FUNCTIONAL METHODS FOR GRAVITY FIELD APPROXIMATION

by

C.C. Tscherning

Geodetic Institute
Gamlehave Allé 22
DK-2920 Charlottenlund, Denmark

Lecture Notes in Earth Sciences, Vol. 7
Mathematical and Numerical Techniques in Physical Geodesy
Edited by H. Sünkel

5. Choice of base functions and inner product

5.1 Introductory remarks.

Suppose we have a set of observations $y_i = L_i(T)$, $i = 1,....n$. Then we have identified 3 situations, namely corresponding to whether we use a number of base functions m with $n > m$, $n = m$ and $n < m$. The last, underdetermined case, is generally reduced to one of the two other cases, using a singular value decomposition of the normal equations and then selecting as base functions the eigenfunctions with eigenvalues numerically larger than a fixed constant.

The selection of base functions may then be done using one of the following considerations:

(1) An approximation is wanted, which describe the gravity field down to a certain resolution, e.g. 500 km, over a given area, despite that maybe the data contains information on even shorter wavelengths.

(2) An approximation is wanted, which represents the data in the best possible manner or which extracts the maximal possible information from the data.

In the first case can we normally have $m < n$ and in the last case $n \geq m$, (which we here will consider equivalent to $n = m$).

5.2 Choice of base functions.

If a global approximation is needed, then case (1) may lead us to use (linear combinations of) spherical harmonics to wavelengths corresponding to the resolution. Or an intermediate solution, containing shorter wavelengths, is constructed and the longer wavelength information is then extracted from this.

In both cases large systems of equations may have to be solved. But if data are globally distributed in a regular manner (equal spacing in longitude for each fixed degree of latitude), the equations may have a Toplitz structure, see Colombo (1979), or be strongly diagonal dominat see Wenzel (1985). In both cases very large systems of equations may be solved with a relatively minor effort.

Also it is possible to use finite elements, taking advantage of the sparceness of the normal equations as discussed in Meissl (1981). However finite elements are not harmonic, so additional equations have to be solved in order to model this property of T.

If a local area is considered, methods similar to spline function or finite elements have been proposed. A set of bell-shaped (generally harmonic) functions, regularity distributed with a spacing equal to the resolution may be used. If harmonic functions are used, we loose some of the advantages which made the original spline functions and finite elements a good choice, namely the property that they were

zero outside a finite interval. The systems of equations to be solved will be full, but a quasi-ortogonallity may be achieved by a careful selection of the functions, which permit large systems of equations to be solved iteratively in very few steps.

Example 5.1 Potentials of point masses may be used as base functions. However they are positive everywhere. If a number of the first terms in the Legendre expansions are eliminated, then the new function will have several zero-points, cf. example 3.11. ●

As actual base functions potentials of point masses or harmonic kernel functions (see example 3.1, 3.3, 3.5, 3.10) have been proposed and used, see Lelgemann (1981). Depending on the data available, it was proposed to use representers of the associated functionals, distributed in a regular grid covering the actual area. If for example the component of the gravity vector is used, then the three functions are orthogonal in each point, if a rotational invariant kernel is used, see Tscherning (1970).

In a first choice of base functions generally only a class of functions are selected. For point mass potentials or kernel functions the depth to the point mass, or the radius of a sphere bounding the set of harmonicity, respectively, may then be determined subsequently. Here a generally non-linear optimization problems must be solved (see (Barthelmes,1986)). For regularly spaced data and certain classes of kernels rules to determine a best "depth" has been found, see Hardy and Göpfert (1975) and Lelgemann (1981). These rules solve the problem that most of the kernels used have strong singularities, i.e. they take on infinite values. Typically, if a point mass is located too close to the Earth's surface, its generated potential will go too fast to ∞ and the determined approximation will look as illustrated in Fig. 3.

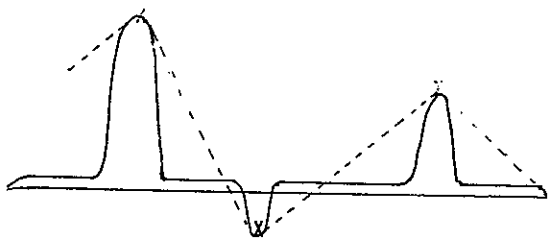

Fig. 3. Approximation obtained using too singular base functions. - and with base functions assuring an exact mid-point interpolation... .

Naturally also the opposite phenomenon may occur, well known from polynomial approximation, that the function has oscillations with large amplitudes between the

data-points. On the other hand we also know how to counteract this phenomenon, namely by using a minimum norm condition, and base functions of a different kind.

In the second case, n=m, the choice of base functions is fixed, if we use minimum norm collocation. Here the representers $K(L_i, P)$ of the linear functionals, L_i, must be used. In some other cases the base functions are selected, because they occur when discretizising an integral formula like Poissons or Stokes equation relating data at the Earth's surface with data in space, cf. some of the methods proposed by Bjerhammar. However, the kernels in most of these integral formulae are reproducing kernels (see section 3.3), so we have in practice a very little difference to the use of minimum norm collocation in the space where these kernels are reproducing kernels or covariance functions for that sake.

Since the integral formulae normally are solutions to boundary value problems with data on a sphere, then the base functions used will be associated with points on the sphere. This sphere will, when being used for selecting base functions, generally have a radius smaller than the Earth's mean radius, and the coefficients then look like "artificial" data which must be estimated explicitly on the sphere, (see e.g. Bjerhammar, 1976). In this way we avoids the problem of the possible singularity of the kernels. If data are not on a sphere, then a non-symmetric system of equations must be solved in order to find the coefficients of the linear combination. But if they are on a sphere, then we have symmetry and an exact equivalence to a minimum norm collocation solution.

Example 5.2 Point mass modelling with carrier points right under the data points on a sphere with radius R.

Then $\tilde{T} = \Sigma a_i / |P - P_i'|$, where P_i' is right under the data point P_i. Suppose the observations are the values $T(P_i)$. Then the equations to be solved have the coefficients

$$1/|P_i' - P_j| = \frac{1}{R} \sum_{k=0}^{\infty} \left(\frac{R}{r_j}\right)^{k+1} P_k(\cos\psi_{ij}) \tag{5.1}$$

If $r_i = r_j = r$ (all data points on a sphere), and $R_B^2 = R \cdot r$, then

$$1/|P_i' - P_j| = \frac{1}{R} \sum_{k=0}^{\infty} \left(\frac{R_B^2}{r^2}\right)^{k+1} P_k(\cos\psi_{ij}). \tag{5.2}$$

In this case the solution is equivalent to the minimum norm collocation solution in the space with the Krarup-kernel, example (3.5). This gives a nice interpolation if the point mass modelling technique as a general approximation method also for non-gravity field related data. The solution will have a maximal smoothness in terms of minimal first order derivatives. ●

Since we in many cases will have a similar close equivalence between the two kinds of collocation solutions, we will in the following only consider the minimum-norm type of solutions.

