10 Least-squares collocation

10.1 Principles of least-squares collocation

The principle of collocation is very simple. The anomalous potential T outside the earth is a harmonic function, that is, it satisfies Laplace's differential equation

$$\Delta T = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = 0 \, . \tag{10–1}$$

An approximate analytical representation of the external potential T is obtained by

$$T(P) \doteq f(P) = \sum_{k=1}^{q} b_k \, \varphi_k(P) \, , \tag{10–2}$$

a linear combination f of suitable base functions $\varphi_1, \varphi_2, \ldots, \varphi_q$ with appropriate coefficients b_k. All these are functions of the space point P under consideration.

As T is harmonic outside the earth's surface, it is natural to choose base functions φ_k which are likewise harmonic, so that

$$\Delta \varphi_k = 0 \, , \tag{10–3}$$

in correspondence to (10–1).

There are many simple systems of functions satisfying the harmonicity condition (10–3), and thus we have many possibilities for a suitable choice of base functions φ_k. We might, for instance, choose spherical harmonics or potentials of suitably distributed point masses, depending on whether we emphasize global or local applications.

The coefficients b_k may be chosen such that the given *observational values are reproduced exactly* – for instance, all deflections of the vertical in a given area. This means that the assumed approximating function f in (10–2) gives the same deflections of the vertical at the observation stations as the actual potential and hence may well be considered a suitable approximation for T. Let us now try to put these ideas into a mathematical form.

Interpolation

Let errorless values of T be given at q spatial points $P_1, P_2, \ldots, P_q$; these points may lie on the earth's surface or in space above the earth's surface. We put

$$T(P_i) = f_i \, , \quad i = 1, 2, \ldots, q \tag{10–4}$$

and postulate that in approximating $T(P)$ by $f(P)$, the observations (10–4) will be reproduced exactly. The condition for this is

$$\sum_{k=1}^{q} b_k\, \varphi_k(P_i) = T(P_i) = f_i\,, \tag{10–5}$$

which may be written as a system of linear equations

$$\sum_{k=1}^{q} A_{ik}\, b_k = f_i \quad \text{with} \quad A_{ik} = \varphi_k(P_i) \tag{10–6}$$

or in matrix notation

$$\mathbf{A\,b} = \mathbf{f}\,. \tag{10–7}$$

If the square matrix $\mathbf{A}$ is regular, then the coefficients b_k are uniquely determined by

$$\mathbf{b} = \mathbf{A}^{-1}\,\mathbf{f}\,. \tag{10–8}$$

This model is suitable, for instance, for a determination of the geoid by satellite altimetry, since this method, rather directly, yields geoidal heights N_i and hence, by Bruns' theorem (2–236), $T(P_i) = \gamma_i\, N_i$. For the astrogeodetic geoid determination, we must generalize this model, which leads us to collocation.

Collocation

Here we wish to reproduce, by means of the approximation (10–2), q measured values which again are assumed to be errorless (this assumption is not essential and will be dropped later). These measured values are assumed to be linear functionals $L_1T, L_2T, \ldots, L_qT$ of the anomalous potential T. "Linear functional" means nothing else than a quantity LT that depends linearly on T but need not be an ordinary function but may, say, also contain a differentiation or an integral; essentially, it is the same as a "linear operator".

In fact, deflections of the vertical,

$$\xi = -\frac{1}{\gamma}\,\frac{\partial T}{\partial x}\,, \qquad \eta = -\frac{1}{\gamma}\,\frac{\partial T}{\partial y}\,, \tag{10–9}$$

but also gravity anomalies,

$$\Delta g = -\frac{\partial T}{\partial z} - \frac{2}{R}\,T\,, \tag{10–10}$$

and gravity disturbances

$$\delta g = -\frac{\partial T}{\partial z} \tag{10–11}$$

are such linear functionals of T; here, x, y, z denotes a local coordinate system in which the z-axis is vertical upwards and the x- and y-axes are directed towards north and east, and $R = 6371$ km is a mean radius of the earth. Equation (10–9) is a consequence of equations such as (2–377), with $\partial s = \partial x$ or ∂y; normal gravity γ may be considered constant with respect to horizontal derivation. Equation (10–10) is the well-known fundamental equation of physical geodesy in spherical approximation (2–263). Equations (10–9) and (10–10) refer to the earth's surface.

