum_k \frac{R_{nm}(\theta_i, \lambda_i) x_i}{r_i^{n+1}} \frac{\bar{R}_{nm}(\theta_k, \lambda_k) x_k}{r_k^{n+1}} \right. \\
&\qquad \left. + \sum_{m=1}^{n} \sum_i \sum_k \frac{\bar{S}_{nm}(\theta_i, \lambda_i) x_i}{r_i^{n+1}} \frac{\bar{S}_{nm}(\theta_k, \lambda_k) x_k}{r_k^{n+1}} \right\} \\
&= \sum_{n=0}^{\infty} A_n R^{2n+2} \left\{ \sum_{m=0}^{n} \left[\sum_i \frac{\bar{R}_{nm}(\theta_i, \lambda_i) x_i}{r_i^{n+1}} \right]^2 + \sum_{m=1}^{n} \left[\sum_i \frac{\bar{S}_{nm}(\theta_i, \lambda_i) x_i}{r_i^{n+1}} \right]^2 \right\}
\end{aligned} \tag{36}$$

If $A_n \geq 0$ for all n, (36) shows that

$$\sum_i \sum_k K(i,k) x_i x_k \geq 0, \tag{37}$$

but if $A_n < 0$ for some n, then for some set of points P_i and weights x_i

$$\sum_i \sum_k K(i,k) x_i x_k < 0.$$

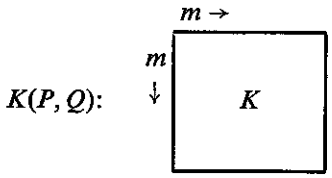
We may conclude that the necessary and sufficient condition of $K(P, Q)$ being of non-negative type is that for all the coefficients A_n in (34)

$$A_n \geq 0.$$

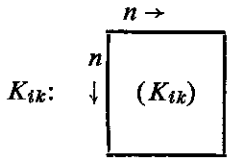
4. I think that the readers will agree with me that the notation used till now is rather clumsy. Therefore I have chosen already here to introduce another notation that only later will be theoretically justified, but which is much more handy.

As long as we are interested only in simple interpolation the new notation is nothing but the classical matrix notation and requires no justification.

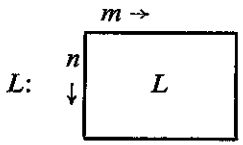
As we can suppose that we are interested only in the interpolated values at a finite set of points we just need to know $K(P, Q)$ at a finite number of points, say m points, i. e. we may represent $K(P, Q)$ by a square matrix:



The square $n \times n$ matrix $K(P_i, Q_k)$ is still represented as an $m \times n$ matrix:

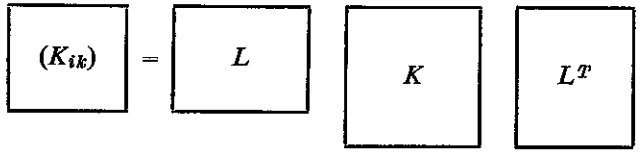


K_{ik} is obtained from $K(P, Q)$ using a projection matrix L. It is of dimensions $m \times n$ and consists of only 0's and 1's with precisely one 1 in each column.

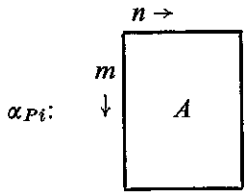


Then we may write:

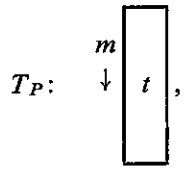
$$(K_{ik}) = LKL^T$$



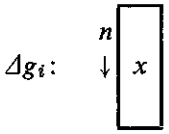
(α_{Pi}) is of the same form as L^T:



(T_P) is represented by a "long" vector:



and Δg_i by a "short" one:



I call the vector representing Δg_i x, because it may represent not only measurements of gravity anomalies, but also components of vertical deflections, pertubations of satellites, and so on.

When we use the method not on a problem of simple interpolation, but on a problem of collocation, the dimensionality m is infinite, and L must be interpreted as a linear operator or, better, as a set of n linear operators operating at different points. However, I do not think that the interpretation will cause any difficulties; but it will be practical for the reader continuously to compare the following example with the formulae (19) – (30).

The problem to be solved using the new notation is this:

If an interpolation or collocation is used on many measurements, the requirement that the result must be exactly compatible with all the results of the measurements may have the effect that the result becomes very oscillating; therefore it is often better to use smoothing than exact interpolation.

Let us suppose that in addition to what was given in the problem treated in section 2 of this chapter we are given an $n \times n$ variance-covariance matrix R for the n measurements, i. e.

$$x = y - v, \tag{38}$$

where x is the vector of the measurements, y is the true value and v is the vector of the corrections.

We put as before, cf. (19),

$$\tilde{t} = Ax. \tag{39}$$

The error of prediction will now be

$$(\varepsilon_P) = t - \tilde{t} = t - Ay + Av \tag{40}$$

and the error covariance

$$\left.\begin{aligned}(\sigma_{PQ}) &= M\{(\varepsilon_P)(\varepsilon_Q)^T\} = M\{(t - Ay + Av)(t^T - y^TA^T + v^TA^T)\}\\ &= M\{tt^T\} - M\{ty^TA^T\} - M\{Ayt^T\} + M\{tv^TA^T\}\\ &\quad + M\{Avt^T\} + M\{Ayy^TA^T\} - M\{Ayv^TA^T\}\\ &\quad - M\{Avy^TA^T\} + M\{Avv^TA^T\}\\ &= K - KL^TA^T - ALK + 0 + 0 + ALKL^TA^T - 0 - 0 + ARA^T\end{aligned}\right\} \tag{41}$$

or

$$(\sigma_{PQ}) = K - KL^TA^T - ALK + A[R + LKL^T]A^T. \tag{42}$$

The condition that ε_P must attain a minimum gives the normal equations

$$A[R + LKL^T] = KL^T \tag{43}$$

or

$$[R + LKL^T]A^T = LK. \tag{44}$$

Here the matrix R is positive definite and LKL^T is always non-negative definite; therefore the matrix of the normal equations

$$R + LKL^T$$

is always positive definite. If we define N by

$$N[R + LKL^T] = I, \tag{45}$$

where I is the $n \times n$ unit matrix, we get

$$A = KL^TN, \qquad A^T = NLK \tag{46}$$

and

$$\tilde{t} = KL^TNx. \tag{47}$$

For the error covariance σ_{PQ} (42) and (46) give

$$(\sigma_{PQ}) = K - KL^TNLK. \tag{48}$$

Also here it is most simple from a computational point of view first to solve the normal equations

$$[R + LKL^T]\xi = x, \tag{49}$$

and then to find $\tilde{t}$ by

$$\tilde{t} = KL^T\xi. \tag{50}$$

The *a posteriori* variance-covariance matrix for the measurements may also be calculated:

$$\left.\begin{aligned}L(\sigma_{PQ})L^T &= LKL^T - LKL^TNLKL^T =\\ LKL^T[I - N(R + LKL^T - R)] &= LKL^TNR =\\ (R + LKL^T - R)NR &= R - RNR.\end{aligned}\right\} \tag{51}$$

5. We return to the collocation problem (not the smoothing problem). Then $R = 0$ and (51) gives

$$L(\sigma_{PQ})L^T = 0, \tag{52}$$

which is not surprising.

I have called σ_{PQ} the error covariance, because it has been called so by others up to now, although I should have preferred to call it the *a posteriori* covariance, and I shall try to explain why.

Imagine the situation that we have made a collocation as in section 2 of this chapter and that afterwards we have made measurements at another set of points and now want to make the interpolation using both sets of measurements. I claim that it can be done in the following way:

Let the $\tilde{t}$ found from the first collocation be called $\tilde{t}_0$ and that found from the second one $\tilde{t}$, and put

$$\tilde{t} = \tilde{t}_0 + \tilde{s}. \tag{53}$$

Let the second set of measurements be called y, and let M play the role of L in the second set. Then $\tilde{s}$ may be found using the ordinary method, if for K we use

$$K' = K - KL^TNLK, \tag{54}$$

for L we use N, and for x we use

$$y - \tilde{t}_0,$$

i. e.

$$\tilde{s} = (K - KL^T NLK)M^T N'(y - \tilde{t}_0), \tag{55}$$

where

$$N' = [M(K - KL^T NLK)M^T]^{-1} \tag{56}$$

This may be verified directly by the well-known technique of manipulation with partitioned matrices, but it will not be done here as I do not think the result will be of any computational importance; nevertheless, I find it rather interesting, especially when it is formulated as follows:

If one has used the prediction method on the result of some measurements and after that made some new measurements (at other points), then the following two procedures will give the same results:

1) Use the prediction method on all measurements with the original covariance function.
2) Use the prediction method on the improvements (the new measured values – the predicted values at the same points) with the *a posteriori* covariance function.

Chapter II

The Least Squares Method in Hilbert Spaces

1. In the first chapter I have carried the generalization of Moritz' prediction formula to a point – or perhaps a little beyond a point – where some difficulties, e. g. with respect to a simple and consistent notation, seem to indicate the need for a more powerful mathematical apparatus.

I have several times stressed that the prediction method may be looked upon in two different aspects: the original prediction aspect, where one formally asks for the predicted value at a single point and, on the other hand, the collocation aspect where one asks for an interpolating function as an entity. Perhaps we can say that the first aspect is a discrete or finite one, while the second is a continuous or infinite one; therefore, it is not very surprising that we have to use methods from the functional analysis in order to make more extensive use of the second aspect.

The form of the normal equations for the second aspect (I, 8) seems to indicate that the interpolating function is a result of an adjustment problem. When we introduced the new notation, we saw that in the case of ordinary interpolation the notation was the ordinary matrix notation, provided that we asked for the value of the interpolating function at a finite set of points only. Let us first investigate this special case to find out which quantity is minimized.

Let us try to find an m-dimensional vector t so that n, say the first n of its coordinates, equal the n coordinates of the n-dimensional vector x and so that

$$t^T G t = \text{Min} \tag{1}$$

where G is an $m \times m$-dimensional positive definite matrix. The first condition can be written

$$Lt = x, \tag{2}$$

where the $m \times n$ matrix L can be partitioned according to

$$\{E_n \mid 0\}, \tag{3}$$

where E_n is the $n \times n$-dimensional unity matrix, and 0 is the $(n \times m - n)$-dimensional zero matrix. The problem is now a classical mean square adjustment problem with the equations of condition (2) and the weight matrix G. Its solution is

$$t = G^{-1}L^T(LG^{-1}L^T)^{-1}x. \tag{4}$$

If

$$G = K^{-1}, \tag{5}$$

then we have (I, 47), i. e. x is the solution to the prediction problem for $K = G^{-1}$. We can explain what we have seen here by saying that the correlation function defines the metric of the vector space of which t is an element, and that the matrix defining this metric is the weight matrix corresponding to the given correlation.

When we have found the t that minimizes

$$t^TK^{-1}t, \tag{6}$$

we can also find the value of the minimum:

$$\left.\begin{aligned} \min_t t^TK^{-1}t &= x^T(LKL^T)^{-1}LKK^{-1}KL^T(LKL^T)^{-1}x \\ &= x^T(LKL^T)^{-1}LKL^T(LKL^T)^{-1}x \\ &= x^T(LKL^T)^{-1}x. \end{aligned}\right\} \tag{7}$$

This result is very interesting because it is independent of the set of $m - n$ points at which we wanted to find the prediction, and because the result is not limited to simple interpolation but may be generalized to the case where the unit matrix E_n in (3) is replaced with another non-singular $n \times m$ matrix.

If we interpret K^{-1} as the matrix that defines the metric in the m-dimensional vector space of the t's, the norm of t is defined by

$$\| t \| = [t^TK^{-1}t]^{\frac{1}{2}}, \tag{8}$$

and using (7) we can define the norm of the interpolating function t for the problem given by L, K and x:

$$\| t \| = [x^T(LKL^T)^{-1}x]^{\frac{1}{2}}. \tag{9}$$

Now we shall find a more explicit expression of the norm. Suppose that $K(P, Q)$ can be expressed as follows

$$K(P, Q) = \sum_n k_n\varphi_n(P)\varphi_n(Q) \text{ for } P, Q \in \Omega, \tag{10}$$

where k_n are positive constants, $\varphi_n(P)$ is a set of functions harmonic in Ω and the range of n is finite or infinite. As we are here interested in simple interpolation, we want to find such a function $t(P)$ that

$$t(P_i) = x_i \qquad i = 1, 2, \ldots, N. \tag{11}$$

As in (I, 8) we put

$$\left.\begin{aligned} t(P) &= \sum_{i=1}^{N} \xi_i K(P, Q_i) = \sum_n k_n\varphi_n(P) \cdot \sum_{i=1}^{N} \xi_i\varphi_n(Q_i) \\ &= \sum_n t_n\varphi_n(P), \end{aligned}\right\} \tag{12}$$

where

$$t_n = k_n \sum_{i=1}^{N} \xi_i\varphi_n(Q_i) \tag{13}$$

are the coefficients in the expansion of $t(P)$ into a series of $\varphi_n(P)$.

The conditions (11) give the normal equations (I, 49 for $R = 0$)

$$LKL^T\xi = x, \tag{14}$$

i. e.

$$\left.\begin{aligned} \| t \|^2 &= \xi^TLKL^T(LKL^T)^{-1}LKL^T\xi \\ &= \xi^TLKL^T\xi = \sum_{i=1}^{N}\sum_{k=1}^{N} \xi_i\xi_kK(P_i, P_k) \\ &= \sum_i \sum_k \xi_i\xi_k \sum_n k_n\varphi_n(P_i)\varphi_n(P_k) \\ &= \sum_n k_n \sum_i \sum_k \xi_i\xi_k\varphi_n(P_i)\varphi_n(P_n) \\ &= \sum_n k_n[\sum_i \xi_i\varphi_n(P_i)]^2 = \sum_n \frac{t_2^n}{k_n}. \end{aligned}\right\} \tag{15}$$

If for all f for which

$$f = \sum_n f_n\varphi_n(P) \tag{16}$$

converges at all points of Ω and for which

$$\sum_n \frac{f_n^2}{k} \tag{16.5}$$

also converges, we define

$$\|f\| = \left[\sum_n \frac{f_2^n}{k}\right]^{\frac{1}{2}}, \tag{17}$$

then our definition is consistent with (9). We shall adopt this definition, and call the set of all such f H_K.