5.3 Choice of inner product.

For the overdetermined problem $n > m$, a solution is found as the "best" approximation of T on the subspace spanned by the base vectors. This involves, cf. eq. (4.5) an evaluation of several inner products, requiring that T or some of its derivatives are known. Hence, the inner product must depend on the data available. If only gravity data are available, then

$$(f, g)_{\frac{1}{2}} = \frac{1}{4\pi}\int_{\omega} \Delta g_f \cdot \Delta g_g \, d\omega \tag{5.3}$$

can be used. The corresponding base vectors are the normalized solid spherical harmonics multiplied by $\frac{R}{I-1}$ and the reproducing kernel

$$K(P, Q)_{\frac{1}{2}} = R^2 \sum_{i=2}^{\infty} \frac{2i+1}{(i-1)^2} \left(\frac{R^2}{rr'}\right)^{i+1} P_i(\cos\psi) \tag{5.4}$$

In practice the integrals of type (5.3) are not evaluated over the whole Earths surface, but only over a limited area. Also sometimes in practice, not the discretized integral is used, but one simply calculates the products of the available data with weights according to their error variances. It is easy to imagine how this can go wrong, if the data are not in some manner regularity distributed. The difficulty is obvious in the case where low order potential coefficients have to be estimated from satellite orbit pertubations. These pertubations are of a similar character as the potential itself (integrals along the orbit of accelerations = velocities) and integrals of the potential (a further integration of the velocities). The minimalization of the differences between observed and computed pertubations must then correspond to one of the negative Sobolev norms. Seen from a functional-analytic standpoint is such a norm much too weak, permitting functions with wildly behaving boundary values. In practice (Lerch et al., 1977, section 3.2) we then also see that the H_0-norm is minimalized simultaneously (the square-sum of the estimated potential coefficients). In the minimum norm collocation situation, $n = m$, the choice of inner product has a consequence, that the base functions are fixed, and the calculation of the inner products does not involve actual observed data. Then, we might use the advice given by for example approximation by spline functions, cf. (Moritz, 1978, p. 44), and select a norm minimalizing the second order derivatives.Now in spline function theory, only values of the approximating function are considered, while we in geodesy

frequently work with first or several order derivatives. We should consequently work with a norm involving the fourth-order derivatives at or close to the Earth's surface. Now, our primary goal in geodesy is to have a good representation of the anomalous potential itself and the gravity anomaly vector. Hence, maybe minimatization of third order derivatives would be reasonable. If the goal is a further geophysical interpretation in terms of density estimates, then the fourth order derivatives might be the right choice. On the other hand, if we as basic function space use a space of functions harmonic in a larger set than the set outside the Earth, then all derivatives are in principle minimalized.

The norms implied using the experiences from spline functions are strong norms. The linear functionals will have increasing correlations and numerical instabilities will be encountered when solving the normal equations. However, this may be counteracted by using a small depth to the Bjerhammar sphere and by removing or decreasing the weight of low degree subspaces. If a subspace corresponding to spherical harmonic of maximal degree i are removed, then the corresponding i first terms in the reproducing kernel are also removed. This has a tremendous effect on the kernel and thereby on the correlation of the various linear functionals. It is illustrated in (Tscherning ,1985, Fig. 5).

If we want a best least-squares approximation, then we also only have one choice: the empirical covariance function. The computation of this function is discussed at many places: Tscherning and Rapp (1974), Lachapelle and Schwarz (1980), Goad at al. (1984), Forsberg (1984) and Tscherning (1985). It is, fortunately, possible to model the covariance function using reproducing kernels, cf. Tscherning (1972), thus giving a compromise between least-squares and minimum norm collocation.

The reproducing kernels, which may be used to model the empirical covariance function(s), correspond to norms minimatizing the second order derivatives on the boundary of a (Bjerhammar) sphere a few kilometers inside the Earth, or third order derivatives over the space outside such a sphere, see Tscherning (1985, Table 1).

It would be interesting to investigate, whether the application of even stronger norms would give better results. Such norms would make it possible to skip the Bjerhammar sphere and we would work directly with the true boundary. In this case T would be an element of the same space as $\tilde{T}$, and the convergence of the collocation solutions towards T is assured for increasing amounts of data, see Tscherning (1978a).

Finally two arguments should be mentioned, which have been put forward, (but which I think are wrong). One argument is that the base functions should be as easy to compute as possible. However, all functions may be tabulized once and for all,and then later evaluated with the same speed as any other functions,cf. Sünkel (1979). The other argument is (see e.g. Barthelmes (1986)) that some basefunctions have an

immediate geophysical interpretation as giving estimates of the density. But as we have seen in section 4.6, then all harmonic functions may be given such an interpretation.

6. Conclusion - Choice of method

We have in chapter 4 described a number of methods and in chapter 5 discussed various choices of base functions and inner products. A comparison of the methods show that several satisfy the following criteria:

all kinds of data can be used,

all kinds of quantities can be predicted,

the methods are fairly easy to implement on a computer,

numerical stability problems can be avoided by proper selections of base functions (Bjerhammar sphere, removal of subspaces).

data noise can be taken into account,

non-gravity field related parameters may be estimated simultaneous with the approximation of T and density estimates may be derived from $\bar{T}$.

However, some of the methods require a considerable numerical effort, especially when solving the normal equations. But also this effort may be reduced, taking in into account repretitive pattens of the data, or by limiting the extend of a solution to a local area, and then using an integration method outside the area, see e.g. (Lachapelle and Tscherning, 1978). Also data selection can be used as described in (Goad et al., 1983) and (Barthelmes, 1986).

Furthermore, some methods delives error-estimates. Here it has been argued that such estimates may be found for all methods, simply by comparing a solution to data not used to construct the solution. But this does not solve the problem we want to solve, namely of knowing error estimates of other data types than these used or available.

Methods are primarily compared in terms of mean square differences between observed and computed quantities. If this is to be taken as the standard for comparison, then least-squares collocation must be the best method, because it has be proved that it (section 4.4) minimalizes exactly this error. But the proper use of this methods requires that a good estimate of the empirical covariance function is available. And this is a difficult quantity to estimate. Lauritzen (1973) has found that it under certain conditions is more difficult to estimate the covariance function than to estimate $\bar{T}$.
However also this paradox has been solved in the Lecture Notes by F. Sansò!

Hence, if it in practical comparisons is found that least squares collocation does not give the best results (see e.g. Bjerhammar, 1985), then the explanation must be that a wrong covariance function was used.