To repeat, by saying that deflections of the vertical and gravity disturbances and anomalies are linear functionals of T, we simply indicate the fact that $\xi, \eta, \delta g, \Delta g$ depend on T by the expressions (10–9) and (10–10), which clearly are linear; they are the linear terms of a Taylor expansion, neglecting quadratic and higher terms. In the above notation $L_i T$, the symbol L_i denotes, for instance, the operation

$$L_i = \frac{1}{\gamma} \frac{\partial}{\partial x} \tag{10–12}$$

applied to T at some point.

Putting

$$L_i f = L_i T = \ell_i \tag{10–13}$$

and substituting (10–2), we get

$$\sum_{i=1}^{q} B_{ik}\, b_k = \ell_i \quad \text{with} \quad B_{ik} = L_i \varphi_k\,, \tag{10–14}$$

where $L_i \varphi_k$ denotes the number obtained by applying the operation L_i to the base function φ_k; the coefficient B_{ik} obtained in this way does not depend on the measured values. Equation (10–14) is a linear system of q equations for q unknowns, which is quite similar to (10–6). This method of fitting an analytical approximating function to a number of given linear functionals is called collocation and is frequently used in numerical mathematics.

It is clear that interpolation is a simple special case of collocation in which

$$L_i f = f(P_i) \tag{10–15}$$

is the "evaluation functional", giving the value of f at a point P_i. Thus we see that in both interpolation and collocation the coefficients b_k require the solution of a linear system of equations (which in general will not be symmetric).

Least-squares interpolation

Let us consider a function

$$K = K(P, Q)\,, \tag{10–16}$$

in which two points P and Q are the independent variables. Let this function K be

- symmetric with respect to P and Q,
- harmonic with respect to both points, everywhere outside a certain sphere, and
- positive-definite (the positive definitiveness of a function is defined similarly as in the case of a matrix).

Then the function $K(P, Q)$ is called a (harmonic) kernel function (Moritz 1980 a: p. 205). A kernel function $K(P, Q)$ may serve as "building material" from which we can construct base functions. Taking for the base functions the form

$$\varphi_k(P) = K(P, P_k)\,, \tag{10–17}$$

where P denotes the variable point and P_k is a fixed point in space, we obtain *least-squares interpolation* already treated by a quite different approach in Chap. 9.

This name originates from the statistical interpretation of the kernel function as a *covariance function* (Sect. 9.2); then least-squares interpolation has some minimum properties (least-error variance, similarly as in least-squares adjustment). This interpretation is not essential, however; one may also work with arbitrary analytical kernel functions, considering the procedure as a purely analytical mathematical approximation technique. Normally one tries to combine both aspects in a reasonable way.

Substituting (10–17) into (10–6), we get

$$A_{ik} = K(P_i, P_k) = C_{ik}\,; \tag{10–18}$$

this square matrix now is symmetric (in the general case, A_{ik} is not symmetric!) and positive definite because of the corresponding properties of the function $K(P, Q)$. Then the coefficients b_k follow from (10–8) and may be substituted into (10–2). With the notation

$$\varphi_k(P) = K(P, P_k) = C_{Pk}\,, \tag{10–19}$$

the result may be written in the form

$$f(P) = \begin{bmatrix} C_{P1} & C_{P2} & \dots & C_{Pq} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1q} \\ C_{21} & C_{22} & \dots & C_{2q} \\ \vdots & \vdots & & \vdots \\ C_{q1} & C_{q2} & \dots & C_{qq} \end{bmatrix}^{-1} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_q \end{bmatrix}, \tag{10–20}$$

formally identical with Eq. (9–67) obtained in a completely different way.

Least-squares collocation

Here we again derive the base functions from a kernel function $K(P,Q)$, but in a way slightly different from (10–17): we put

$$\varphi_k(P) = L_k^Q K(P,Q)\,, \tag{10–21}$$

where L_k^Q means that the functional L_k is applied to the variable Q; the result no longer depends on Q (since the application of a functional results in a definite number). Thus, in (10–14) we must put

$$B_{ik} = L_i^P L_k^Q K(P,Q) = C_{ik}\,, \tag{10–22}$$

which gives a matrix which again is symmetric. Solving (10–14) for b_k and substituting into (10–2) gives with

$$\varphi_k(P) = L_k^Q K(P,Q) = C_{Pk} \tag{10–23}$$

the formula

$$f(P) = \begin{bmatrix} C_{P1} & C_{P2} & \dots & C_{Pq} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1q} \\ C_{21} & C_{22} & \dots & C_{2q} \\ \vdots & \vdots & & \vdots \\ C_{q1} & C_{q2} & \dots & C_{qq} \end{bmatrix}^{-1} \begin{bmatrix} \ell_1 \\ \ell_2 \\ \vdots \\ \ell_q \end{bmatrix} . \tag{10–24}$$