H_K is a linear space, i. e. for $f,g \in H_K$ and for a being any number $f+g \in H_K$ and $a \cdot f \in H_K$.

Let us for $f,g \in H_K$ define the scalar product $\langle f,g\rangle$ by

$$2\cdot\langle f,g\rangle = \|f+g\|^2 - \|f\|^2 - \|g\|^2. \tag{18}$$

It is evident that $\langle f,g\rangle = \langle g,f\rangle$, that the scalar product is bilinear, and that

$$\|f\|^2 = \langle f,f\rangle. \tag{19}$$

If $\langle f,g\rangle = 0$, we say that f and g are orthogonal.

For

$$f = \varphi_n + \varphi_m$$

we find by (18)

$$\|f\|^2 = \|\varphi_n + \varphi_m\|^2 = \|\varphi_n\|^2 + \|\varphi_m\|^2 + 2\langle\varphi_n, \varphi_m\rangle \tag{20}$$

and using (17)

$$\frac{1}{k_n} + \frac{1}{k_m} = \frac{1}{k_n} + \frac{1}{k_m} + 2\langle\varphi_n, \varphi_m\rangle \text{ for } m \neq n \tag{21}$$

and

$$\frac{4}{k_n} = \frac{2}{k_n} + 2\langle\varphi_n, \varphi_n\rangle \quad \text{for } m = n. \tag{22}$$

Equation (21) gives

$$\langle\varphi_n, \varphi_m\rangle = 0, \qquad \text{for } m \neq n, \tag{23}$$

and (22) gives

$$\|\varphi_n\|^2 = \langle\varphi_n, \varphi_n\rangle = \frac{1}{k_n}, \tag{24}$$

which shows that $\varphi_2, \varphi_1, \ldots$ is a system of orthogonal functions.

As a function of φ the covariance function $K(P,Q)$ is an element of the space H_K and for f being any element of H_K we can calculate

$$\left.\begin{aligned}\langle K(P,Q), f(Q)\rangle &= \Big\langle \sum_n k_n\varphi_n(P)\varphi_n(Q),\ \sum_m f_m\varphi_m(Q)\Big\rangle \\ &= \sum_n\sum_m k_m f_m \varphi_n(P)\langle\varphi_n(Q), \varphi_m(Q)\rangle \\ &= \sum_n k_n f_n \varphi_n(P)\cdot\frac{1}{k_n} = \sum_n f_n\varphi_n(P) = f(P).\end{aligned}\right\} \tag{25}$$

Specially for $f(Q) = K(Q,R)$ we get

$$\langle K(P,Q),\ K(Q,R)\rangle = K(P,R). \tag{26}$$

What we have found so far may be summarized as follows: The covariance function given by (10) defines a Hilbert space H_K consisting of the functions f satisfying (16) – (17), and $K(P,Q)$ is the reproducing kernel for H_K. From the theory of Hilbert spaces with kernel function it follows that it is not necessary to demand the convergence of (16) – it follows from that of (16.5).

As to the rotation invariant case we have found the form of the covariance function (I, 34):

$$\left.\begin{aligned}K(P,Q) &= \sum_{n=0}^{\infty}(2n+1)A_n\left[\frac{R^2}{r_P r_Q}\right]^{n+1} P_n(\cos w) \\ &= \sum_{n=0}^{\infty} A_n\left[\frac{R^2}{r_P r_Q}\right]^{n+1}\sum_{m=-n}^{n} E_{nm}(\theta_P,\lambda_P)E_{nm}(\theta_Q,\lambda_Q),\end{aligned}\right\} \tag{27}$$

where $A_n \geq 0$, and I have defined

$$\left.\begin{aligned}E_{nm}(\theta,\lambda) &= \bar{R}_{nm}(\theta,\lambda) \qquad \text{for } m \geq 0; \\ E_{nm}(\theta,\lambda) &= \bar{S}_{nm}(\theta,\lambda) \qquad \text{for } m \leq 0.\end{aligned}\right\} \tag{28}$$

H_K consists here of all functions of the form

$$f(P) = \sum_n{}' \sum_{m=-n}^{n} f_{nm}\left(\frac{R}{r_P}\right)^{n+1} E_{nm}(\theta,\lambda), \tag{29}$$

for which

$$\|f\|^2 = \sum_n{}' \sum_{m=-n}^{n} \frac{f_{nm}^2}{A_n} \tag{30}$$

converges. $\sum_n{}'$ denotes here and below that the sum is to be taken over the set of n for which $A_n > 0$.

Equations (23) and (24) give here

$$\left\langle \left(\frac{R}{r}\right)^{i+1} E_{ij}(\theta,\lambda), \left(\frac{R}{r}\right)^{k+1} E_{ke}(\theta,\lambda) \right\rangle = \begin{cases} \frac{1}{A_i} \text{ for } i = k \text{ and } j = e; \\ 0 \text{ in all other cases} \end{cases} \quad (31)$$

2. In this section the two classical least-squares adjustment problems will be generalized to Hilbert spaces.

In both cases we have two Hilbert spaces H_1 and H_2, and each of them may independent of the other be finite- or infinite-dimensional. The scalar products and the norms in the different spaces will be distinguished by lower indices, so that for instance $\| \ \|_2$ is the norm in H_2. The two spaces may be identical.

First I shall draw the attention of the reader to some definitions concerning linear operators in Hilbert spaces.

A linear operator A: $H_1 \to H_2$ is a function on H_1 to H_2 so that if A is defined for two elements x_1 and $y_1 \in H_1$ in such a way that

$$\left.\begin{aligned} Ax_1 &= x_2, \\ Ay_1 &= y_2, \end{aligned}\right\} \quad (32)$$

where x_2 and $y_2 \in H_2$, then A is also defined for $x_1 + y_1$, and

$$A(x_1 + y_1) = x_2 + y_2, \quad (33)$$

and for ax_1, where a is any real number, and

$$A(ax_1) = aAx_1. \quad (34)$$

A linear operator A: $H_1 \to H_2$ is said to be bounded if it is defined for every $x \in H_1$ and there exists such a positive number b that for all $x \in H_1$

$$\| Ax \|_2 \leq b \| x \|_1 \quad (35)$$

(in fact the fulfilment of the second condition follows from that of the first).

If $H_1 = H_2$, the bounded linear operators A: $H_1 \to H_2$ are analogous to square matrices; if $H_1 \neq H_2$, they are analogous to rectangular matrices. There is also the following analogy to transposition:

If A: $H_1 \to H_2$ is a bounded operator, then there exists one and only one bounded linear operator A^T: $H_2 \to H_1$ so that for all $x_1 \in H_1$ and $x_2 \in H_2$

$$\langle Ax_1, x_2 \rangle_2 = \langle x_1, A^T x_2 \rangle_1. \quad (36)$$

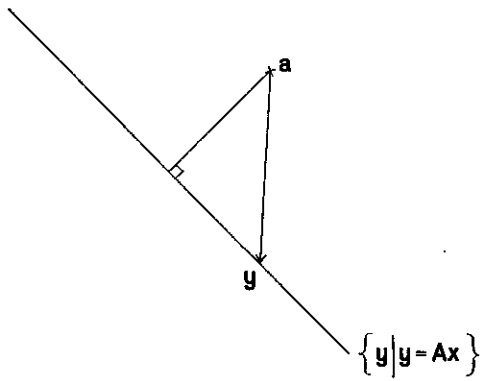
When I use the word operator, I generally mean bounded linear operator.

The first problem:

Given a bounded operator A: $H_1 \to H_2$ and an element $a \in H_2$. Find such an element $x \in H_1$ that $\|z\|_2$, where

$$z = Ax - a, \quad (37)$$

is as small as possible.



$\|z\|_2$ attains a minimum if, and only if, z is orthogonal to Ay for all $y \in H_1$:

$$\left.\begin{aligned} \langle z, Ay \rangle_2 &= 0 \\ \langle A^T z, \ y \rangle_1 &= 0 \end{aligned}\right\} \text{ for all } y \in H_1; \quad (38)$$

that is

$$\left.\begin{aligned} 0 = A^T z &= A^T Ax - A^T a, \\ A^T Ax &= A^T a. \end{aligned}\right\} \quad (39)$$

If the operator $A^T A$ is invertible, the unique solution is

$$x = (A^T A)^{-1} A^T a \quad (40)$$

and

$$\left.\begin{aligned} &\|z\|_2^2 \\ &= \langle A(A^TA)^{-1}A^Ta - a, A(A^TA)^{-1}A^Ta - a\rangle_2 \\ &= \langle A(A^TA)^{-1}A^Ta, A(A^TA)^{-1}A^Ta\rangle_2 - 2\langle A(A^TA)^{-1}A^Ta, a\rangle_2 + \langle a,a\rangle_2 \\ &= \langle A^TA(A^TA)^{-1}A^Ta, (A^TA)^{-1}A^Ta\rangle_1 - 2\langle A(A^TA)^{-1}A^Ta, a\rangle_2 + \langle a,a\rangle_2 \\ &= \langle A^Ta, (A^TA)^{-1}A^Ta\rangle_1 - 2\langle (A^TA)^{-1}A^T, A^Ta\rangle_1 + \langle a,a\rangle_2 \\ &= \langle a,a\rangle_2 - \langle a, A(A^TA)^{-1}A^Ta\rangle_2 = \langle a,a\rangle_2 \\ &\qquad - \langle A(A^TA)^{-1}A^Ta, A(A^TA)^{-1}A^Ta\rangle_2 \\ &= \|a\|_2^2 - \|A(A^TA)^{-1}A^Ta\|_2^2, \end{aligned}\right\} \quad (41)$$

because

$$[A(A^TA)^{-1}A^T]^2 = A(A^TA)^{-1}A^TA(A^TA)^{-1}A^T = A(A^TA)^{-1}A^T. \quad (42)$$

The operator of the normal equations (39) is non-negative definite:

$$\langle x, A^TAx\rangle_2 \geq 0 \qquad \text{for all } x \in H_2, \quad (43)$$

because

$$\langle x, A^TAx\rangle_2 = \langle Ax, Ax\rangle_1 = \|Ax\|_1^2 \geq 0. \quad (44)$$

The operator A^TA is symmetric:

$$(A^TA)^T = A^TA. \quad (45)$$

The second problem:

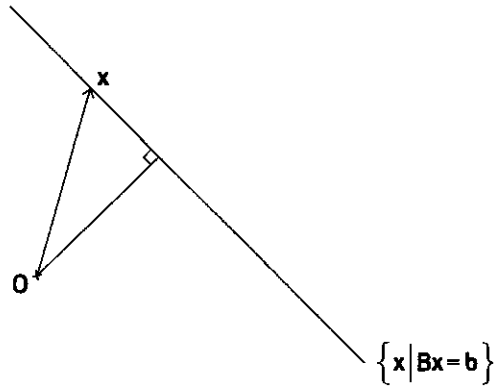
Given a bounded operator B: $H_2 \to H_1$ and an element $b \in H_1$. Find such an element $x \in H_2$ that

$$Bx = b \quad (46)$$

and that

$$\|x\|_2 \quad (47)$$

is as small as possible.



Any $x \in H_2$ can be written as

$$x = B^T\xi + y \quad (48)$$

with a suitable $\xi \in H_2$ and $y \in H_2$ so that y is orthogonal to B^Ts for all $s \in H_1$, that is

$$0 = \langle B^Ts, y\rangle_2 = \langle s, By\rangle_1 \text{ for all } s \in H_1 \quad (49)$$

or

$$By = 0.$$

Now (48) and (50) give

$$Bx = BB^T\xi + By = BB^T\xi. \quad (51)$$

The condition of x, written in the form (48), satisfying (46) is that ξ is determined by

$$BB^T\xi = b. \quad (52)$$

If BB^T, which definitely is symmetric and non-negative definite, is invertible, we have

$$\xi = (BB^T)^{-1}b. \quad (53)$$

In (48) x is expressed as a sum of two orthogonal elements of which the first is given; the norm of x therefore attains its minimum for $y = 0$, and we have the solution:

$$x = B^T(BB^T)^{-1}b. \quad (54)$$

For the minimum value $\|x\|_2$ we have

$$\left.\begin{aligned} \|x\|_2^2 &= \langle B^T(BB^T)^{-1}b, B^T(BB^T)^{-1}b\rangle_2 \\ &= \langle BB^T(BB^T)^{-1}b, (BB^T)^{-1}b\rangle_1 \\ &= \langle b, (BB^T)^{-1}b\rangle_1. \end{aligned}\right\} \quad (55)$$

As in the finite-dimensional case we can also here define and solve the more sophisticated mean squares problems.

There is a method of including in the two problems treated here problems that look more general by using direct sums of Hilbert spaces.

For two (finite- or infinite-dimensional) Hilbert spaces H_1 and H_2 we can define a third Hilbert space H_+:

$$H_+ = H_1 \oplus H_2 \quad (56)$$

called the direct sum of the spaces H_1 and H_2. It consists of all ordered pairs

$$(x_1, x_2) \quad (57)$$

of elements $x_1 \in H_1$ and $x_2 \in H_2$, and the scalar product in H_+ is defined as

$$\langle (x_1,x_2), (y_1,y_2)\rangle_+ = \langle x_1, y_1\rangle_1 + \langle x_2, y_2\rangle_2. \quad (58)$$

If we have a fourth Hilbert space H_3 and an operator A: $H_+ \to H_3$, it will often be practical to partition A:

$$A = (A_1 \mid A_2), \tag{59}$$

where A_1: $H_1 \to H_3$ and A_2: $H_2 \to H_3$.

If, in the same way, B: $H_3 \to H_+$, then

$$B = \begin{pmatrix} B_1 \\ \hline B_2 \end{pmatrix}, \tag{60}$$

where B_1: $H_3 \to H_1$ and B_2: $H_3 \to H_2$. The analogy to matrix notation is striking.

3. To be able to use the adjustment formulae we must know how to obtain explicit forms of the operators A and B_1 and here the covariance function or kernel $K(P, Q)$ may help us very much.