From theoretical considerations and practical experiences I am convinced that

least squares collocation is the best method for local gravity field approximation. The difficulties accounted when using the method with a large number of data will also be solved, cf. the work by Colombo (1979), Jekeli (1985). For global approximation, I would judge, that least square collocation also would work very well, but comparisons between methods are lacking.

This does not mean that all problems have been solved in global and local gravity field approximation. In Tscherning (1986) I have listed a large number of current problems: how to do we fit local solutions together, why is there a height-dependent error in geoid heights computed from spherical harmonic expansions, is it possible to give the error estimates a statistical interpretation?

There are still many problems to be solved. Their solution requires a good understanding not only of the theory of functional analysis, but also numerical experiments and a careful comparison of the results of several methods as done e.g. in Kearsly et al. (1985).

Appendix: Hints to the solution of the exercises.

Exercise	Hint
3.1	See Barthelmes (1986).
3.3	See Tscherning (1985, example 2.2)
3.4	Solution: The product sum of the gravity values produced by each block evaluated in P and Q divided by the volume of the block.
3.5	See Tscherning (1985, example 2.6)
3.6	See Tscherning (1985, example 2.7)
3.7	See Tscherning (1985, example 2.7)
4.1	Use eq. (3.24).
4.2	See Tscherning (1985, example 2.3)
4.3	See Tscherning (1985, example 2.4 and 3.1).

References:

Aubin, J.-P.: Approximation of elliptic boundary-value problems. Wiley-Interscience, 1972.

Barthelmes,F.: Untersuchungen zur Approximation des aeusseren Gravitationsfeldes der Erde durch Punktmassen mit optimierten Positionen. Akademie der Wissenschaften der DDR, ZIPE, Potsdam 1986.

Bjerhammar, A.: A Dirac Approach to Physcial Geodesy. Z. f. Vermessungswesen, 101 Jg., no. 2, pp. 41-44, 1976.

Bjerhammar,A.: A robust approach to global problems in physical geodesy. Bulletin Geodesique, Vol. 59, pp. 303-315, 1985.

Colombo, O.: Optimal estimation from data regularly sampled on a sphere with applications in geodesy. Reports of the Dep. of Geodetic Science, No. 291, The Ohio State University, Columbus, 1979.

Forsberg, R.: Local Covariance Functions and Density Distributions. Reports of the Department of Geodetic Science and Surveying No. 356, The Ohio State University, Columbus, 1984.

Freeden, W.: On the Permanence Property in Spherical Spline Interpolation. Reports of the Department of Geodetic Science and Surveying, No. 341, The Ohio State University, Columbus, Ohio, 1982.

Garabedian, P.R.: Partial Differential Equations. John Wiley & Son, New York, 1964.

Goad, C.C., C.C.Tscherning and M.M.Chin: Gravity Empirical Covariance values for the Continental United States. J.Geophys.Res., Vol. 89, No. B9, pp. 7962-7968, 1984.

Hardy,R.L. and W.M.Goepfert: Least squares prediction of gravity anomalies, geoid undulations and deflections of the vertical with multiquadratic harmonic functions. Presented XVI General Assembly IUGG, Grenoble 1975.

Heiskanen, W.A. and H. Moritz: Physical Geodesy. W.H. Freeman & Co, San Francisco, 1967.

Jekeli, C.: On Optimal Estimation of Gravity from Gravity Gradients at Aircraft Altitude. Rev. of Geophysics, Vol. 23, pp. 301-311, 1985.

Kearsley, A.H.W., M.G.Sideris, J.Krynski, R.Forsberg and K.P.Schwarz: White Sands Revisited - A Comparison of Techniques to Predict Deflections of the Vertical. Report 30007, Division of Syrveying Engineering, University of Calgary, 1985.

Krarup, T.: A Contribution to the Mathematical Foundation of Physical Geodesy. Meddelelse no. 44, Geodætisk Institut, København 1969.

Krarup, T.: Some remarks about collocation. In: H.Moritz and H.Suenkel (Ed.): Approximation methods in Geodesy, pp. 193-209, H.Wichmann Verlag, Karlsruhe, 1978.

Lachapelle, G. and K.-P.Schwarz: Empirical Determination of the gravity anomaly covariance function in mountainous areas. The Canadian Surveyor, Vol. 34, no. 3, pp. 251-264, 1980.

Lachapelle, G. and C.C. Tscherning: Use of Collocation for Predicting Geoid Undulations and related Quantities over Large Areas. Proceedings Int. Symposium on the Geoid in Europe and Mediterranean Area, Ancona, Italy, September 25-29, 1978, pp. 1-19, 1978.

Lauritzen, S.L.: The Probabilistic Background of Some Statistical Methods in Physical Geodesy. Meddelelse no. 48, Geodætisk Institut, 1973.

Lelgemann, D.: On the numerical Properties of Interpolation with Harmonic Kernel Functions. Manuscripta Geodaetica, Vol. 6, no. 2, pp. 157-191, 1981.

Lerch, F.J., S.M.Klosko, R.E.Laubscher and C.A.Wagner: Gravity Model Improvement using GEOS-3 (GEM 9 & 10), Goddard Spece Flight Center, X-921-77-246, 1977.

Meissl, P.: The use of finite elements in physical geodesy. Reports of the Dep. of Geodetic Science and Surveying, No. 313, The Ohio State University, Columbus, 1981.

Moritz,H.: Introduction to Interpolation and Approximation. In: Moritz and Suenkel (Ed.): Approximation methods in Geodesy, pp. 1-45, H.Wichmann Verlag, Karlsruhe, 1978.

Moritz, H.: Advanced Physical Geodesy. H.Wichmann Verlag, Karlsruhe, 1980.

Parzen, E.: Statistical Inference on Time Series by Hilbert Space Methods, I. 1959. (Reprinted in "Time Series Analysis Papers", Holden-Day, San Francisco, 1967, pp. 251-282).

Sanso',F., R.Barzaghi and C.C.Tscherning: Choice of Norm for the Density Distribution of the Earth. In print Geoph.Jour.Royal Astr. Soc., 1986.

Sanso, F. and C.C.Tscherning: Mixed Collocation: A proposal. Quaterniones Geodasiae, Vol. 3, no. 1, pp. 1-15, 1982.

Suenkel, H.: A Covariance Approximation Procedure. Reports of the Department of Geodetic Science, No. 286, The Ohio State University, Columbus, Ohio, 1979.

Tscherning,C.C.: Bestemmelse af lodafvigelser ved collokation. (In Danish, "Determination of Deflections of the Vertical using Collocation", Masters thesis, University of Copenhagen, 1970.

Tscherning, C.C.: Representation of Covariance Functions Related to the Anomalous Potential of the Earth using Reproducing Kernels. The Danish Geodetic Institute Internal Report No. 3, 1972.