This is formally the same expression as (10–20), but with f_i replaced by ℓ_i and with "covariances" C_{ik} and C_{Pi} defined by *"covariance propagation"* (10–22) and (10–23). The concept of covariance propagation is a straightforward generalization of the formal structure of error propagation known from adjustment computations. However, this structure as such is purely mathematical rather than statistical. We know that a "linear functional" is the continuous analogue (in infinite-dimensional Hilbert space) to the usual concept of a linear function in n-dimensional vector space. We try not to burden the reader with too much mathematical formalism, but this is treated in great detail in Moritz (1980 a) and in Moritz and Hofmann-Wellenhof (1993: Chap. 10). We cannot, however, resist the temptation to compare the structure

$$b_i = L_i^j a_j \tag{10–25}$$

leading to

$$\operatorname{cov}(b_i, b_j) = L_i^k L_j^l \operatorname{cov}(a_k, a_l) \tag{10–26}$$

for finite-dimensional vectors a and matrix L using the usual summation over two equal indices, and $N_i = L_i^P \Delta g_P$ leading to

$$\mathrm{cov}(N_P, N_Q) = L_i^P L_j^Q \,\mathrm{cov}(\Delta g_P, \Delta g_Q)\,, \tag{10–27}$$

where N_i denotes the geoidal height at point i and Δg is the gravity anomaly at point P, and L denotes the Stokes formula. Explicit expressions are found in Moritz (1980 a: Sect. 15).

In this statistical interpretation, we take the kernel function $K(P, Q)$ as the covariance function $C(P, Q)$. Then $f(P)$ is an optimal estimate (in the sense of least variance) for the anomalous potential T and hence for the height anomaly $\zeta = T/\gamma$, on the basis of arbitrary measurement data. For geoid determination in mountainous areas, relevant terrestrial measurement data primarily are ξ, η, and Δg. The covariances C_{ik} and C_{Pi} are given by known analytical expressions, see Tscherning and Rapp (1974) or Moritz (1980 a: Sect. 15). A general computer program for collocation is described in Sünkel (1980).

Least-squares collocation may easily be generalized to observational data affected by random errors; systematic effects may also be taken into consideration. In addition to the estimated quantities (f in our present case) we may also compute their standard error by a formula similar to (10–24). A comprehensive presentation of a least-squares collocation may be found in Moritz (1980 a). *You cannot learn collocation from this slight chapter only!*

Harmonicity of the covariance functions.

In three-dimensional space, the covariance functions, being kernel functions and their linear functional transformations L, are always harmonic. If we have (9–25),

$$C(\psi) = \sum_{n=2}^{\infty} c_n P_n(\cos\psi) \tag{10–28}$$

on the sphere, then in space there will be

$$C(r, r', \psi) = \sum_{n=2}^{\infty} c_n \left(\frac{R^2}{r\,r'}\right)^{n+2} P_n(\cos\psi) \tag{10–29}$$

(Moritz 1980 a: Sect. 23, Eq. (32-1)). The point $P(r, \theta, \lambda)$ is the computation point, and $Q(r', \theta', \lambda')$ is a current data point; ψ is the spherical distance between (θ, λ) and (θ', λ'), and R is the mean radius of the earth. The dependence on r is given by the factor

$$r^{-(n+2)} \tag{10–30}$$

because $r\Delta g$ is harmonic, and similarly for r'. The factor

$$\left(\frac{R^2}{r\,r'}\right) \tag{10–31}$$

is chosen to become equal to 1 if both points P and Q lie on sea level; in this case, Eq. (10–29) reduces to (10–28).

So, each of the terms of (10–29) is harmonic, that is, it satisfies Laplace's equation. Thus, the whole series (10–29) is harmonic (if it converges), being a linear combination of harmonic terms. This is a well-known consequence of the linearity of Laplace's equation: the linear combination of solutions of any linear equation is itself a solution of this equation.