Formula (25) shows that $K(P, Q)$ can operate on a function $f(P)$. We shall now write (25) in this way:

$$Kf = f. \tag{61}$$

We see that in reality this operator is the unity operator in the Hilbert space H_K.

If $P_1, P_2, \ldots, P_N$ is any ordered set of points in Ω, then the operator L defined by

$$L_i = \langle K(P_i, Q), f(Q) \rangle = f(P_i) \tag{62}$$

is a bounded linear operator L: $H_K \to E_N$, where E_N is the N-dimensional Euclidean space, which to an element of H_K assigns the vector consisting of its values at the points P_i. This can be generalized, defining L by

$$L_i = \langle \mathscr{L}_{P_i} K(P, Q), f(Q) \rangle = \mathscr{L}_i f(P), \tag{63}$$

where $\mathscr{L}_i$ are N-linear functionals, e. g. differential operators operating at discrete points.

If (V_i) is a vector in E_N, then L^T is defined:

$$\langle \mathscr{L}_{Q_i} K(P, Q), V \rangle_N = \sum_i \mathscr{L}_{Q_i} K(P, Q) V_i, \tag{64}$$

and it follows that

$$LL^T V = \sum_k \mathscr{L}_{P_i} \mathscr{L}_{Q_k} K(P, Q) V_k, \tag{65}$$

i. e. LL^T is the same operator (matrix) as LKL^T in for example (I, 43). This is not surprising, as we have seen that the operator K is equivalent to the unity operator in H_K.

Now we can solve the problem which was previously solved by reasoning on covariance as an example of the second problem of adjustment.

H_2 is now the infinite-dimensional Hilbert space H_K, and H_1 is the N-dimensional space of measurements E_N, where N is the number of scalar measurements involved.

B: $H_2 \to H_1$ becomes here L: $H_K \to E_N$, and b is the N-vector of the results of the measurements. The normal equations are

$$LL^T \xi = b, \tag{66}$$

and now they really are normal equations. The solution is given by (54) or by

$$x = L^T (LL^T)^{-1} b. \tag{67}$$

We notice that we can compute the solution to the problem without explicitly using the norm in the potential space, but only the covariance function K. As we see from (55), the norm $\|x\|$ of the result can also be calculated without our knowing the explicit expression of the norm in the potential space.

An instructing example of the adjustment technique is the smoothing problem from the first chapter.

Here we have again the measurement equations

$$(Lf)_i \equiv \mathscr{L}_i f(P) = b_i, \qquad i = 1, 2, \ldots, n; \tag{68}$$

but now we do not want them to be satisfied exactly. We ask for such a potential f that

$$Lf - v = b, \tag{69}$$

and

$$\|f\|^2 + v^T P v = \text{minimum}, \tag{70}$$

where P is a positive definite $n \times n$ matrix.

This problem may be treated as an adjustment problem of the second type.

Our unknown quantities consist of the potential f and the vector v, and we may look upon them as a single quantity

$$x = f \oplus v \tag{71}$$

or

$$x = \begin{Bmatrix} f \\ \hline v \end{Bmatrix}. \tag{72}$$

If we have another element y of the same type

$$y = \begin{Bmatrix} g \\ \cdots \\ u \end{Bmatrix}, \tag{73}$$

we can define the scalar product

$$\langle x,y\rangle_2 = \langle f,g\rangle + v^T P u \tag{74}$$

and the norm

$$\| x \|_2 = \langle x,x\rangle_2^{\frac{1}{2}}. \tag{75}$$

Now we have defined the space H_2.

H_1 is the space of the measurements, i. e. the n-dimensional Euclidean space, and the operator B: $H_2 \to H_1$ is defined by (69) or by

$$Bx = \{L \mid -I_n\} \begin{Bmatrix} f \\ \cdots \\ v \end{Bmatrix} = Lf - v, \tag{76}$$

where I_n is the unity matrix in H_1.

The problem is now: Find such an element $x \in H_2$ that

$$Bx = b \tag{46}$$

and that

$$\| x \|_2 \tag{47}$$

is as small as possible.

We first have to find the normal equations

$$BB^T\xi = b, \tag{52}$$

but what is B^T?

From the definition of transposed operators it follows that B^T: $H_1 \to H_2$ is given by the equation

$$\langle Bx_2,x_1\rangle_1 = \langle x_2,B^T x_1\rangle_2 \tag{77}$$

for all $x_1 \in H_1$ and $x_2 \in H_2$, or if we write

$$x_2 = \begin{Bmatrix} f_2 \\ \cdots \\ v_2 \end{Bmatrix} \tag{78}$$

and

$$B^T x_1 = \begin{Bmatrix} y \\ \cdots \\ z \end{Bmatrix}, \; y \in H, \; z \in H, \tag{79}$$

and use (74) and (76)

$$\langle Lf_2 - v_2, x_1\rangle_1 = \langle f_2, y\rangle + v_2^T P z \tag{80}$$

or

$$\langle f_2, L^T x_1\rangle - \langle v_2, x_1\rangle_1 = \langle f_2, y\rangle + \langle v_2, Pz\rangle_1. \tag{81}$$

Equation (81) will only hold for all x_2, that is for all f_2 and v_2, if

$$y = L^T x_1 \tag{82}$$

and

$$z = -P^{-1}x_1, \tag{83}$$

i. e.

$$B^T x_1 = \begin{Bmatrix} L^T x_1 \\ \cdots \\ -P^{-1}x_1 \end{Bmatrix} = \begin{Bmatrix} L^T \\ \cdots \\ -P^{-1} \end{Bmatrix} x_1. \tag{84}$$

Then

$$BB^T = \{L \mid -I_n\} \begin{Bmatrix} L^T \\ \cdots \\ -P^{-1} \end{Bmatrix} = LL^T + P^{-1}, \tag{85}$$

where P^{-1} is the variance-covariance matrix R of the measurements. If this matrix is a diagonal matrix, as it is commonly assumed to be, it will have an ameliorating influence on the condition number of the normal equations

$$(R + LL^T)\xi = b. \tag{86}$$

The solution to the problem is now given by

$$x = B^T\xi \tag{87}$$

or more explicitly

$$f = L^T\xi, \; v = -R\xi. \tag{88}$$

Equation (55) gives here

$$\|x\|_2^2 = \|f\|^2 + v^T P v = b^T\xi = b^T(R + LL^T)^{-1}b. \tag{89}$$

The reader should carefully compare this solution of the smoothing problem with that given in the first chapter (I, 38 – I, 51).

The same problem could also be solved as a problem of the first type.

Chapter III

Hilbert Spaces with Kernel Function and Spherical Harmonics

1. As in the first approximation the earth is spherical, it is tempting to look for its gravitation potential among the potentials which are regular outside some sphere. There has been some discussion among geodesists as to the permissibility thereof, but before answering this important question we shall first study such sets of potentials and their connection with the spherical harmonics.

Let Σ be the part of the space outside a sphere with radius R and surface σ, which surface is not included in Σ. We shall be interested in several sets of potentials φ all of which are regular in Σ including infinity so that

$$\lim_{P \to \infty} \varphi(P) = 0. \tag{1}$$

The first set S of these potentials consists of those which are continuous in $\Sigma + \sigma$, i. e. those which have continuous boundary values on the surface σ of the sphere. For such potentials we can define a scalar product

$$\langle \varphi, \psi \rangle = \frac{1}{4\pi} \int_\sigma \varphi(P)\psi(P)d\sigma, \quad \text{for} \quad \varphi, \psi \in S. \tag{2}$$

This is the mean value of the product of the boundary values for the two potentials on the surface of the sphere. The corresponding norm is

$$||\varphi|| = \langle \varphi, \psi \rangle^{\frac{1}{2}} = \left(\frac{1}{4\pi} \int_\sigma \varphi(P)^2 d\pi \right)^{\frac{1}{2}}. \tag{3}$$

If for a potential $\varphi \in S$ $||\varphi|| = 0$, then φ must be zero on σ, because φ is continuous and

$$\int_\sigma \varphi(P)^2 d\sigma = 0. \tag{4}$$

But a potential which is regular in Σ and continuous in $\Sigma + \sigma$ is zero at all points of Σ if it is zero on σ. So $||\varphi|| = 0$ if, and only if, $\varphi = 0$: (3) defines really a norm, and the set S is a pre-Hilbert space.

It is well-known [4, p. 34–35] that, if the boundary values of a potential $\varphi \in S$ on σ are known, then φ can be found in Σ by Poisson's integral:

$$\varphi(P) = \frac{R(r^2{}_P - R^2)}{4\pi} \int_\sigma \frac{\varphi(Q)}{l^3} d\sigma, P \in \sum, Q \in \sigma, \tag{5}$$

where

$$l = \sqrt{r_P^2 + R^2 - 2Rr_P \cos \psi}, \tag{6}$$

and that "Poisson's kernel" can be expressed by spherical harmonics in the following way

$$\frac{R(r_P^2 - R^2)}{l^3} = \sum_{n=0}^{\infty} (2n+1) \left(\frac{R}{r_P} \right)^{n+1} P_n(\cos \psi). \tag{7}$$

Poisson's kernel is a function of the two points P and Q, of which one is on σ and the other in Σ. We can define a symmetrical kernel $K(P, Q)$ by putting:

$$K(P, Q) = \frac{r_P^2 r_Q^2 - R^4}{RL^3}, \tag{8}$$

where

$$L = \sqrt{\frac{r_P^2 r_Q^2}{R^2} - 2r_P r_Q \cos \psi + R^2}. \tag{9}$$

As $K(P, Q)$ is the result of substituting

$$R^2 \text{ for } R$$

and

$$r_P r_Q \text{ for } r_P,$$

in the formula for Poisson's kernel, the same substitution in (7) will give the expansion of $K(P, Q)$ into spherical harmonics:

$$K(P, Q) = \sum_{n=0}^{\infty} (2n+1) \left(\frac{R^2}{r_P r_Q} \right)^{n+1} P_n(\cos \psi). \tag{10}$$

$K(P, Q)$ is defined for P and $Q \in \Sigma$ and also for $P \in \sigma$ and $Q \in \Sigma$ (or $P \in \Sigma$ and $Q \in \sigma$), but for both P and $Q \in \sigma$ $K(P, Q)$ is zero for $P \neq Q$ and not defined for $P = Q$.

From (10) it follows, at least formally, that for either P or Q being fixed $K(P,Q)$ is a regular potential as a function of the other variable. A straightforward differentiation will verify that. Therefore, we have that for a fixed $P \in \Sigma$ $K(P,Q)$ as a function of Q is a member of the set S.

Now we may calculate the scalar product of $K(P,Q)$ and a potential $\varphi(Q) \in S$. Here we shall only use the values of $K(P,Q)$ for $Q \in \sigma$, in which case $r_Q = R$ and $K(P,Q)$ has the same values as Poisson's kernel. Therefore, the scalar product is exactly the right-hand member of (5), and we have

$$\langle K(P,Q), \varphi(Q)\rangle = \varphi(P), \text{ for } P \in \Sigma. \tag{11}$$

i. e. $K(P,Q)$ is the reproducing kernel for the set S of potentials.

With the metric defined in (3) S is not a Hilbert space, but only a pre-Hilbert space, i. e. not every sequence

$$\{\psi_n\} \quad \psi_n \in S \text{ for } n = 0, 2, \ldots$$

for which

$$\lim_{n\,m\to\infty} \|\psi_n - \psi_m\| = 0,$$

has a limit $\psi \in S$. Therefore we shall complete S to a Hilbert space H, which can be proved to consist of potentials regular in Σ and having square integrable boundary values on σ. For this Hilbert space $K(P,Q)$ is the reproducing kernel. Now the values of the integrals in (2), (3), etc. must be understood as the limits for $r > R$, $r \to R$ of the corresponding integrals over spheres with radius r.

If we define the functions with two indices

$$\{\varphi_n^m(P)\} \qquad \begin{matrix} m = -n, -n+1, \ldots, n-1, n \\ n = 0, 1, 2, \ldots, \end{matrix}$$

by

$$\varphi_n^m(P) = \left\{ \begin{matrix} \left(\frac{R}{r_P}\right)^{n+1} \bar{R}_{nm}(Q_P, \lambda_P) & \text{for} \quad m \geq 0; \\ \left(\frac{R}{r_P}\right)^{n+1} \bar{S}_{nm}(Q_P, \lambda_P) & \text{for} \quad m < 0, \end{matrix} \right\} \tag{12}$$

where $\bar{R}_{nm}$ and $\bar{S}_{nm}$ are the fully normalized harmonics [4, p. 31], then we can write (10) as follows:

$$K(P,Q) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \varphi_n^m(P)\varphi_n^m(Q), \tag{13}$$

which shows that the spherical harmonics $\{\varphi_n^m\}$ form a complete orthonormal system for the Hilbert space H. This means that every $\varphi \in H$ may be represented by a series expansion:

$$\varphi(P) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_n^m \varphi_n^m(P), \quad \text{for} \quad P \in \Sigma, \tag{14}$$

where

$$a_n^m = \langle \varphi, \varphi_n^m \rangle. \tag{15}$$

This, however, is to be understood in the sense of convergence in the Hilbert space metric

$$\lim_{N\to\infty} \|\varphi(P) - \sum_{n=0}^{N} \sum_{m=-n}^{n} a_n^m \varphi_n^m(P)\| = 0, \tag{16}$$

and we want a theorem which secures uniform convergence of the series. Fortunately the theory of reproducing kernels may help us here. For every $\psi \in H$ we can write

$$\left. \begin{aligned} |\varphi(P)| &= |\langle \psi(Q), K(P,Q)\rangle| \leq \|\varphi\| \cdot \|K(P,Q)\|_Q \\ &= \|\varphi\| \cdot \langle K(P,Q), K(P,Q)\rangle_Q^{\frac{1}{2}} = \|\varphi\| \cdot K(P,P)^{\frac{1}{2}}; \end{aligned} \right\} \tag{17}$$

here we have used the fact that $K(P,Q)$ is a reproducing kernel, Schwartz' inequality, the expression of the norm by the scalar product (the index Q signifies that the norm and the scalar product are to be understood with respect to Q) and once more the fact that $K(P,Q)$ is a reproducing kernel.