Tscherning, C.C.: Application of Collocation: Determination of a Local Approximation to the Anomalous Potential of the Earth using "Exact" Astro-Gravimetric Collocation. In: Brosowski, B. and E. Martensen (Ed's): Methoden und Verfahren der Mathematischen Physik, Vol. 14, pp. 83-110, 1975.

Tscherning, C.C.: Models for the Auto- and Cross Covariances between Mass Density Anomalies and First and Second Order Derivatives of the Anomalous Potential of the Earth. Proceedings 3rd. Int. Symposium "Geodesy and Physics of the Earth", Weimar, October, 1976, pp. 261-268, Potsdam, 1977.

Tscherning, C.C.: On the Convergence of Least Squares Collocation. Bolletino di Geodesia e Scienze Affini, Vol. XXXIII, No. 2-3, pp. 507-516, 1978.

Tscherning, C.C.: Introduction to Functional Analysis with a View to its Application in Approximation Theory. In: Moritz, H. and H.Suenkel (Ed's): Approximation Methods in Geodesy, H.Wichmann Verlag, Karlsruhe, pp. 157-192, 1978a.

Tscherning, C.C.: Local Approximation of the Gravity Potential by Least Squares Collocation. In: K.P.Schwarz (Ed.): Proceedings of the International Summer School on Local Gravity Field Approximation, Beijing, China, Aug. 21 - Sept. 4, 1984. Publ. 60003, Univ. of Calgary, Calgary, Canada, pp. 277-362, 1985.

Tscherning, C.C.: GEOCOL - A FORTRAN-program for Gravity Field Approximation by Collocation. Technical Note, Geodætisk Institut, 3.ed., 25 MAR 1985a.

Tscherning, C.C.: Current Problems in Gravity Field Approximation. Proceedings 1. Hotime-Marussi Symposium, Rome, June 3-6, 1985, pp. 363-384. Politecnico di Milano, 1986.

Tscherning, C.C. and R.H.Rapp: Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations, and Deflections of the Vertical Implied by Anomaly Degree-Variance Models. Reports of the Department of Geodetic Science No. 208, The Ohio State University, Columbus, Ohio, 1974.

Tscherning, C.C. and H. Suenkel: A Method for the Construction of Spheroidal Mass Distributions consistent with the harmonic Part of the Earth's Gravity Potential. Manuscripta Geodaetica, Vol. 6, pp. 131-156, 1981.

Wenzel,H.-G.: Hochaufloesende Kugelfunktionsmodelle fuer das Gravitationspotential der Erde. Wiss. Arb. Fachrichtung Vermessungswesen der Universitaet Hannover, (in print), 1985.

4. Approximation in a Hilbert space

4.1. Approximation in a vector space

The most simple form for an approximation is constructed as the linear combination ($\tilde{f}$) of a given set of m functions, f_j, which gives the best least squares fit to the data, $y_i = L_i(f)$, i=1,...,n ,

$$\sum_{i=1}^{n} (L_i(\sum_{j=1}^{m} f_j a_j) - y_i)^2 \cdot w_i = \min_{\{a_j\}} \tag{4.1}$$

where w_i are (positive) weights, depending e.g. on the quality of the data. There are then three different situations, depending on whether n > m, n = m or n < m.

If we have an overdetermined problem, the solution is well known

$$\{a_j\} = (A^T W A)^{-1} A^T W y \ , \ A = \{L_i f_j\} \tag{4.2}$$

if $A^T WA$ is of full rank. (W is a diagonal matrix with w_i in the diagonal).

Note, that if the observations are values of f, $L_i(f) = f(P_i)$, regularily distributed with unit weights, then we here have a discretization of the minimum norm condition

$$\int_{\Omega} (\sum_{j=1}^{m} a_j f_j(P) - f(P))^2 d\Omega = \min_{a_j} \tag{4.3}$$

In case n=m, the minimum is simply found using

$$\{a_j\} = A^{-1} \{y_j\} . \tag{4.4}$$

These solutions are also in certain cases equivalent to mimimum norm solutions, where $\| f-\tilde{f} \|$ or $\| \tilde{f} \|$ is minimalized, as we shall see in section 4.3. In these cases A will be symmetric and positive definite.

The underdetermined problem is generally reduced to the problem n=m by finding the eigenvalues of the m x m matrix AA^T which has non-zero eigenvalues (singular value decomposition). These (n) eigenvectors are then used as new base vectors, see (Sansó et al., 1986, Appendix 1). Other alternatives are found, if we work inside a Hilbert space, see the following sections.

4.2 Best linear approximation in an inner product space, H

Given a set of linear independent elements $g_i \in H$, i=1,...,n it is possible to find a unique "best" linear approximation $\tilde{f}$ to a function $f \in H$,

$$\tilde{f} = \sum_{i=1}^{n} a_i g_i$$

in the sense that $\tilde{f}$-f has the smallest possible norm. This means that for any set b_i, i=1,...,n

$$\| f-\tilde{f} \| = \| f - \sum_{i=1}^{n} a_i g_i \| \leq \| f - \sum_{i=1}^{n} b_i g_i \| .$$

Using Gram-Schmidt orthonormalization, we can findan orthonormal set g_i^*, and it is easily seen that a best linear approximation is given by

$$\tilde{f} = \sum_{i=1}^{n} (f_i, g_i^*) g_i^*,$$

the projection of f on the subspace spanned by $\{g_i\}$. The difference between $\tilde{f}$ and f is orthogonal on all g_i^* because

$$(f- \sum_{j=1}^{n} (f,g_j^*) g_j^*, g_i^*) = (f,g_i^*) - \sum_{k=1}^{n} (f,g_k^*)(g_k^*,g_i^*) = (f,g_i^*)-(f,g_i^*) = 0$$

and must therefore always be orthogonal in all g_k, see Fig. 1.

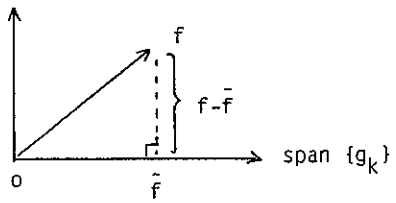

Fig. 1. Construction of $\tilde{f}$ as a projection.

We can use this to write down a system of normal equation which directly will determine the constants $\{a_i\}$. We must have

$$(f-\tilde{f}, g_i) = (f-a_1 g_1 \cdots -a_n g_n, g_i) = 0$$

or

$$a_1(g_1,g_i) + a_2(g_2,g_i) + \cdots a_n(g_n,g_i) = (f,g_i) .$$

On matrix form

$$\{ (g_i,g_j) \} \{ a_j \} = \{(f,g_i)\} \qquad (4.5)$$

Note, that in order to find $\tilde{f}$ we must require that f is an element of the same Hilbert space as the functions g_i so that (f,g_i) can be calculated. Also data must be available permitting the calculation. Or we must select the inner product, so that it uses the available data. When approximating T using gravity data, for example, an inner product

$$(f,g)_{\Delta g} = \int_{\omega} (\frac{\partial f}{\partial r}|_P + \frac{2}{r}f(P))(\frac{\partial g}{\partial r}|_P + \frac{2}{r}g(P))d\omega_P \qquad (4.6)$$

could "easily" be used.