Thus, also the spherical harmonics series of $T = r\Delta g$ is harmonic down to the reference sphere $r = R$, with respect both to r and r'. Harmonic functions, by their very definition, are *regular analytic functions* down to $r = R$, so T and all its linear combinations are regular and thus *admit downward continuation down to the reference sphere* (cf. Sect. 8.6).

10.2 Application of collocation to geoid determination

It is well known that the direct interpolation of free-air gravity anomalies, which essentially are surface gravity anomalies (8–128) in high mountains, e.g., by least-squares interpolation, leads to relatively poor results because of the correlation of the free-air anomalies with elevation (Sect. 9.7). This correlation with elevation constitutes a considerable trend which must be removed before the interpolation. Bouguer anomalies take care of the dependence on the local irregularities of elevation; isostatic anomalies are, in addition, also largely independent on the regional features of topography; in Sect. 11.1 we shall consider, in addition, also the removal of global trends by spherical-harmonic earth gravity models (e.g., EGM 96, see www.iges.polimi.it/index/geoid_repo/global_models.htm) obtainable from the internet.

In exactly the same way we must remove the main trend of the vertical deflections ξ, η and the gravity anomalies Δg by an isostatic reduction before applying collocation. Thus, isostatic reduction, pragmatically regarded as trend removal, is essential for the practical application of least-squares collocation in mountainous regions (Forsberg and Tscherning 1981).

Physically speaking, we transport the topographic masses to the interior of the geoid in such a way that the isostatic mass deficiencies are filled. The observation point P remains in its position on the earth's surface. In this way, not only the harmonic character of the anomalous potential T outside the earth's surface is preserved, but in addition, the computational removal of the topographic masses above sea level makes the function T harmonic down to sea level. Hence, the collocation formula (10–24) can be applied also at sea level, giving cogeoid heights N^c. By applying the inverse reduction (the indirect effect) to the computed height anomalies ζ^c and cogeoid heights N^c, we get actual ζ and N. It can be expected that errors in the isostatic model used (e.g., an Airy–Heiskanen model) will largely cancel in this combined procedure of reduction and "anti-reduction" (remove-restore technique; see Sect. 11.1).

The procedure is theoretically optimal and practically well suited for computer use. The integrability conditions, which in Helmert integration are represented by the closures of the individual triangles (see Sect. 5.14), are automatically taken into account. The fact that the deflections of the vertical are given only in a certain region has the effect that the geoid can only be computed in that region. Since, even by collocation, differences in geoidal heights between two neighboring stations A and B depend essentially only on the deflections in those two stations, the lack of data outside the region under consideration will hardly cause a noticeable distortion. Note, however, that the addition of a constant to all geoidal heights N will not affect the deflections of the vertical; hence, astrogeodetic data determine the geoidal heights only up to an additive constant. This constant may be chosen such that the average value of the computed N is zero, and the result of collocation comes near to this case.

To get immediately almost geocentric geoidal heights, it is appropriate to take into consideration a global trend which mainly affects ζ and N itself, by subtracting the effect of a suitable global gravity field, e.g., the gravity earth model given as a spherical-harmonic expansion up to degree $180° \times 180°$ of Rapp (1981), say, following Sünkel (1983). This will be described in the next section; in the present section we limit ourselves to the isostatic reduction.

Computational procedure

The computational procedure consists of the following steps:

1. Transformation of the astrogeodetic surface deflections ξ, η from the local datum used for the geocentric Geodetic Reference System 1980 by the well-known differential formulas of Vening Meinesz (see Heiskanen and Moritz 1967: Eq. (5-59)). This is necessary since collocation requires a reference system which is as realistic as possible.

2. Application of the normal plumb line curvature (8–137) to the "geometric" surface deflections ξ, η gives the "dynamic" surface deflections

$\bar{\xi}$, $\bar{\eta}$ by (8–136).

3. Computation of the gravity anomalies Δg, also referred to the earth's surface according to (8–128).

4. The topographic-isostatic reduction of $\bar{\xi}$, $\bar{\eta}$, Δg by (8–154) and (8–101) gives values ξ^c, η^c, Δg^c which continue to refer to the surface point P.

5. The application of collocation to ξ^c, η^c, Δg gives height anomalies ζ^c and cogeoid heights N^c, by simply varying the elevation parameter (h and zero, respectively) in the collocation program (see Sünkel 1983).

6. By applying the indirect effect (10–2) and (8–153), we get actual height anomalies ζ and geoidal heights N.