Let us use (17) on

$$\left. \begin{aligned} \|\varphi(P) - \sum_{n=0}^{N} \sum_{m=-n}^{n} a_n^m \varphi_n^m(P)\| &= \|\sum_{n=N+1}^{\infty} \sum_{m=-n}^{n} a_n^m \varphi_n^m(P)\| \\ &\leq (\sum_{n=N+1}^{\infty} \sum_{m=-n}^{n} (a_n^m)^2)^{\frac{1}{2}}, \end{aligned} \right\} \tag{18}$$

to get

$$|\varphi(P) - \sum_{n=0}^{N} \sum_{m=-n}^{n} a_n^m \varphi_n^m(P)| \leq (\sum_{n=N+1}^{\infty} \sum_{m=-n}^{n} (a_n^m)^2)^{\frac{1}{2}} \cdot K(P,P)^{\frac{1}{2}}. \tag{19}$$

A simple calculation gives that

$$0 < K(P,P)^{\frac{1}{2}} \leq \frac{R r_P}{r_P^2 - R^2}\sqrt{2}, \tag{20}$$

and then (19) shows that the series (14) converges uniformly for all P so that

$$r_P \geq r_0 > R. \tag{21}$$

We now have to go back to the formulae (14) and (15).

From the general theory for Hilbert spaces we know that from (14) and (15) follows

$$\|\varphi\| = \sum_{m=0}^{\infty} \sum_{m=-n}^{n} (a_n^m)^2)^{\frac{1}{2}}, \tag{22}$$

and that for every sequence $\{a_n^m\}$ for which

$$\sum \sum (a_n^m)^2$$

converges such an element $\varphi \in H$ exists that (14), (15) and (22) hold.

If we have a φ given by (14), it might be of interest to know if the series converges for points P so that $r_P < R$.

Let us put

$$A_n = (\sum_{m=-n}^{n} (a_n^m)^2)^{\frac{1}{2}}, \tag{23}$$

then we have from (22)

$$\sum_{n=0}^{\infty} A_n^2 = \|\varphi\|^2, \tag{24}$$

and therefore we must have

$$\lim_{n \to \infty} A_n \varrho^n = 0 \tag{25}$$

for $0 \leq \varrho \leq 1$. Let the least upper bound of ϱ for which (25) is valid be called ϱ_0. (ϱ_0 may be ∞).

Then (25) is valid for every ϱ so that $0 < \varrho < \varrho_0$.

If $\varrho_0 > 1$, then take two members ϱ_1 and ϱ_2 so that

$$0 < \varrho_1 < \varrho_2 < \varrho_0. \tag{26}$$

Then we have

$$\sum_{n=0}^{\infty} A_n \varrho_1{}^n < \infty. \tag{27}$$

for since

$$\lim_{n \to \infty} A_n \varrho_2^n = 0 \tag{28}$$

there will be some N so that

$$|A_n \varrho_2^2| < 1 \quad \text{for} \quad n > N. \tag{29}$$

We may then write

$$\sum_{n=0}^{\infty} (A_n \varrho_1^n)^2 \equiv \sum_{n=0}^{N} (A_n \varrho_1^2)^2 + \sum_{n=N+1}^{\infty} \left(\frac{\varrho_1}{\varrho_2}\right)^{2n} = \sum_{n=0}^{N} (A_n \varrho_1^n)^2 + \frac{\left(\frac{\varrho_1^2}{\varrho_2^2}\right)^{N+1}}{1 - \frac{\varrho_1^2}{\varrho_2^2}}, \tag{30}$$

and may also state that

$$\sum_{n=0}^{\infty} \sum_{m=-n}^{n} (a_n^m \varrho_1^n)^2$$

will converge for every ϱ_1 so that (20) or

$$0 < \varrho_1 < \varrho_0 \tag{31}$$

holds.

Let us now use functions $\{\psi_n^m\}$ defined as $\{\varphi_n^m\}$ by (12) only with R/ϱ_1 substituted for R; then we have the result:

$$\psi(P) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (\varrho_1^{n+1} a_n^m) \varphi_n^m(P) \quad \text{for} \quad r_P > \frac{R}{\varrho_1}. \tag{32}$$

If we call the sphere with centre at the centre of σ and radius $\frac{R}{\varrho_0}$ the limit sphere, we can express our result as follows:

The series expansion of a potential into spherical harmonics with O as origin will converge uniformly on the surface of and in the space outside any sphere with O as centre and so that all the singularities of the potential are in the interior of the sphere. There exists such a radius R' that for $R > R'$ the series will converge uniformly on the surface of and outside any sphere with radius R and centre O. This, however, is not true for $R < R'$.

The last (the negative) part of the theorem will be proved a few pages further on.

2. Geodesists are interested in exterior potential fields; therefore I have only treated such fields here. But by means of inversion with respect to a sphere we get corresponding results for interior spherical regions, which will recall the well-known theorems on the convergence of power series in the complex plane. There we have a limiting circle and uniform convergence in circles inside and divergence at all points outside the circle. I have the impression that many geodesists believe that it should, correspondingly, be so that the spherical harmonics series diverge at all points inside the limit sphere, but they cannot give any proof of it.

I shall not give the proof – on the contrary, I shall give an example that shows that the conjecture is false.

Consider the following potential

$$\varphi = \frac{1 - x}{(1 - x)^2 + y^2} \tag{33}$$

in a three-dimensional space. It corresponds to a uniform mass distribution on the line

$$\left.\begin{aligned} x &= 1 \\ y &= 0 \end{aligned}\right\} \quad (34)$$

and is only a function of x and y. It is elementary to show that it can be expanded into a series

$$\varphi = \sum_{n=0}^{\infty} a_n r^n \cos n\,\theta \cos n\lambda = \sum_{n=0}^{\infty} a_n (x^2 + y^2)^{\frac{n}{2}} \cos n\lambda \quad (35)$$

or

$$\varphi = \sum_{n=0}^{\infty} b_n r^n P_n^n (\sin\theta) \cos n\lambda. \quad (36)$$

$$(P_n^n (\sin\theta) = (-1)^n 1\cdot 3\cdot 5 \ldots (2n+1) \cos n\theta). \quad (37)$$

Neither the potential φ nor the coefficients in (35) or (36) depend on the coordinate z; therefore the series (36) will be convergent in the cylinder with $x = 0$, $y = 0$ as axis and with radius 1. If we make an inversion with respect to the sphere

$$x^2 + y^2 + z^2 = 1, \quad (38)$$

then

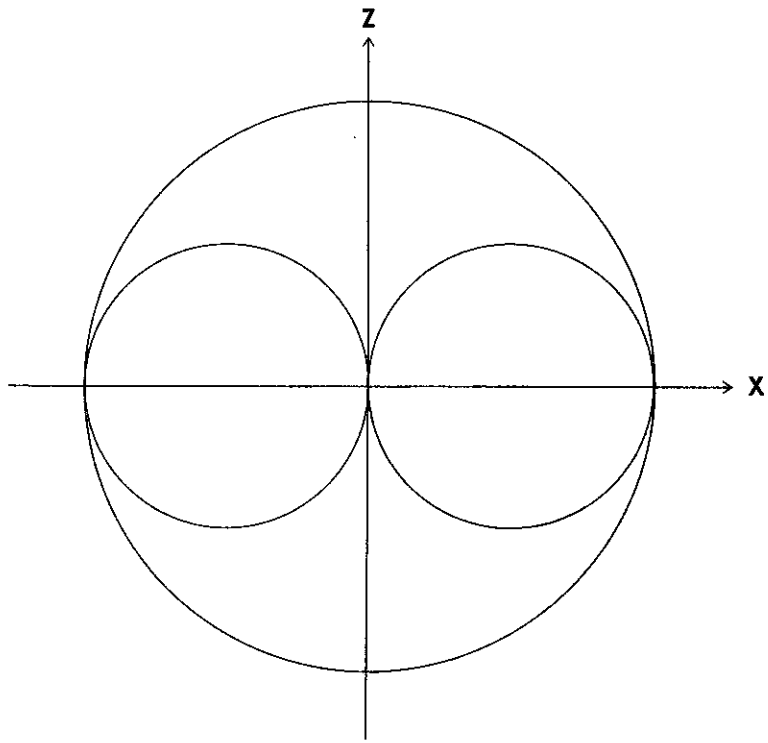
$$\Phi(x, y, z) = \frac{1}{r}\varphi\left(\frac{x}{r^2}, \frac{y}{r^2}, \frac{z}{r^2}\right) = \frac{r(r^2 - x)}{(r^2 - x)^2 + y^2} \quad (39)$$

is also a potential, and it is regular at infinity. The series expansion of Φ corresponding to (36) becomes

$$\Phi = \sum_{n=0}^{\infty} b_n \frac{1}{r^{n+1}} P_n^n(\sin\theta) \cos n\lambda, \quad (40)$$

which therefore will converge at all points of the space outside the torus into which the cylinder is transformed. The torus is that described by the circle:

$$(x - \tfrac{1}{2})^2 + z^2 = \tfrac{1}{4}:$$
$$y = 0;$$



when its plane is rotated around the z-axis. By using the idea of this example it is not difficult to construct other expansions into spherical harmonics that converge in regions that are not spherical. I have nevertheless still the feeling that normally the limit sphere bounds the region of convergence, but this question requires a closer study.

The criterion given here for the limit sphere is merely a theoretical one, as the limit radius can only be found when all the coefficients of the series are known. Therefore, we must have a criterion that can give more practical information.

From the above it follows that outside the limit sphere the series represents a regular potential, provided that the radius of the limit sphere is smaller than the radius of the smallest sphere containing the points at which the originally given potential was undefined or irregular. In other words, the series gives an analytic continuation of the potential. It is a well-known fact that the potential of a homogeneous sphere can be analytically continued to the whole space except the centre and that the normal potential of an oblate ellipsoid of revolution can also be

continued to the whole space except the "focal disc". Naturally the continuation of the outer potential into the gravitating body has nothing to do with the (irregular) potential that exists physically in the body. Two potentials which are regular in some domain and are identical in some open set of points in that domain are known to be identical in the whole domain of definition. Therefore, we may define a *maximal potential* as a regular potential defined in a connected region so that there does not exist any larger connected region in which the potential could be defined and be regular, and we may claim that to each potential there corresponds one and only one maximal potential continuation of the given. In the vicinity of every point of the boundary of the region of definition of a maximal potential there are points at which the potential is singular. Consequently, we may claim:

To every potential regular in the vicinity of infinity and to every sphere Σ containing all the singular points of the corresponding maximal potential there exists a series expansion of the form (14) which converges uniformly on and outside every sphere concentric with Σ so that no point of its surface is in the vicinity of any of the singular points. If we call the minimum value of the radii of such spheres R_0, then the series must be divergent at some points of every sphere with radius less than R_0 and concentric with Σ.

The remaining part of the claim runs like this:

1) For fixed θ and λ a series as (14) is a power series in $\frac{1}{r}$; therefore, if it diverges for some value of $\frac{1}{r}$, say α, it will diverge for every value of r so that $\frac{1}{r} > \alpha$ and the same θ and λ, or in other words, if it diverges at one point, it will diverge at all points of the line segment connecting that point and the origin.

2) From the definition of the limit sphere it follows that there are singularities in every vicinity of some of its points: therefore, given any $\varepsilon > 0$ we can find a sphere concentric with Σ and with radius $r > R_0 - \varepsilon$, so that there are singular points on its surface, and then the series cannot be uniformly convergent on such a sphere (for if it were, it would represent a continuation of the potential which was supposed to be maximal). The surface of a sphere is a closed set of points; therefore, convergence at all points would simply be uniform convergence.

3) From 2) it follows that there are spheres with radii arbitrarily near R_0 where the series diverges (at least at some points), and from 1) it follows that all the spheres with radii less than R_0 will also have divergence points on their surfaces.

The potential which is of most interest to geodesists is the disturbing potential T; therefore, it would be interesting to know whether it is reasonable to hope that the limit sphere for the corresponding maximal potential would be located below the surface of the earth. Using an idea from [10] I shall show that the basis for such a hope is slender. More precisely, I shall show that if to a gravitating body for which the corresponding maximal potential is defined below the surface we locally add some mass distribution, e. g. a grain of sand, above the surface at a place where the original maximal potential is defined below the surface, then the resulting maximal potential will have a singularity in the interior of the added mass distribution.

The proof is *almost* trivial. The resulting potential is the sum of the original maximal potential Φ and the potential φ of the added mass distribution and is regular at least where both are defined. Let us enclose the added mass distribution with a single closed surface σ leaving the singularities of Φ outside (this is possible according to what we have supposed). φ is regular outside σ and must therefore have a singularity inside σ (since every regular potential defined in the whole space vanishes). On the other hand Φ is regular inside σ, and therefore $\Phi + \varphi$ must have a singularity inside σ, and as we can select σ approximately, we have the theorem.

Here I must warn the reader that if he has not noticed the importance of the word "locally", then he has not really understood the proof. (The added mass distribution must not cover the whole surface of the original body, for then we should not be able to find σ).

If as an example of an added mass distribution I mentioned a grain of sand and not, as Moritz did, a mountain, it was in order to push the discussion on the series of harmonic functions near the gravitating masses ad absurdum. A popular way of expressing the theorem would be: even if the series were convergent at the surface of the earth, a displacement of a single grain of sand should spoil the convergence.

The consequence of this must be that the convergence of series of spherical harmonics near the surface of the earth is such an unstable property that it has no physical meaning at all.

I know that here many geodesists will argue that we may use the series without knowing anything about their convergence. I must confess that I do not understand what they mean by the little word "use".

Let us take an example. Let us regard the "Poisson kernel" (7). It has a singularity at the "north pole" – $r_P = R$, $\psi = 90°$ – and is regular at all other points in the space. For the singular point the series gives

$$\sum_{n=0}^{\infty} (2n + 1). \tag{41}$$

The sum of the first N members is N^2, and this is all right; for all other points of the sphere, $r_P = R$, the kernel is zero. But for the "south pole" – $\psi = -90°$ – the series gives

$$\sum_{n=0}^{\infty} (-1)^n(2n + 1), \tag{42}$$

and the sum of the first N members is $(-1)^{N+1} \cdot N$.