Exercise 4.1: Calculate the reproducing kernel corresponding to this norm used on the space of example 2.1 with the subspaces of degree 0 and 1 removed. Show that the gravity anomaly functionals evaluated at the boundary will have infinite norm, and hence do not belong to the dual space.●

4.3 Approximation in a reproducing kernel Hilbert space (RKHS)

A method to circumvent the actual evaluation of the inner product in eq. (4.5) is to use for g_i the representers of the functionals corresponding to the observations, $g_i = K(P,L_i)$. Eq. (4.5) becomes

$$\{(g_i,g_j)\}\ \{a_j\} = \{K(L_i,L_j)\}\ \{a_j\} = \{(f,g_i)\} = \{(f,K(P,L_i))\} = \{L_i f\}\ ,$$

where $L_i f$ are the observations. In this case we also have

$$L_i\tilde{f} = \sum_{k=1}^{n} a_k L_i(g_k) = \sum_{k=1}^{n} \{L_k f\}^T \{K(L_k,L_j)\}^{-1} \{K(L_i,L_j)\} = L_i f \qquad (4.7)$$

i.e. there is an exact agreement between observations and values computed using the best approximation $\tilde{f}$.

The fulfilment of this condition is the basis for minimum norm collocation. Here a function $\tilde{f}$ is obtained so that eq. (4.7) is fulfilled and so that $\tilde{f}$ has the minimum norm in between the elements of a RKHS which fulfil this equation. Since we do not require that $\|f-\tilde{f}\|$ is minimal (or that it can be computed) we do not even need to require that f is an element of the RKHS, H. Only the linear functionals associated with the observations must be elements of H^*. (On the other hand if $f\in H$, then $\|f-\tilde{f}\|$ will be minimal).

The situation is shown in Fig. 2, where $\tilde{f}$ must be an element of an affine subspace, $\tilde{f}\in A = \{g \mid L_i g = L_i f,\ i=1,\ldots,n\}$, where again $L_i f$ are the observed values.

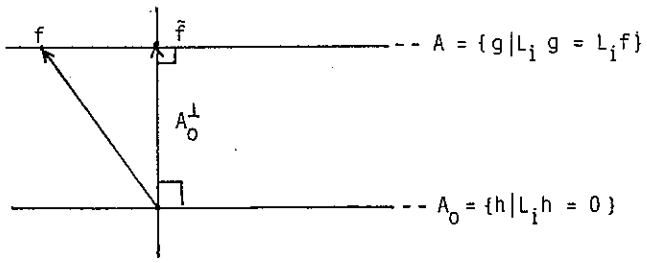

Fig. 2. The construction of $\tilde{f}$ as the intersection of two subspaces. Note that f is supposed to be in H in the figure.

$\tilde{f}$ must be the element in the affine subspace, which has the shortest distance from zero, i.e. it must be located on the n-dimensional subspace orthogonal to the affine subspace. This subspace must also be orthogonal on the subspace $A_0 = \{h \,|\, L_i h = 0, i=1,\ldots, n\}$, parallel to the affine subspace A.

If we regard the representers $K(L_i,P)$, then if $h \in A_0$

$$(h(P),K(L_i,P)) = L_i(h(P),K(P,Q)) = L_i h = 0,$$

i.e. the functions $K(L_i,P)$ span the subspace orthogonal to A_0,

$$A_0^{\perp} = \{g = \sum_{i=1}^{n} a_i K(L_i P),\ \{a_i\} \in R^n\} .$$

The function $\tilde{f}$ must therefore be equal to the intersection between $A_0^{\perp}$ and A, so

$$\tilde{f}(P) = \sum_{i=1}^{n} a_i\ K(L_i,P) \tag{4.8}$$

and

$$\{L_i\tilde{f}\} = \{K(L_i,L_j)\}\ \{a_j\} = \{L_i f\},$$

$$\{a_j\} = \{K(L_i,L_j)\}^{-1}\ \{L_i f\} \tag{4.9}$$

A further condition for finding $\tilde{f}$ is then naturally that $\{K(L_i,L_j)\}$ is positive definite, or that the functionals as elements of H^* are linearily independent.

It can easily be proved rigorously (cf. Tscherning, 1975, p. 89) that $\tilde{f}$ has the minimum norm. In fact the norm is

$$\| \tilde{f} \|^2 = \left(\sum_{i=1}^{n} a_i \, K(L_i,P), \; \sum_{j=1}^{n} a_j K(L_j,P) \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j (K(L_i,P), \, K(L_j,P))$$

$$= \left(\{L_i f\}^T \{K(L_i,L_k)\}^{-1} \right)^T \{K(L_k,L_p)\} \{K(L_p,L_j)\}^{-1} \{L_j f\}$$

$$= \{L_i f\}^T \{K(L_i,L_j)\}^{-1} \{L_j f\} \tag{4.10}$$

Using $\tilde{f}$, the approximate value corresponding to any functional $L \in H^*$ can be calculated,

$$L\tilde{f} = \sum_{i=1}^{n} a_i \, K(L,L_i) \; . \tag{4.11}$$

If $f \in H$ we may also calculate an upper limit for the error of prediction using eq. (3.2)

$$|Lf - L\tilde{f}| = |L(f-\tilde{f})| = |L(f) - \{K(L,L_i)\}^T \{K(L_i,L_j)\}^{-1} \{L_j f\}|$$

$$= |(L - \{K(L,L_i)\}^T \{K(L_i,L_j)\}^{-1} \{L_j\})(f)|$$

$$\leq \|f\| \cdot \|L - \{K(L,L_i)\}^T \{K(L_i,L_j)\}^{-1} \{L_j\}\|_*$$

or

$$|L(f) - L(\tilde{f})| \leq \|f\| \left(\|L\|_*^2 - \{K(L,L_i)\}^T \{K(L_i,L_j)\}^{-1} \{K(L,L_j)\} \right)^{\frac{1}{2}} \; . \tag{4.12}$$

An alternative expression for the upper limit of the error is found in (Krarup, 1978, eq. (12)). However, such equations have a limited use since in all cases $\|T\|$ must be known.

Exercise 4.2. Consider the case where we have only one observation $L_p f = f(P)$. Write down the minimum norm collocation solution and the expression for the error bound. ●

Exercise 4.3. Consider the case where the n observations are the n first coefficients $\{b_i\}$ with respect to an orthogonal base f_i. Show that the minimum norm collocation solution is the Fourier expansion

$$\tilde{f}_n(P) = \sum_{i=1}^{n} b_i f_i(P)$$

as we should expect. Then suppose we have a second set of observations as well, $y_i = L_i(f)$, $i=n+1,\ldots,N$. Show that the collocation solution in this case is identical to the determination of an approximation $\tilde{\tilde{f}}_N$, using a reproducing kernel with the first n terms removed, and new data values $x_i=y_i-L_i(\tilde{f}_n)$, $\tilde{\tilde{f}}=\tilde{f}_n+\tilde{\tilde{f}}_N$. ●

Example 4.1. Cholesky decomposition of the normal equations. Let $C = \{K(L_i,L_j)\}$, $C_o = K(L,L) = \|L\|_*^2$ and $C_p = \{K(L,L_i)\}$. Then the solution to the normal equations $\{a_i\} = C^{-1}\{y_i\}$ may be calculated using Cholesky decomposition of C,

$$C = U^T U$$

where U is an upper triangular matrix, called the Cholesky factor. {a} is found by solving the simple system of equations

$$U\,a = (U^T)^{-1}y \ .$$

The Cholesky factor of the matrix

$$\begin{Bmatrix} C & C_P \\ C_P^T & C_o \end{Bmatrix}$$

is

$$\begin{Bmatrix} U(U^T)^{-1}C_P \\ 0 \quad (C_o-C_P^T C^{-1} C_P)^{\frac{1}{2}} \end{Bmatrix} \ ,$$

which shows that Cholesky reduction produces the quantity needed in the estimate of the error bound eq. (4.12) as its diagonal element no. n+1.

As a part of the solution a new vector of observations $(U^T)^{-1}y$ is computed. It is easily seen that the representers of the linear functionals corresponding to these new observations are orthonormal. Hence, the addition of new observations does not require a new calculation of the Cholesky factor of the first n columns of the new normal equation matrix. (This is the basis for the re-start possibility implemented in the program GEOCOL, (Tscherning, 1985a), and corresponds to the so-called permanency property, cf. (Freeden, 1982).) ●

Example 4.2. Approximation of the function $g(x) = -2x+1$, using minimum norm collocation, and computation of error bound. We use the Hilbert space of example 3.4, and suppose we have two observations $g(0) = 1$ and $g(1) = -1$. We use the results of the examples 2.9, 3.9 and 4.1.

The representer of L_o and L_1 are f_o and f_1, where

$$f_o = -\frac{15}{4}x^2 + \frac{9}{4} \text{ and } f_1 = -\frac{15}{2}x^2 + 3x - \frac{3}{2}.$$

The approximation is then

$$\tilde{g}(x) = b_o f_o + b_1 f_1 \quad \text{with} \quad \{b\} = \{K(L_i,L_i)\}^{-1} \{ {}_{-1}^{1} \}.$$

Then

$$\begin{Bmatrix} \frac{9}{4} & -\frac{3}{2} \\ -\frac{3}{2} & 9 \end{Bmatrix} \begin{Bmatrix} b_o \\ b_1 \end{Bmatrix} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}$$

and from this $b_o = \frac{5}{12}$ and $b_1 = -\frac{1}{24}$, hence

$$\tilde{g}(x) = -\frac{15}{8}x^2 - \frac{1}{8}x + 1.$$

The difference $|\tilde{g}(x)-g(x)| = \frac{15}{8}|(x-1)x|$, is equal to zero for $x = 0,1$ as expected. The error expression (4.12) gives for the evaluation functional $L_x(f)=f(x)$

$$|g(x) - \tilde{g}(x)| \leq \|g\| \left(K(x,x) - \begin{Bmatrix} K(0,x) \\ K(1,x) \end{Bmatrix}^T \begin{Bmatrix} \frac{9}{4} & -\frac{3}{2} \\ -\frac{3}{2} & 9 \end{Bmatrix}^{-1} \begin{Bmatrix} K(0,x) \\ K(1,x) \end{Bmatrix} \right)^{\frac{1}{2}}$$

The error expression is according to example 4.1 the Cholesky factor of the 3x3 matrix

$$\begin{Bmatrix} \frac{9}{4} & -\frac{3}{2} & K(0,x) \\ -\frac{3}{2} & 9 & K(1,x) \\ K(0,x) & K(1,x) & K(x,x) \end{Bmatrix}$$

which is easily seen to be

$$\begin{Bmatrix} \frac{3}{2} & -1 & (-\frac{5}{2}x^2 + \frac{3}{2}) \\ 0 & 2\sqrt{2} & \frac{\sqrt{2}}{4}(5x^2+3x) \\ 0 & 0 & (\frac{15}{8}x^4 - \frac{30}{8}x^3 + \frac{15}{8}x^2)^{\frac{1}{2}} \end{Bmatrix}$$

Hence, since $\|g\| = (7/3)^{\frac{1}{2}}$,

$$|g(x) - \tilde{g}(x)| \le (\tfrac{7}{3})^{\frac{1}{2}} (\tfrac{15}{8})^{\frac{1}{2}} |x(x-1)| = (\tfrac{35}{8})^{\frac{1}{2}} |x(x-1)|$$

which is slightly more pessimistic than the true error. ●

Exercise 4.4. Repeat the calculations in example 4.2 with the empirical covariance function of g(x) as reproducing kernel,

$$K(x,y) = 1^2 \cdot 1 \cdot 1 + (\frac{2}{\sqrt{3}})^2 3xy = 1+4xy . \quad \bullet$$

4.4. Approximation using the empirical covariance function

Besides the reqirement of minimum norm of $\tilde{f}$, another (similar) requirement may lead to a collocation solution. The requirement is, that we want to find a function, so that the overall mean square prediction error becomes minimal. Let the error of prediction be e_P , and suppose our observations are $T(Q_i)$ $i=1, \ldots,n$, then

$$e_P = \tilde{T}(P) - T(P) = \sum_{i=1}^{n} b_{Pi} T(Q_i) - T(P)$$

then

$$e_P^2 = T(P)^2 - 2 \sum_i b_{Pi} T(P) T(Q_i) + \sum_i \sum_k T(Q_i) T(Q_k) b_{Pi} b_{Pk} .$$

We want $M(e_P^2)$ to be minimal. But with

$$M(T(Q_i)T(Q_k)) = C_{ik}$$
$$M(T(P)\ T(Q_i)) = C_{Pi} , \qquad M(T(P)^2) = C_o$$

we see

$$m_P^2 = M(e_P^2) = C_o - 2 \sum b_{Pi} C_{pi} + \sum \sum b_{Pi} b_{Pk} C_{ik} .$$