For a point having $r_P = \dfrac{R}{\varrho}$ and $\psi = -90°$ $(\varrho > 1)$ the series gives

$$\sum_{n=0}^{\infty} (-1)^n(2n + 1)\cdot\varrho^n, \tag{43}$$

and the sum of the first N members is

$$\frac{1 - \varrho + (-1)^N\varrho^{2N}(2N + 1 + (2N - 1)\varrho^2)}{(1 + \varrho)^2} \tag{44}$$

How can you "use" such a result?

I see very well that if we multiply the kernel by a small constant and add a dominating "normal potential", then it is a simple task in the resulting series to 'filter' the disturbing part with the increasing coefficients from the well-behaved part with decreasing coefficients. But if we happen to be interested mainly in the disturbing potential, how are we then to "use" the series?

I do like this example very much, so I ask the reader to have patience enough to follow an other experiment with it.

The series (7) may be written like this:

$$F_0 = \sum_{n=0}^{\infty} (2n + 1)\left(\frac{R}{r_P}\right)^n P_n(\cos\psi). \tag{45}$$

We define another series

$$F_\lambda = \sum_{n=0}^{\infty} \frac{2n + 1}{1 + 2^{n+1}\lambda}\left(\frac{R}{r_P}\right)^{n+1} P_n(\cos\psi), \quad \lambda \geq 0. \tag{46}$$

It is evident that for $\lambda = 0$ $F_\lambda = F_0$, and it seems likely that for small λ F_λ should approximate F_0; in fact we have:

$$\left.\begin{aligned} |F_0 - F_\lambda| &\leq \sum_{n=0}^{\infty}(2n + 1)\left[1 - \frac{1}{1 + 2^{n+1}\lambda}\right]\left(\frac{R}{r_P}\right)^{n+1} |P_n(\cos\psi)| \\ &\leq \sum_{n=0}^{\infty}(2n + 1)\frac{2^{n+1}\lambda}{1 + 2^{n+1}\lambda}\left(\frac{R}{r_P}\right)^{n+1}. \end{aligned}\right\} \tag{47}$$

For every fixed $r_P > R$ the series in the second line is uniformly convergent for $\lambda \geq 0$, and, consequently, it represents a continuous function. As this function is zero for $\lambda = 0$, we have

$$\lim_{\lambda \to 0} F_\lambda = F_0 \quad \text{for} \quad r_P > R. \tag{48}$$

The interesting thing is that for $\lambda > 0$ the series for F_λ (46) is convergent outside a sphere with radius $\dfrac{R}{2}$ and not R as in the case of the series for F_0.

It is not difficult to see that the method used in this example can be used generally to solve the following problem:

Given a potential φ defined in the space outside a sphere σ. Find a sequence of potentials $\{\varphi_n\}$ regular outside a sphere concentric with σ and with half the radius so that for all points outside σ

$$\lim_{n \to \infty} \varphi_n = \varphi. \tag{49}$$

If from the series expansions for the potentials $\{\varphi_n\}$ we take only the members up to the n'th degree φ'_n, then we also have

$$\lim_{n \to \infty} \varphi'_n = \varphi, \tag{50}$$

and here we have an approximation of φ by "polynomials" of spherical harmonics (i. e. series with a finite number of members).

We have seen a very important new aspect of the instability of convergence for series of spherical harmonics. We saw before (Moritz' theorem) that in the vicinity of every potential which can be expressed

by a series convergent at the surface of the earth there is another potential regular outside the earth but for which the series diverges at the surface. Now we see that *perhaps* there exists also another theorem (Runge's theorem) expressing that in the vicinity of every potential φ regular in the space outside the earth there is a potential for which the series of spherical harmonics converges down to the surface of some sphere in the interior of the earth; that is to say: φ can be approximated arbitrarily well by polynomials of spherical harmonics.

In fact there exists a Runge's theorem for physical geodesy (as to the name I have given it, see [1, p. 275–278]), but as the proof of it is rather technical, I have given it in an appendix and shall only state the result here:

> *Runge's theorem*: Given any potential regular outside the surface of the earth and any sphere in the interior of the earth. For every closed surface surrounding the earth (which surface may be arbitrarily near the surface of the earth) there exists a sequence of potentials regular in the whole space outside the given sphere and uniformly converging to the given potential on and outside the given surface.

This theorem is extremely important. In fact it permits a mathematical treatment of physical geodesy by giving a good compensation for the possibility of using series of harmonic functions converging down to the surface of the earth.

3. Runge's theorem merely establishes the existence of a sequence with the wanted properties. I shall now, at least theoretically, show how such a sequence can be found by means of the adjustment method from chapter II.

Let the given potential be φ and the domain on which it is given be Ω. The part of the space outside the sphere σ is called Σ, and we have $\Omega \in \Sigma$. Let us then define a metric $<,>_\Sigma$ for potentials regular in Σ, and let us call the corresponding Hilbert space H_Σ. We suppose that H_Σ has a reproducing kernel K. We know that it has one if the metric in H_Σ is that defined in the first part of this chapter (3). Let ω be a closed smooth surface in Ω surrounding and arbitrarily near to the boundary of Ω. Then we can define the norm

$$\| \varphi \|_\Omega = \left(\frac{1}{4\pi} \int_\omega \varphi^2 \, d\omega \right)^{\frac{1}{2}} \tag{51}$$

for potentials regular in Ω. By this norm and the corresponding scalar product we have defined a Hilbert space H_Ω consisting of potentials regular in Ω.

Then the problem is:

Find $\psi \in H_\Sigma$ so that

$$\| \psi \|_\Sigma^2 + \lambda \| \psi - \varphi \|_\Omega^2 = \text{minimum}, \tag{52}$$

where λ is a not yet specified constant. Let us rewrite (52) as

$$\| \frac{1}{\sqrt{\lambda}} \varphi - 0 \|_\Sigma^2 + \| \psi - \varphi \|_\Omega^2 = \text{minimum}. \tag{53}$$

Then the problem is an adjustment problem of the first type treated in chapter II (formulae (37)–(45)), provided that we put

$$\left.\begin{aligned} H_1 &= H_\Sigma, \\ H_2 &= H_\Sigma \oplus H_\Omega. \end{aligned}\right\} \tag{54}$$

The operator A: $H_1 \to H_2$ is defined as follows:

$$\left.\begin{aligned} A &= \begin{Bmatrix} A_1 \\ \hline A_2 \end{Bmatrix} \\ A_1 &: H_\Sigma \to H_\Sigma \\ A_1 &= \frac{1}{\sqrt{\lambda}} \\ A_2 &: H_\Sigma \to H_\Omega \\ A_2 \psi &= \psi' \text{ where } \psi' \text{ is the restriction of } \psi \text{ to } \Omega. \end{aligned}\right\} \tag{55}$$

$$a = \begin{Bmatrix} 0 \\ \hline \varphi \end{Bmatrix}. \tag{56}$$

The normal equations (II, 39):

$$A^T A \psi = A^T a \tag{57}$$

become here

$$\{ A_1{}^T \mid A_2{}^T \} \begin{Bmatrix} A_1 \\ \hline A_2 \end{Bmatrix} \psi = \{ A_1{}^T \mid A_2{}^T \} \begin{Bmatrix} 0 \\ \hline \varphi \end{Bmatrix} \tag{58}$$

or

$$\{ A_1^T A_1 + A_2^T A_2 \} \psi = A_2^T \varphi. \tag{59}$$

Here A_1 is a scalar operator, i. e. an operator indicating multiplication by a scalar, and thus identical with its transpose so that

$$A_1^T A_1 \psi = \frac{1}{\lambda}\psi. \tag{60}$$

For the restriction ψ' of ψ to Ω we have

$$\psi'(Q) = \langle K(Q,P), \psi(P)\rangle_\Sigma \quad \text{for} \quad Q \in \Omega, P \in \Sigma, \tag{61}$$

which follows from the trivial fact that

$$\psi'(Q) = \psi(Q), \quad \text{for} \quad Q \in \Omega, \tag{62}$$

and from the defining property of the reproducing kernel $K(Q,P)$.

To find the transpose A_2^T we remember the definition of the transposed operator (II. 36) and write for $\xi \in H_\Omega$:

$$\left.\begin{aligned} &\langle A^T\xi(Q), \psi(Q)\rangle = \langle \xi(Q), A(\psi Q)\rangle \\ &= \int_\omega \xi(Q)\langle K(Q,P), \psi(P)\rangle_\Sigma d\omega_Q = \Big\langle \int_\omega \xi(Q)K(Q,P)d\omega_Q, \psi(P)\Big\rangle_\Sigma, \end{aligned}\right\} \tag{63}$$

so that

$$A^T\xi(Q) = \int_\omega \xi(Q)K(Q,P)d\omega_Q \quad \text{for} \quad Q \in \Omega, P \in \Sigma \tag{64}$$

follows.

Now we can write (59):

$$\left.\begin{aligned} &\frac{1}{\lambda}\psi(P) + \int_\omega \psi(Q)K(Q,P)d\omega_Q = \int_\omega \varphi(Q)K(Q,P)d\omega_Q, \\ &\qquad Q \in \Omega, P \in \Sigma. \end{aligned}\right\} \tag{65}$$

Let us define $\xi \in H_\Omega$ by

$$\xi(Q) = \varphi(Q) - \psi(Q); \tag{66}$$

then for $P, Q \in \Omega$ (54) becomes

$$\xi(P) + \lambda \int_\omega K(Q,P)\xi(Q)d\omega_Q = \varphi(P). \tag{67}$$

Equation (67), which is analogous to the normal equations, is an integral equation. As we shall see, $\xi(P)$ can be found from (67), and then $\psi(Q)$ can be found by

$$\psi(Q) = \varphi(Q) - \xi(Q) \tag{68}$$

for $Q \in \Omega$; for $P \in \Sigma$ $\psi(P)$ can be found from (65):

$$\psi(P) = \lambda \int_\omega K(Q,P)\xi(Q)d\omega_Q \quad \text{for} \quad P \in \Sigma, Q \in \Omega. \tag{69}$$

The integral equation (67) is a Fredholm integral equation of the second kind with bounded continuous positive definite symmetric kernel. Normally such an integral equation is written with a minus and not a plus before the λ, and then one of the many elegant theorems on equations of this kind shows that all the eigenvalues are positive or zero; therefore (67) can have no eigensolutions for positive λ's. So for a given φ it will have a unique solution ξ for every positive λ, and then by (69) also ψ will be uniquely determined, and the least squares problem (52) has been solved.

But what does that mean?

Let us start with a "small" $\lambda > 0$, and let then λ increase. Then we may expect to find potentials $\psi \in H_\Sigma$ which, in H_Ω, approximate φ better and better so that in the part of Σ outside Ω ψ increase and, if we have luck,

$$\|\psi - \varphi\|_\Omega \to 0 \quad \text{for} \quad \lambda \to \infty. \tag{70}$$

I shall prove that this is in fact so, and for the proof I shall use the theory of integral equations and Runge's theorem.

Let the homogeneous integral equation

$$\Phi(P) - \lambda \int_\omega K(Q,P)\Phi(Q)d\omega_Q = 0; P, Q \in \Omega \tag{71}$$

have the eigenvalues $\{\lambda_n\}$ and the corresponding eigenfunctions $\{\varphi_n\}$; we may suppose that

$$0 < \lambda_1 \le \lambda_2 \le \lambda_3 \le \dots. \tag{72}$$

and that the eigenfunctions are normalized and orthogonal so that

$$\langle \varphi_n, \varphi_m\rangle_\Omega = \begin{cases} 1 \text{ for } m = n; \\ 0 \text{ for } m \neq n. \end{cases} \tag{73}$$

The eigenfunctions $\{\varphi_n\}$ are only defined in Ω, but they may also be defined in Σ and on σ by

$$\varphi_n(P) = \lambda_n \int_\omega K(Q,P)\varphi_n(Q)d\omega_P, P \in \Sigma + \sigma, Q \in \Omega, \quad (74)$$

because $K(Q,P)$ is regular for $Q,P \in \sigma + \Sigma$ provided that not both Q and $P \in \sigma$. From Mercer's theorem it follows that

$$K(P,Q) = \sum_n \frac{\varphi_n(P)\varphi_n(Q)}{\lambda_n} \quad (75)$$

for P, $Q \in \Omega$ and by the method of proof used for Mercer's theorem (75) can be proved for P, $Q \in \Sigma$.

Now $K(P,Q)$ is the reproducing kernel for H_Σ and, therefore,

$$\left.\begin{aligned}\varphi_n(P) &= \Big\langle \sum_m \frac{\varphi_m(P)\varphi_n(Q)}{\lambda_m}, \varphi_n(Q)\Big\rangle_\Sigma \\ &= \sum_m \frac{\varphi_m(P)}{\lambda_m}\langle \varphi_m(Q), \varphi_n(Q)\rangle_\Sigma.\end{aligned}\right\} \quad (76)$$

As $\{\varphi_n\}$ are orthogonal in H_Ω, they must be linearly independent, and so (76) implies

$$\frac{1}{\lambda_m}\langle \varphi_m, \varphi_n\rangle_\Sigma = \left\{\begin{array}{l}1 \text{ for } m = n; \\ 0 \text{ for } m \neq n,\end{array}\right\} \quad (77)$$

or

$$\langle \varphi_m, \varphi_n\rangle = \left\{\begin{array}{l}\lambda_n \text{ for } m = n; \\ 0 \text{ for } m \neq n,\end{array}\right\} \quad (78)$$

i. e. the functions $\left\{\frac{\varphi_n}{\sqrt{\lambda_n}}\right\}$ form a system of orthonormal functions in H_Σ. If we write

$$\varphi'_n = \frac{\varphi_n}{\sqrt{\lambda_n}}, \quad (79)$$

then (75) becomes

$$K(P,Q) = \sum_n \varphi'_n(P)\varphi'_n(Q). \quad (80)$$

A well-known theorem from the theory of Hilbert spaces with reproducing kernel establishes that, if a reproducing kernel can be expressed by (80), i. e. by a set of orthonormal functions, then this set will be complete. Therefore the set of functions $\{\varphi_n\}$ is complete in H_Σ.