If we want to minimize this expression, then

$$\frac{\partial m_P^2}{\partial b_{Pi}} = -2 C_{Pi} + 2 \sum_{k=1}^{n} b_{Pk} C_{ik} = 0, \quad i=1,\ldots,n ,$$

or

$$\sum_{k=1}^{n} C_{ik} b_{Pk} = C_{Pi} , \qquad (4.13)$$

so that

$$\tilde{T}(P) = \sum_{i=1}^{n} a_i \, C_{Pi}$$

with

$$\{a_i\} = \{C_{ik}\}^{-1} \{T(Q_k)\} \tag{4.14}$$

We see this is the collocation solution with $K(P,Q) = C(P,Q)$. Unfortunately $\|T\|$ is infinite in this norm, but we may compute a mean square error

$$m_L^2 = C(L,L) - \{C(L,L_i)\}^T \{C(L_i,L_j)\}^{-1} \{C(L,L_j)\}. \tag{4.15}$$

4.5 Treatment of noise and parameters in minimum norm collocation

The data we have will generally not be directly related to T through a linear functional. It contain measurement errors (e_i) and may be affected by parameters such as an incompletely known relationship between a local geodetic datum and a geocentric, correctly oriented datum. Let us denote the observations y_i, the m parameters $\{X_j\} = X$ and suppose the observations are related to the parameters through a vector A_L. Then

$$y_i = L_i(T) + e_i + A_L^T X \, . \tag{4.16}$$

Let us suppose, that the noise vector e has the variance-covariance matrix D. A minimum norm collocation solution may then be obtained so that the sum of three quantities are minimized, namely $\|\tilde{T}\|^2$, $e^T D^{-1} e$ and $X^T P X$, where P is a positive definite matrix expressing some a-priori weights of the parameters,

$$\|\tilde{T}\|^2 + e^T D^{-1} e + X^T P X = \min \, . \tag{4.17}$$

The solution may be found using Lagrange multipliers as described in (Moritz, 1980, Chapter 29 and 30).

$$\begin{aligned} \tilde{X} &= P^{-1} A^T \tilde{C}^{-1} (y + A\tilde{X}) \\ &= (A^T \tilde{C}^{-1} A + P)^{-1} A^T \tilde{C}^{-1} y \end{aligned} \tag{4.18}$$

$$\tilde{T} = \{K(L_i,P)\}^T \tilde{C}^{-1} (y - A\tilde{X}) \, , \tag{4.19}$$

with $\tilde{C} = \{K(L_i,L_j) + D_{ij}\}$.

The fulfilment of the last two equations are only a necessary condition for a minimum. However, it is possible to show that we have obtained a minimum.

It is still possibly to write down expressions for maximal error bounds similar to eq. (2.34), now using both the norm in the reproducing kernel Hilbert space, and the norms implicitly introduced in the n-dimensional space of measurements by D and the m-dimensional space of parameters by P, see Moritz (1980, p. 128). For the mean square error of prediction we find with

$$H = \{ K(L,L_i) \}^T \bar{C}^{-1}$$

that

$$m_x^2 = (A^T\bar{C}^{-1} A + P)^{-1} \tag{4.20}$$

and

$$m_L^2 = K(L,L) - H\{K(L,L_i)\} + HAm_x^2 (HA)^T \tag{4.21}$$

We will end this section with two examples showing that the addition of parameters makes no problem from the computational standpoint, but that the estimated results maybe should be evaluated very carefully using the error estimates.

Example 4.3. Modified Cholesky factorization.

Like in example 4.1 we will now modify the Cholesky decomposition of C, so that we directly may obtain the estimates of the parameters and the error estimates. The solution is obtained from the extended matrix

$$\begin{Bmatrix} \bar{C} & A & y \\ A^T & P & 0 \\ y & 0 & 0 \end{Bmatrix}$$

by defining the Cholesky factor as

$$\begin{Bmatrix} U & (U^T)^{-1}A & (U^T)^{-1}y \\ 0 & V & (V^T)^{+1}A^T\bar{C}^{-1}y \\ 0 & 0 & (y^TC^{-1}y - y^T\bar{C}^{-1}A(VV^T)^{-1}A^T\bar{C}^{-1}y)^{\frac{1}{2}} \end{Bmatrix}$$

where V is the Cholesky reduced of $(P+A^T\bar{C}^{-1}A)$. The modification with respect to an ordinary Cholesky reduction is that this would produce $P-A^T\bar{C}^{-1}A$, which might be negative definite, since P might be the 0-matrix.

It is then easily seen, that the X vector is obtained by solving

$$VX = (V^T)^{-1} A^T C^{-1} y$$

and that the error estimate is obtained by computing the Cholesky factor of

$$\begin{Bmatrix} \bar{C} & A & C_P \\ A^T & P & 0 \\ C_P^T & 0 & C_0 \end{Bmatrix} \bullet$$

Example 4.4. Suppose we have observed two sea-surface heights (ζ_1 and ζ_2) by satellite radar altimeter measurements. However we will treat the measurements as if they were geoid heights biased by a constant X due to orbit error and sea surface topography,

$$L_i(T) + X + e_i = \zeta_i \quad , \qquad i=1,2,$$

where e_i is the error and L_i is given by eq. (2.10). Then suppose $K(L_1,L_2) = 0.60\ m^2$, $K(L_i,L_i) = 0.99\ m^2$, and the variance of the (supposedly uncorrelated) noice is $0.01\ m^2$. Suppose $P = 0$ in eq. (4.17). Then we will show how an estimate of T and of the bias may be obtained, and we will predict the geoid height in a point Q where $K(L_1, \zeta_Q) = K(L_2, \zeta_Q) = 0.8\ m^2$, and compute the error estimate.

The estimate of T is

$$\tilde{T}(P) = b_1 K(P,L_1) + b_2 K(P,L_2).$$

As shown in example 4.3 we will get the solution (b_1,b_2,X) by modified Cholesky reduction of the matrix

$$\begin{Bmatrix} \bar{C} & A^T & y \\ A & 0 & 0 \\ y & 0 & 0 \end{Bmatrix} = \begin{Bmatrix} 1 & 0.6 & 1 & \zeta_1 \\ 0.6 & 1 & 1 & \zeta_2 \\ 1 & 1 & 0 & 0 \\ \zeta_1 & \zeta_2 & 0 & 0 \end{Bmatrix}$$

and then execute the back-substitution using the last column as the reduced right-hand side. The Cholesky factor U is

$$U = \begin{Bmatrix} 1 & 0.6 & 1 & \zeta_1 \\ 0 & 0.8 & 0.5 & (\zeta_2 - 0.6\zeta_1)\frac{5}{4} \\ 0 & 0 & \sqrt{5}/2 & (\zeta_1 + \zeta_2)\sqrt{5}/4 \\ 0 & 0 & 0 & ((\zeta_1 - \zeta_2)\frac{5}{4})^{\frac{1}{2}} \end{Bmatrix}$$