The system $\{\varphi_n\}$ will also be complete in H_Ω. For if it were not, there would exist such an element $\eta \in H_\Omega$ that

$$\|\eta\|_\Omega = 1 \text{ and } \langle \eta, \varphi_n\rangle_\Omega = 0 \text{ for all } \varphi_n. \quad (81)$$

From Runge's theorem it follows that for any given $\varepsilon > 0$ there exists such an element $\mu \in H_\Sigma$ that

$$|\eta(Q) - \mu(Q)| < \varepsilon \text{ for } Q \in \omega. \quad (82)$$

Being an element of H_Σ $\mu(Q)$ may be expressed as

$$\mu(Q) = \sum_n a_n\varphi_n, \quad (83)$$

where the sum converges uniformly on ω. Therefore,

$$\|\eta(Q) - \sum_n a_n\varphi_n(Q)\|^2_\Omega = 1 + \sum a_n^2 > 1, \quad (84)$$

(from (81)). On the other hand,

$$\|\eta(Q) - \sum_n a_n\varphi_n(Q)\|^2_\Omega = \int_\omega (\eta(Q) - \sum_n a_n\varphi_n(Q))^2 d\omega \leq \varepsilon^2 \cdot A, \quad (85)$$

where A is the area of the surface ω. If we choose $\varepsilon < A^{-\frac{1}{2}}$, we have a contradiction, and, consequently, $\{\varphi_n\}$ will be complete also in H_Ω.

Now we can go back to the integral equation (67). Here we can express φ and ξ by the complete set $\{\varphi_n\}$:

$$\xi = \sum_n x_n\varphi_n, \varphi = \sum_n f_n\varphi_n \quad (86)$$

so that, using (75), we have

$$\sum_n x_n\varphi_n + \lambda\sum_n \frac{x_n}{\lambda_n}\varphi_n = \sum_n f_n\varphi_n, \quad (87)$$

or

$$x_n = \frac{\lambda_n}{\lambda_n + \lambda}f_n, \quad (88)$$

$$\xi = \sum_n \frac{\lambda_n}{\lambda_n + \lambda}f_n\varphi_n. \quad (89)$$

ξ is defined by (66) so that

$$\|\varphi(Q) - \psi(Q)\|_\Omega = \left(\sum_n \left(\frac{\lambda_n}{\lambda_n + \lambda}\right)^2 f_n^2\right)^{\frac{1}{2}}. \quad (90)$$

(We should remember that $\psi(Q)$ is a function of λ).

The series in the right-hand member of (90) is uniformly convergent in λ for $\lambda > 0$ since

$$\sum f_n^2 = \|\varphi\|_\Omega^2. \tag{91}$$

and

$$\frac{\lambda_n}{\lambda_n + \lambda} < 1 \quad \text{for} \quad \lambda > 0. \tag{92}$$

However, every member of the series converges to zero when $\lambda \to \infty$, and, therefore,

$$\lim_{\lambda \to \infty} \|\varphi(Q) - \psi(Q)\|_\Omega = 0, \tag{93}$$

from which uniform convergence follows in the usual way (17).

If we put

$$\psi = \sum_n P_n \varphi_n, \tag{94}$$

we find from (69)

$$P_n = \frac{\lambda}{\lambda_n + \lambda} f_n \tag{95}$$

or

$$\psi(P) = \sum_n \frac{\lambda}{\lambda_n + \lambda} f_n \varphi_n(P), \tag{96}$$

and

$$\|\psi\|_\Sigma^2 = \sum_n \left(\frac{\lambda}{\lambda_n + \lambda}\right)^2 f_n^2. \tag{97}$$

Only if the series Σf_n^2 is convergent, i. e. if the definition of φ can be extended to Σ, does the series in (97) converge for $\lambda \to \infty$. Only in this case does

$$\lim_{\lambda \to \infty} \psi = \varphi \tag{98}$$

hold in H_Σ. $\psi(P)$ does not in general converge for $\lambda \to \infty$ (for $P \in \Sigma$ but not $\in \Omega$), as can be seen from the example (46).

Now we have – at least theoretically – solved the problem of approximation of potentials down to the surface of the earth by potentials regular down to a Bjerhammar sphere and thus given a sound mathematical foundation of the method described in the previous two chapters. Moreover – as far as I can see – only this method presents a way in which one can find such approximations from physical measurements of the effects of the potential.

The proof given here of the convergence of the function ψ for $\lambda \to \infty$ might be used as a model for proofs of the convergence of the results of the application of the adjustment method on concrete problems in physical geodesy, provided that the number and the quality of the measurements increase until we have enough exact measurements. However, I shall not give such a proof for any practical case here, because I do not really see the value thereof. I believe that sufficient information about the reliability of the results can be found by means of the statistical method mentioned in the first chapter. The important information that Runge's theorem gives us in this connection is the method of approximation, which does not introduce any form of systematic error in the result.

4. Before leaving the question about the convergence of series of spherical harmonics I should like to advance a few naive considerations.

It has often been said that generally the convergence of series of spherical harmonics is slow. (I often wonder if those who say so have ever tried to calculate e^{-100} using directly the very well-known power series for e^x which converges for all x).

The reason why the series of spherical harmonics used here are so slowly converging is rather that the functions we want to represent are very complicated (i. e. contain a large amount of information) than that the spherical harmonics are not well suited for the set of functions in which we are interested.

If we try to describe some function defined on a sphere by a series of spherical harmonics of up to the 36th degree, then we must have $36^2 = 1296$ parameters, but we cannot expect that details of a magnitude less than $180°/36 = 5°$ can be sharply mapped. If we want a more detailed mapping, the price to be paid is more parameters and this is relatively independent of the type of function used for the mapping. By local interpolation it is of course possible to map much smaller details by a suitable choice of the 1296 first coefficients in the series, but then we are to expect a very "wild" behaviour of the series outside the local domain in which it has been forced to follow the details.

After this warning I shall say something about criteria for the choice of kernels and the corresponding metrics.

There exists an infinite number of metrics symmetric with respect to rotation for sets of potentials regular outside a given Bjerhammar sphere and having corresponding reproducing kernels. One of them was

treated in the first few pages of this chapter. There is another metric which has been mentioned sometimes in literature. It is defined by the scalar product:

$$\langle \varphi, \psi \rangle = \frac{1}{4\pi} \int_{\Sigma} \operatorname{grad} \varphi \cdot \operatorname{grad} \psi \, d\Sigma, \tag{99}$$

where the domain of integration is the whole exterior space. Using Green's theorem and $\Delta \psi = 0$ we find, however:

$$\langle \varphi, \psi \rangle = -\frac{1}{4\pi} \int_{\sigma} \varphi \frac{\partial \psi}{\partial r} d\sigma. \tag{100}$$

This property of the special metric has made it suitable for the study of the classical boundary value problems related to Laplace's equation. But for our purpose it is not of very great interest, since it has the drawback that the expression of the reproducing kernel is a rather complicated one containing a logarithm.

A metric that has a slightly more complicated scalar product but a much simpler reproducing kernel is defined by the following scalar product:

$$\langle \varphi, \psi \rangle_L = \frac{1}{4\pi} \int_{\Sigma} \frac{1}{r} \operatorname{grad} \varphi \cdot \operatorname{grad} \psi \, d\Sigma \tag{101}$$

and has the reproducing kernel

$$K(P, Q) = \frac{2R}{L},$$

where L is given by (9)

$$L = \sqrt{\frac{r_P^2 r_Q^2}{R^2} - 2 r_P r_Q \cos \psi + R^2}. \tag{9}$$

If P is a point in Σ, then

$$P' = \left\{ \frac{x_P}{R^2}, \frac{y_P}{R^2}, \frac{z_P}{R^2} \right\} \tag{102}$$

is the point in the interior of the sphere at which P is mapped by inversion with respect to the sphere. We have

$$L = \frac{r_P}{R} \sqrt{r_Q^2 - 2 r_Q r_{P'} \cos \psi + r_{P'}^2}, \tag{103}$$

i. e. for a fixed P $\frac{1}{L}$ is proportional to $\frac{1}{QP'}$, and therefore $\frac{1}{L}$ is a regular potential in Σ as a function of one of the points P, Q, the other being fixed, and

$$\left.\begin{aligned} \frac{2R}{L} &= \frac{2R^2}{r_P} \frac{1}{QP'} = \frac{R^2}{2 r_P} \frac{1}{r_P'} \sum_{n=0}^{\infty} \left(\frac{r_P}{r_Q} \right)^{n+1} P_n(\cos \psi) \\ &= 2 \sum_{n=0}^{\infty} \left(\frac{R^2}{r_P r_Q} \right)^{n+1} P_n(\cos \psi). \end{aligned}\right\} \tag{104}$$

I shall prove that $\frac{2R}{L}$ is the reproducing kernel corresponding to $\langle , \rangle_L$. Since

$$\varphi_i^k = \left(\frac{R}{r} \right)^{i+1} E_i^k \text{ for } k = -i, -i+1, \ldots i \text{ and } i = 0, 1, \ldots, \tag{105}$$

where E_i^k are $2i + 1$ fully normalized spherical harmonics for $i = 0, 1, 2, \ldots$, is a complete orthogonal system, a scalar product is known once its effect on $\{\varphi_i^k\}$ is known. We find

$$\left.\begin{aligned} \langle \varphi_i^k, \varphi_j^l \rangle_L &= \frac{1}{4\pi} \int_{\Sigma} \frac{1}{r} \operatorname{grad} \varphi_i^k \cdot \operatorname{grad} \varphi_j^l \, d\Sigma \\ &= \frac{1}{4\pi} \int_{\Sigma} \frac{1}{r} \left\{ \frac{\partial}{\partial r} \varphi_i^k \frac{\partial}{\partial r} \varphi_j^l + \operatorname{grad}_2 \varphi_i^k \operatorname{grad}_2 \varphi_j^l \right\} d\Sigma, \end{aligned}\right\} \tag{106}$$

(where grad_2 is the two-dimensional gradient on the sphere)

$$= \frac{1}{4\pi} \int_{\Sigma} \frac{1}{r} \left\{ \frac{d}{dr} \left(\frac{R}{r} \right)^{i+1} \frac{d}{dr} \left(\frac{R}{r} \right)^{j+1} E_i^k E_j^l + \left(\frac{R}{r} \right)^{i+1} \left(\frac{R}{r} \right)^{j+1} \operatorname{grad}_2 E_i^k \operatorname{grad} E_i^l \right\} d\Sigma, \tag{106}$$

(now we shall make use of Green's theorem for the surface of the sphere)

$$= \frac{1}{4\pi} \int_{\Sigma} \frac{1}{r} \left\{ \frac{(i+1)(j+1)}{r^2} \left(\frac{R}{r} \right)^{i+j+2} E_i^k E_j^l - \left(\frac{R}{r} \right)^{i+j+2} E_j^l \Delta_2 E_i^k \right\} r \sin \theta \, d\theta \, d\lambda \tag{106}$$

(where Δ_2 is the Laplace-Beltrami operator on the sphere)

$$= \frac{1}{4\pi}\int_{\Sigma}\frac{1}{r}\left(\frac{R}{r}\right)^{i+j+2}\{(i+1)(j+1)+i(i+1)\}E_i^kE_j^l\sin\theta d\theta d\lambda$$

$$= \frac{(i+1)(i+j+1)}{4\pi}\int E_i^kE_j^l\sin\theta d\theta d\lambda\int_R^\infty\frac{R^{i+j+2}dr}{r^{i+j+3}} \qquad (106)$$

$$= (i+1)(i+j+1)\delta_{ij}\cdot\delta_{kl}\cdot\frac{1}{2i+2} = \begin{cases}\frac{2i+1}{2} \text{ for } i=j,\ k=l\\ 0 \quad \text{in all other cases.}\end{cases}$$

We can now write:

$$\langle\frac{2R}{L},\varphi_j^l\rangle_L = \langle 2\sum_{i=0}^{\infty}\left(\frac{R^2}{r_Pr_Q}\right)^{i+1}P_i(\cos\psi),\varphi_j^l\rangle_L = \qquad (107)$$

$$\langle 2\sum_{i=0}\frac{1}{2i+1}\varphi_i^k(P)\varphi_i^k(Q),\varphi_j^l(Q)\rangle_L = \varphi_j^l(P),$$

and so we have proved that $\frac{2R}{L}$ is the reproducing kernel in the Hilbert space H_L with the scalar product $\langle\ ,\rangle_L$.

This chapter will be concluded by a short discussion of the important problem: how to choose the metric (or the kernel) for practical computations.

The most obvious idea is perhaps to use as kernel the finite series of spherical harmonics which corresponds to Kaula's expansion of the correlation function $C(P,Q)$. [7]. (In Kaula's publications one of the coefficients in this expansion is negative, but it cannot be so; is it an iterated printing error?).

But if one uses a kernel with only a finite number of members, one has limited the solution to a finite-dimensional space consisting of potentials expressible by spherical harmonics of the same degrees as those occurring in the kernel. To get good local results one, consequently, has to use a very large number of members.

Another possibility is to use a kernel with a simple closed expression, e. g. $\frac{2R}{L}$ (or $\frac{2R}{L}$ multiplied by a suitable constant). If the radius R of the Bjerhammar sphere is chosen a few per cent smaller than the mean radius of the earth, a good local approximation to Kaula's correlation function is achieved.

As far as I can see the best thing to do is to use a combination of these two ideas, i. e. $\frac{2R}{L}$ multiplied by some constant + a correction consisting of a finite sum of spherical harmonics so that the corresponding correlation function is sufficiently similar to Kaula's correlation function. The scalar product will then be a constant multiple of $\langle\ ,\rangle_L$ + a correction which is simple to calculate, but as the scalar product is not explicitly used in the calculations I shall not give the result here. From a theoretical point of view the important thing is that since this correction is finite, the Hilbert space corresponding to the corrected kernel consists of the same elements as does H_L.

Chapter IV

Application of the Method

Already today Molodenskiy's problem is looked upon as the classical problem of physical geodesy; it is therefore reasonable to start the discussion on the application of the kernel-method with this problem.

As I do not find that the mathematical aspects of Molodenskiy's problem are clearly formulated in literature, I shall first propose another formulation of the problem.

Molodenskiy's problem is the problem of finding a better approximation of the potential of the earth – the normal potential – from a given approximation by geodetic measurements.

Let us assume that we have measured the following data for points on the physical surface of the earth:

1) the astronomical geographical coordinates, i. e. the direction of the plumb line.
2) the potential of gravity of the earth W, and
3) the gravity g.

Let us also assume that we have given a normal potential U; we can then for each point P on the surface of the earth find such a point Q that

a) the normal potential at Q equals the potential at P:

$$U_Q = W_P, \tag{1}$$

and that

b) the geodetic normal coordinates of Q are equal to the astronomical coordinates of P, which can be expressed by the vector equation

$$\frac{1}{\gamma_Q}(\text{grad } U)_Q = \frac{1}{g_P}(\text{grad } W)_P; \tag{2}$$

both members of this equation represent unity vectors defining the directions in question. The locus of the points Q is the telluroid.

We want to find the vector $\vec{QP}$ and $T = W - U$. Here T, the disturbance

potential, is a regular potential, because the potential of the centrifugal force is the same in W and U.

Since $T = W - U$, (1) can now be written

$$T_P + U_P - U_Q = 0, \tag{3}$$

and (2) can be written

$$(\text{grad } T)_P + (\text{grad } U)_P - \frac{g_P}{\gamma_Q}(\text{grad } U)_Q = 0, \tag{4}$$

or

$$(\text{grad } T)_P + (\text{grad } U)_P - (\text{grad } U)_Q = \frac{g_P - \gamma_Q}{\gamma_Q}(\text{grad } U)_Q. \tag{5}$$

Now

$$g_P - \gamma_Q = \Delta g \tag{6}$$

is the gravity anomaly at the point in question, so that (5) becomes:

$$(\text{grad } T)_P + (\text{grad } U)_P - (\text{grad } U)_Q = \frac{\Delta g}{\gamma}(\text{grad } U)_Q. \tag{7}$$

If we introduce a Cartesian system of coordinates (x_1, x_2, x_3) and call the components of the vector $\vec{QP}$ p_1, p_2, and p_3, we can write (3) and (7) in linear approximation

$$T + \sum_{i=1}^{3} p_i \frac{\partial U}{\partial x_i} = 0, \tag{8}$$

and

$$\frac{\partial T}{\partial x_j} + \sum_{i=1}^{3} p_i \frac{\partial^2 U}{\partial x_i \partial x_j} = \frac{\Delta g}{\gamma}\frac{\partial U}{\partial x_j}, \quad j = 1, 2, 3. \tag{9}$$

These four equations must be satisfied for all points of the telluroid, where all variables except T, $\frac{\partial T}{\partial x_j}$, and p_j are known.

If T and $\frac{\partial T}{\partial x_j}$ were known at a point Q of the telluroid, then (8) and (9) would be four linear equations in the three variables p_j, which could be determined if, and only if, (8) and (9) were compatible, i. e.:

$$\begin{vmatrix} T & \frac{\partial U}{\partial x_i} \\ \frac{\partial T}{\partial x_j} & \frac{\partial^2 U}{\partial x_i \partial x_j} \end{vmatrix} = \frac{g}{\gamma} \begin{vmatrix} 0 & \frac{\partial U}{\partial x_i} \\ \frac{\partial U}{\partial x_j} & \frac{\partial^2 U}{\partial x_i \partial x_j} \end{vmatrix} \quad \begin{matrix} i = 1, 2, 3 \text{ (columns)} \\ j = 1, 2, 3 \text{ (rows)} \\ (4 \times 4 \text{ determinants}); \end{matrix} \tag{10}$$

in other words (10) has to be satisfied for all points of the telluroid.

Equation (10) is the correct form of the boundary conditions for T in Molodenskiy's problem.

The boundary value problem for T is not one of the classical boundary value problems for potentials, i. e. problems where the potential, the normal derivative of the potential, or a linear combination thereof is given at the boundary. Ours is the so-called oblique derivative problem where a linear combination of the potential and the derivative of it in some direction is given at the boundary. It can be proved that the direction in question is that of the normal line through the point Q, the normal line being the curve consisting of points having the same normal coordinates as Q; the normal lines are approximately vertical. If this direction does not at any point of the boundary coincide with the direction of a tangent to the boundary at the same point and if for the whole boundary the direction is to the same side of the boundary surface, then we have the regular oblique derivative problem, provided, however, that the boundary surface and the coefficients in equation (10) satisfy some very weak regularity conditions.

The oblique derivative problem is in general very complicated, but if it is regular it has been proved ([5, p. 265], [2, p. 82]) that the theorem called Fredholm's alternative applies. Since we have a certain liberty in choosing the mathematical model for the surface of the earth, we can and shall assume that we have to do with the regular oblique derivative problem.

Fredholm's alternative runs as follows:

Either

a) there is no regular potential T different from zero which satisfies the homogeneous boundary value equation corresponding to (10) (i. e. (10) with the right-hand member equal to zero); if so, (10) has for all right-hand members a unique solution T that is a regular potential (outside the boundary surface)

or

b) the homogeneous problem has a finite number n of linearly independent solutions; if so, the inhomogeneous problem is solvable only if the right-hand member satisfies n linearly independent linear homogeneous conditions, and then it has n linearly independent solutions so that the difference between two arbitrary solutions is a solution to the corresponding homogeneous problem.

In pure mathematics it makes good sense to work with clear alternatives – in applied mathematics and in numerical mathematics the facts are more blurred. There we often have a situation where it is practically impossible to tell whether we are in case a) or case b); we have the same situation when we are to solve a system of linear algebraic equations so that the coefficient matrix has a "small" eigenvalue: the system is unstable – a "small" change in the input values may cause a change in the result that is not "small".

To find out which case applies to Molodenskiy's problem, we shall first consider the simplified situation where we have a non-rotating planet.

Here the normal potential U is regular also at infinity as are all its derivatives with respect to the Cartesian coordinates; thus,

$$\frac{\partial U}{\partial x_1}, \frac{\partial U}{\partial x_2}, \frac{\partial U}{\partial x_3}$$

are regular potentials.

If in the left-hand determinant of (10) we substitute $\frac{\partial U}{\partial x_h}$ for T, then the first and the $h+1$'th column are identical and the determinant vanishes, i. e. $\frac{\partial U}{\partial x_h}$ is a solution to the homogeneous problem corresponding to Molodenskiy's problem for $h = 1, 2, 3$; that is to say we are in case b) with n being at least three.

Let us first suppose that $n = 3$.

Then, if the boundary value problem has a solution T_0 – and so it has if the gravity anomalies satisfy three linear equations –,

$$T = T_0 + a\frac{\partial U}{\partial x_1} + b\frac{\partial U}{\partial x_2} + c\frac{\partial U}{\partial x_3} \tag{11}$$

is a solution for all values of a, b and c, and we may choose the constants so that the gravity centre for T coincides with the gravity centre for U, i. e. so that the three first-order members of the expansion of T into spherical functions in the vicinity of infinity vanish.

For $n > 3$ the anomalies must satisfy more than three conditions, but if there are solutions, there will always be such which have gravity centres coinciding with that of the earth. However, the solutions will be $n - 3$ times indeterminate.

Now we shall discuss the interesting case where the earth is rotating. Let us fix the coordinate system so that the origin is at the gravity

centre of the earth and so that the x_3-axis coincides with the axis of rotation. Then

$$\frac{\partial U}{\partial x_3}$$

is still a solution to the homogeneous problem, whereas

$$\frac{\partial U}{\partial x_1} \quad \text{and} \quad \frac{\partial U}{\partial x_2}$$

are only formal solutions; they are not zero at infinity; they are not even bounded. Therefore, we can only say that n is *at least* one and, if the problem has a solution T_0, we cannot generally obtain coincidence between the gravity centres of T and the earth. So we have again the situation that in order to obtain a usable solution, we must have a set of Δg satisfying (at least) three conditions. And even more: it is known that also two of the first-order members in the expansion must be zero [4, p. 62], so in reality we have (at least) five conditions that should be satisfied.

The result of this investigation must be that Molodenskiy's problem is ill-posed, in the terminology of J. Hadamard [8]. For a problem to be well-posed it should according to Hadamard have one and only one solution for arbitrarily given data, and small variations of the given data should cause reasonably small variations in the solution. The correct formulation of Molodenskiy's problem would be to ask for a potential T that is regular outside the telluroid, that satisfies five conditions at infinity (the vanishing of the first-order members and two of the second-order members in the expansion into spherical harmonics) and that satisfies equation (10) where in the right-hand member $\Delta g + V$ is substituted for Δg, where

$$\int_{\omega} p(Q) V^2(Q)\, d\omega = \text{minimum},$$

the integral to be taken over the telluroid and p being a given positive weight function. This form of boundary value problem might be called a least-squares boundary value problem.

I shall not follow up this idea here, since we are not in possession of a continuous field of boundary data, but I have tried to point out that adjustment methods, also from an abstract theoretical point of view, are more realistic than the classical approach.

In the computation for the determination of the potential it may be practical to include results concerning the deflection of the vertical on one hand and to make it possible to find deflections of the vertical on the other hand. Therefore, we must find the differential operators that give these quantities from the representation of the disturbing potential, and I shall here derive those corresponding to the same model that led to the formula (10) for the differential operator giving Δg.

At the point P the direction of the physical vertical is – grad W and the direction of the normal vertical is – grad U. We are interested in the projection on a horizontal plane through P of the difference between the unity vectors in the two verticals. This difference is

$$\left.\begin{aligned} -\frac{1}{g}\operatorname{grad} W + \frac{1}{\gamma}\operatorname{grad} U &= -\frac{1}{g}\operatorname{grad} W + \frac{1}{\gamma}(\operatorname{grad} W - \operatorname{grad} T) \\ &= \left(\frac{1}{\gamma} - \frac{1}{g}\right)\operatorname{grad} W - \frac{1}{\gamma}\operatorname{grad} T. \end{aligned}\right\} \tag{12}$$

Let us use a coordinate system with the z-axis in the direction of the physical vertical, the x-axis in the west-east direction and the y-axis in the south-north direction. The horizontal plane through P has the equation z = constant, i. e. the projection on this plane of grad W is zero, and we have for the west-east and the south-north components of the vertical deflection:

$$\xi = \frac{1}{\gamma}\frac{\partial T}{\partial x} \quad \text{and} \quad \eta = -\frac{1}{\gamma}\frac{\partial T}{\partial y}. \tag{13}$$

If the interpolation method is used for the interpolation of vertical deflections using gravity anomalies, we have at least two advantages over the classical method: 1) the theoretical advantage that all the measurements enter into the calculation in the same way, 2) the practical advantage that there is no integration in the process; the anomalies at the measured stations enter directly into the calculations, which, therefore, can be automatized.

I find it a very attractive thought that the problem of local interpolation of vertical deflections can be solved by the smoothing method so that the measurements of vertical deflections and of gravity anomalies enter formally into the calculations in the same way. This problem is in fact the first one on which we have planned to use the present theory at the Danish Geodetic Institute.

But still more attractive is it to use this method on an integrated adjustment of dynamical satellite measurements and measurements referring to the potential at the surface of the earth.

I cannot write down yet practical formulae to be used in such calculations; I think that special research work is needed on this problem and I can only offer some theoretical comments on the question.

It is important to remember that the disturbing potentials T and R used for measurements relative to the surface of the earth and to the satellites respectively are not the same, but as the difference between them is known, this fact does not cause severe difficulties. Also the slightly more complicated case where the difference is not completely known but is dependent on one or more unknown parameters may be dealt with by letting these parameters enter into the adjustment as unknowns. This problem is perhaps not unrealistic.

As a starting point for the discussion of the explicit form of the normal equations I take the formulae from [6]:

$$\left.\begin{aligned}
\frac{da}{dt} &= \frac{2}{na}\frac{\partial R}{\partial M},\\
\frac{de}{dt} &= \frac{1-e^2}{na^2e}\frac{\partial R}{\partial M} - \frac{(1-e^2)^{\frac{1}{2}}}{na^2e}\frac{\partial R}{\partial \omega},\\
\frac{d\omega}{dt} &= \frac{\cos i}{na^2(1-e^2)^{\frac{1}{2}}\sin i}\frac{\partial R}{\partial i} + \frac{(1-e^2)^{\frac{1}{2}}}{na^2e}\frac{\partial R}{\partial e},\\
\frac{di}{dt} &= \frac{\cos i}{na^2(1-e^2)^{\frac{1}{2}}\sin i}\frac{\partial R}{\partial \omega} - \frac{1}{na^2(1-e^2)^{\frac{1}{2}}\sin i}\frac{\partial R}{\partial \Omega},\\
\frac{d\Omega}{dt} &= \frac{1}{na^2(1-e^2)^{\frac{1}{2}}\sin i}\frac{\partial R}{\partial i},\\
\frac{dM}{dt} &= n - \frac{1-e^2}{na^2e}\frac{\partial R}{\partial e} - \frac{2}{na}\frac{\partial R}{\partial a}.
\end{aligned}\right\} \qquad (14)$$

Here, as in the previous problems, the operations on the right-hand sides of the equations (14) are not be performed directly on the disturbing potential R but on the reproducing kernel with respect to the first point P or the second point Q. If the kernel is given explicitly, e. g. as $\alpha \cdot \frac{2R}{L}$, it is not difficult to express it by the elements corresponding to the two points P and Q and perform the differentiations. If the kernel contains correction members in the form of spherical harmonics, the differential coefficients with respect to the elements can be derived in the traditional way or by using the generalized spherical harmonics [12], [3].

Now, if we could measure directly the rates of variation of the elements in short intervals of time, then the problem would be solved, but we can only find the resulting perturbations during long intervals of time, and, consequently, we have to use integrations or mean values. This situation, however, is not pecular to our problem, so it should be possible to overcome also that difficulty.

The most severe draw-back of the method is that it results in very large systems of normal equations – one equation for each measurement – and that these equations are not sparse, as are for instance the normal equations used for the adjustment of geodetic networks. It is a consolation that the matrix of the normal equations is positive definite, so that the equations may be solved without using pivoting, and that the adjustment procedure – like adjustment procedures in general – is relatively simple to automatize.

I believe it is necessary to find some trick that may reduce the number of normal equations or at least the number of coefficients different from zero; I have some ideas in this respect, but I think it is too early to go into computational details.

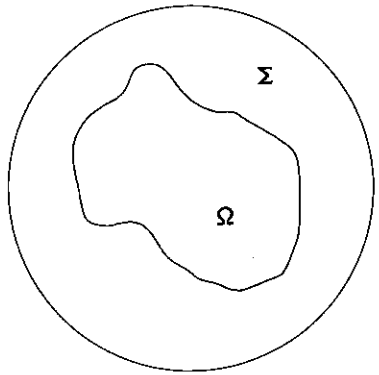
The adjustment technique introduced here is a typical data-processing method giving a formally correct result which is absolutely independent of the meaning given (or not given) to the input data. It is in my opinion a dangerous draw-back of this method, as of all adjustment methods, that it gives an answer to even the most foolish question if it is only asked in a formally correct way.

Therefore, I hope that the time gained by this and other forms of mechanization of tedious calculations will not be used exclusively for the production of more figures but for a better formulation of the problem, so that the questions we ask may be more realistic, from the physical as well as from the numerical point of view. I think that some of the thoughts expressed in this paper may be helpful in that respect.

Appendix

Proof of Runge's Theorem

I want to prove that any potential regular in an open bounded region Ω can be approximated by potentials regular in an open sphere Σ containing Ω in its interior. The region Ω is supposed to be bounded by a surface ω which is sufficiently regular, e. g. by having a finite curvature all over. This condition could be weakened very much, but I do not think it would be of much interest in this connection. It is, however, important that $\Sigma - \Omega$ is connected.



The theorem will be proved through a series of lemmas.

Lemma 1. For every function f continuous in $\Omega + \omega$ and 0 on ω we have that from

$$\int_{\Omega} f \cdot \psi d\Omega = 0, \tag{1}$$

for all potentials ψ regular in Σ follows

$$\int_{\Omega} f \cdot \varphi d\Omega = 0, \tag{2}$$

for all potentials φ regular in Ω.

It is obvious that, if we can prove the lemma for certain potentials ψ regular in Σ, then the lemma will be true as it stands. The set of ψ I shall use is that represented by continuous single layer distributions on the surface σ of Σ, i. e. for potentials representable by

$$\psi_P = \int_{\sigma} \frac{1}{r_{PQ}} \varkappa(Q) d\sigma_Q, \tag{3}$$

where

$$P \in \Sigma, \quad Q \in \sigma$$

and r_{PQ} is the distance between the points P and Q. $\varkappa$ is the density of the single layer distribution on σ.

For (1) to be satisfied for all ψ represented by (3) it is necessary and sufficient that

$$F(Q) \equiv \int_{\Omega} f(P) \frac{1}{r_{PQ}} d\Omega_P = 0 \quad \text{for all } Q \in \sigma. \tag{4}$$

F is here a regular potential in the space outside ω. In Ω F is a solution to the Poisson equation $\Delta F = f$.

Now (4) says that F is zero on the surface of the sphere and that it has the finite mass

$$M = \int_{\Omega} df \Omega, \tag{5}$$

i. e. F is zero at infinity; therefore it will vanish in the space outside the sphere, but then it must be zero in the whole space outside ω and on ω.

For every potential φ regular in Ω we may, consequently, write:

$$\int_{\Omega} f \cdot \varphi d\Omega = \int_{\Omega} \Delta F \cdot \varphi d\Omega = \int_{\Omega} F \cdot \Delta \varphi d\Omega = 0, \tag{6}$$

(here we have used Green's formula) and the lemma is proved.

Let us now consider the Hilbert space H consisting of functions, not necessarily potentials, f defined in Ω so that the integral

$$\|f\|^2 = \int_{\Omega} f^2 d\Omega \tag{7}$$

is finite. The scalar product in H is

$$\langle f_1, f_2 \rangle = \int_{\Omega} f_1 \cdot f_2 d\Omega. \tag{8}$$

Since functions of the type of f in Lemma 1 are dense in H, we have

Lemma 2. In the Hilbert space H any element orthogonal to every potential ψ regular in Σ and restricted to Ω is orthogonal to every element φ of H that is a regular potential in Ω.

Now, according to the elementary theory of Hilbert spaces lemma 2 implies:

Theorem 1. Any element φ of H which is a regular potential in Ω can be approximated *in the strong topology* in H by restriction to Ω of potentials ψ regular in Σ.

But what we wanted was not a theorem on approximation in the strong topology in H, but in the uniform topology on all closed subsets of Ω.

If for a moment we assume that the elements φ of H which are regular potentials in Ω form a Hilbert space, say H_0, with the reproducing kernel $K(P, Q)$, then we can deduce the following theorem using a technique used already in chapter III.

Theorem 2. Any element φ of H_0 (i. e. any φ of H which is a regular potential in Ω) can be approximated uniformly on all closed subsets of Ω by restriction to Ω of potentials ψ regular in Σ, so that also any derivative of φ with respect to the coordinates in Ω is uniformly approximated by the corresponding derivatives of the ψ's on the same closed subsets of Ω.

Corollary. By using inversion with respect to the sphere Σ this theorem gives a strengthened form of Runge's theorem, at least for $\varphi \in H_0$.

Proof of theorem 2: For a function $f(P)$ that is an element of a Hilbert space with reproducing kernel $K(P, Q)$, we have (cf. (III. 17)):

$$\left.\begin{aligned} |f(P)| &= |\langle f(Q), K(Q,P)\rangle| \leq \|f\| \cdot \|K(Q,P)\| \\ &= \|f\| \cdot \langle K(P,Q), K(Q,P)\rangle_Q^{\frac{1}{2}} = \|f\| \cdot K(P,P)^{\frac{1}{2}} \end{aligned}\right\} \tag{9}$$

and

$$\left.\begin{aligned} \left|\left(\frac{\partial f}{\partial x_P}\right)_P\right| &= \left|\left\langle f(Q), \frac{\partial}{\partial x_P} K(Q,P)\right\rangle\right| \leq \|f\| \cdot \left\|\frac{\partial}{\partial x_P} K(Q,P)\right\|_Q \\ &= \|f\| \cdot \left\langle \frac{\partial K(P,Q)}{\partial x_P}, \frac{\partial K(P,Q)}{\partial x_P}\right\rangle_Q^{\frac{1}{2}} \end{aligned}\right\} \tag{10}$$

as well as similar formulae for the higher derivatives. From the properties of the reproducing kernel and the boundedness of Ω it follows that

$$\langle K(P,Q), K(Q,P)\rangle_Q^{\frac{1}{2}}$$

is finite. From the same premises and from the fact that $K(P, Q)$, a function of P or Q, is a potential regular in Ω it follows that the same applies to

$$\left\langle \frac{\partial K(P,Q)}{\partial x_P}, \frac{\partial K(P,Q)}{\partial x_P}\right\rangle_Q^{\frac{1}{2}}$$

and also to the higher derivatives of the kernel.

Now, from theorem 1 it follows that, given a $\varphi \in H_0$, we can find a sequence $\{\varphi_n\}$ of potentials ψ_n regular in Σ so that for every $\varepsilon > 0$ there is such an N that

$$\|\varphi - \psi_n\| < \varepsilon \quad \text{for } n > N. \tag{11}$$

By putting $f = \varphi - \psi_n$ theorem 2 follows from (9), (10), etc.

In order to get rid of the restriction that φ must have a finite H-norm, we may use the spaces H_p instead of H. For any twice continuously differentiable positive function defined on Ω H_p is given by the scalar product

$$\langle f, g\rangle_p = \int_{\Omega} p(P) \cdot f(P) \cdot g(P) d\Omega \tag{12}$$

and the corresponding norm

$$\|f\|_p = \left(\int_{\Omega} p \cdot f^2 d\Omega\right)^{\frac{1}{2}}. \tag{13}$$

Given any potential φ regular in Ω, we can find such a p that

$$\varphi \in H_p.$$

(Take $p = (1 + \varphi^2)^{-\frac{1}{2}}$).

The reader is invited to prove lemma 1, lemma 2 and theorem 1 for H_p instead of H, which is quite simple. Then the restriction is removed and Runge's theorem is proved as soon as we have proved:

Lemma 3. The subset of H_p consisting of potentials regular in Ω is a Hilbert space with reproducing kernel.

But we shall first prove

Lemma 4. Any Hilbert space consisting exclusively of potentials regular in Ω has a reproducing kernel.

If φ is an element of the Hilbert space in question and P is a fixed point in Ω, then $\varphi(P)$ is finite. The linear operator A_P from the given Hilbert space to the real numbers which to φ assigns the value $\varphi(P)$ of φ at P is therefore defined for all φ of the Hilbert space. From a well-known theorem from functional analysis (Hellinger and Toeplitz' Theorem or the Closed Graph Theorem) it follows that A_P is a bounded operator or, in our case where the range is the real numbers, a bounded linear functional, and from this it follows again from one of the fundamental theorems of Hilbert spaces with reproducing kernel that the Hilbert space in question has a reproducing kernel.

As to the proof of lemma 3 it remains only to be shown that the subset of H_p consisting of potentials regular in Ω forms a Hilbert space, i. e. a *closed* linear subspace of H_p. That it is a linear subspace is evident; the only difficulty is to prove that the subspace is closed.

Let us use M to denote the set of twice continuously differentiable functions defined on Ω and zero outside some closed subset of Ω. It is evident that $M \in H_p$ and that any function of H_p which is orthogonal to every $f \in M$ is equivalent to zero, so that a necessary and sufficient condition of $\varphi \in H_p$ being a regular potential in Ω is that

$$\int_{\Omega} p \cdot f \Delta \varphi d\Omega = 0 \quad \text{for all } f \in M. \tag{14}$$

From (14) follows

$$\int_{\Omega} \Delta(p \cdot f) \cdot \varphi d\Omega = 0 \quad \text{for all } f \in M, \tag{15}$$

Green's formula having been used.

If φ is twice differentiable, (14) follows from (15), and as the $\varphi \in H_p$ for which (15) holds form a closed linear subset of H_p, lemma 3 follows from the famous Weyl's lemma, which shows that the φ for which (15) holds are not only twice but arbitrarily often differentiable.

Finally, I shall outline a short proof of Weyl's lemma.

Let S be the unit sphere, and let us suppose that $\Phi(P)$ is an n times continuously differentiable function which is zero outside S and which depends only on the distance of P from the origin. Let us also suppose that

$$\int_S \Phi d\Omega = 1. \tag{16}$$

For every function $u(P) \in H_p$ and every $\varepsilon > 0$ we can now define:

$$u_\varepsilon(P) = \int_S \Phi(Q) u(P - \varepsilon Q) d\Omega_Q = \varepsilon^{-3} \int \Phi\left(\frac{P - Q}{\varepsilon}\right) u(Q) d\Omega_Q, \tag{17}$$

where the latter integral is taken over the domain where the integrand is different from zero.

It is easy to prove that u_ε is n times continuously differentiable, and that

$$\lim_{\varepsilon \to 0} u_\varepsilon(P) = u(P) \tag{18}$$

almost everywhere.

We shall now prove that, given two positive numbers ε_1 and ε_2, it follows from (15) that

$$\varphi_{\varepsilon_1}(P) = \varphi_{\varepsilon_2}(P) \tag{19}$$

for all $P \in \Omega$ so that the distance from P to the boundary of Ω is larger than both ε_1 and ε_2.

Let us define the function F by

$$F(P) = \int \left\{ \varepsilon_1^{-3} \Phi\left(\frac{P - Q}{\varepsilon_1}\right) - \varepsilon_2^{-3} \Phi\left(\frac{P - Q}{\varepsilon_2}\right) \right\} \frac{1}{r_Q} d\Omega_Q. \tag{20}$$

For F we have immediately

$$\Delta F(P) = \varepsilon_1^{-3} \Phi\left(\frac{P}{\varepsilon_1}\right) - \varepsilon_2^{-3} \Phi\left(\frac{P}{\varepsilon_2}\right) \tag{21}$$

and

$$F(P) = 0 \tag{22}$$

for P outside the spheres with centres at the origin and radii ε_1 and ε_2.

Therefore, for a given $Q \in \Omega$ and for ε_1 and ε_2 sufficiently small $\Delta F(P - Q)$ is a function of P in the set M, and so (15) implies:

$$\int F(P - Q) \varphi(Q) d\omega_Q = 0, \tag{23}$$

which in its turn is equivalent to (19).

Equations (18) and (19) now give

$$\varphi_\varepsilon(P) = \varphi(P) \tag{24}$$

almost everywhere for ε sufficiently small (dependent on P).

But a consequence of (24) is that φ is equivalent to a function which is n times continuously differentiable. For $n = 2$ this means that we may deduce (14) from (15), and we have Weyl's lemma.

References

[1] Davis, P. J.: Interpolation and Approximation. – New York, 1963.

[2] Fichera, G.: Linear Elliptic Differential Systems and Eigenvalue Problems. – Berlin 1965.

[3] Gelfand, J. M. and Z. Šapiro: Representations of the Group of Rotations of 3-dimensional Space. – American Mathematical Society Translations, Series 2, Vol. 2. p. 207–316.

[4] Heiskanen, W. A. and H. Moritz: Physical Geodesy. – 1967.

[5] Hörmander, L.: Linear Partial Differential Operators. – Berlin 1964.

[6] Kaula, W. M.: Theory of Satellite Geodesy. – San Francisco, 1966.

[7] Kaula, W. M.: Statistical and Harmonic Analysis of Gravity. – Journal of Geophysical Research, Vol. 64, 1959.

[8] Lavrientiev, M. M.: Some Improperly Posed Problems of Mathematical Physics. – Berlin 1967.

[9] Meschkowski, H.: Hillbertsche Räume mit Kernfunktion. – Berlin 1962.

[10] Moritz, H.: Uber die Konvergens für das Aussenraumpotential an der Erdoberfläche. – Österreichische Zeitschrift für Vermessungswesen, 49. Jahrgang, Nr. 1.

[11] Moritz, H.: Schwerevorhersage und Ausgleichungsrechnung. – Zeitschrift für Vermessungswesen, 90 Jahrgang 1965.

[12] Stiefel, E.: Expansion of Spherical Harmonics on a Satellite Orbit Based on Spinor Algebra. – Mathematische Methoden der Himmelsmechanik und Astronautik. Mannheim 1965. p. 341–350.