The back-substitution is then a solution of the equations

$$\begin{Bmatrix} 1 & 0.6 & 1 \\ 0 & 0.8 & 0.5 \\ 0 & 0 & \sqrt{5}/2 \end{Bmatrix} \begin{Bmatrix} b_1 \\ b_2 \\ X \end{Bmatrix} = \begin{Bmatrix} \zeta_1 \\ (\zeta_2 - 0.6\,\zeta_1)\frac{5}{4} \\ (\zeta_1 + \zeta_2)\sqrt{5}/4 \end{Bmatrix}$$

or

$$X = (\zeta_1 + \zeta_2)/2 \quad \text{with} \quad (A^T C^{-1} A)^{\frac{1}{2}} = \frac{\sqrt{5}}{2} \simeq 1.12 \text{ m}$$

$$b_2 = \frac{5}{4}(\zeta_2 - \zeta_1), \quad b_1 = \frac{5}{4}(\zeta_1 - \zeta_2) .$$

Note that the prediction back in P_1 gives

$$\tilde{\zeta}_1 = K(L_1, L_1)b_1 + K(L_1, L_2)\,b_2 = 0.99\frac{5}{4}(\zeta_1 - \zeta_2) + \frac{3}{5}\frac{5}{4}(\zeta_2 - \zeta_1) \simeq \frac{1}{2}(\zeta_1 - \zeta_2)$$

i.e. the bias has been removed. The prediction in Q becomes

$$\tilde{\zeta}_Q = \frac{4}{5}\frac{5}{4}(\zeta_2 - \zeta_1 + \zeta_1 - \zeta_2) = 0 .$$

The error of prediction may be computed from the Cholesky reduced of the right-hand side with elements $K(L_i, \zeta_Q)$, 0 and $K(\zeta_o, \zeta_o)$:

$$\left\{\begin{array}{ccc:c} \overbrace{1 & 0.6 & 1}^{\text{reduced part}} & & 0.8 \\ 0 & 0.8 & 0.5 & & 0.8 \\ 0 & 0 & \sqrt{5}/2 & & 0 \\ \hdashline 0.8 & 0.8 & 0 & & -0.99 \end{array}\right\} \begin{array}{c} \text{reduction of} \\ \text{last column} \\ \longrightarrow \end{array} \left\{\begin{array}{c:c} & 0.8 \\ & 0.4 \\ & 2.4/\sqrt{5} \\ \hdashline & \sqrt{0.99} \end{array}\right\}$$

i.e. a standard deviation of nearly 1 m. Note that the bias is estimated to be the mean value of the observations as could be expected. ●

Exercise 4.5. Repeat the example (1) without assuming a bias and (2) with only one observation, but with a bias. ●

We have here seen that noise and parameters may be easily treated in the minimum norm collocation set-up. The same is true if $\tilde{T}$ is estimated using least squares with fewer base functions than observations. For the best approximation methods (section 4.2) it seems more difficult.

4.6 Density estimation

We know (cf. example 2.10) that it in principle is impossible to determine the Earth's density distribution from information on the outer potential only. But we will here show that an estimate may be made anyway using principles which might be given a physical interpretation in the form of implied elastic properties.

If potentials of point masses or rectangular boxes (examples 3.5,3.6) are used, then the interpretation is immediate using the coefficients of the linear combination obtained. Volume distributions filling the whole space inside the Earth may be estimated using the harmonic density functions discussed in section 3.3.

The estimation of the harmonic density may be made if the inverse of the Newton-functional eq. (3.17) can be constructed. This is easily done if the boundary surface is a sphere, because the exterior spherical harmonic functions are mapped into the interior spherical harmonic functions, multiplied by a constant:

$$N_P(V_{ij}^{\,o}) = \frac{4\pi G R^2}{(2i+3)(2i+1)} V_{ij}(P) ,$$

cf. (Tscherning, 1977). Using the inverse of this equation, we can find the representer of the functional N^{-1} for a reproducing kernel belonging to a space of functions harmonic outside a sphere with radius R by

$$K(N_Q^{-1},P) = \sum_{i=2}^{\infty} \sigma_i \frac{(2i+3)(2i+1)}{4\pi G R} \frac{(r')^i}{r^{i+1}} P_i(\cos\phi), \quad r'<R .$$

For a non-spherical boundary a Hilbert space may be constructed using the product of two Hilbert spaces. One must consist of functions harmonic outside a sphere enclosed in the Earth and the other must be the finite dimensional space of potentials of indicator functions filling the space between the sphere and the actual boundary used, for details see Sansó and Tscherning (1980).

Hence, not only point mass models, but (nearly) all methods can be given a geophysical interpretation in the form of an implied volume density distribution for the Earth. Further possibilities for determining a unique density distribution (besides the use of harmonic densities) are discussed in Tscherning (1977) and Tscherning and Sünkel (1981).

3. Hilbert spaces and reproducing kernels

3.1 Banach and Hilbert spaces

A sequence $\{f_i,\ i=1,\ldots,\infty\}$ in a normed (linear) space is called a Cauchy sequence if it has the property that for all $\varepsilon > 0$ there is an integer n so that $d(f_i,f_j) \leq \varepsilon$ for $i,j \geq n$. The space is called complete if every Cauchy sequence has a limit which belongs to the space. A complete, normed linear space is called a Banach space, and if the norm is derived from an inner product, then it is a Hilbert space.

Example 3.1. $C^0(\Omega)$, with Ω being the closed interval from a to b, may be equipped with the norm $\| f \| = \max_{a\leq x\leq b} |f(x)|$, and it is complete in this norm. But using $\| f \|^2 = \int_a^b f(x)^2 dx$ makes the space incomplete because the limits of continuous functions need not to be continuous using this norm. ●

Since the anomalous gravity potential, T, always may be considered an element of a vector space with a countable basis, we need only consider so-called separable Hilbert spaces. In such a space the expansion of an element f with respect to an orthonormal basis $\{f_1, f_2, \ldots\}$ converge in the norm towards f, $\lim_{n\to\infty} \| f - \Sigma_{i=1}^{n} a_i f_i \| = 0$, (but we still do not necessarily have $f(P) = \sum_{i=1}^{\infty} a_i f(P)$.) The spaces have a lot of nice properties, which makes them nearly look like a finite dimensional Euclidean space:

The inner product of two elements may be calculated as scalar products of their coordinates,

$$(f,g) = \sum_{i=1}^{\infty} (f,f_i)\cdot(g,f_i).$$

Also an element is uniquely determined by its coordinates

$$(g,f_i) = (f,f_i),\ i=1,\ldots,\infty \Rightarrow g = f$$
$$(f_i,f) = 0,\ i=1,\ldots,\infty \Rightarrow f = 0\ .$$

Example 3.2. The space of harmonic functions introduced in example 2.1 may as mentioned in Chapter 2 be equipped with the inner product

$$(f,g)_0 = \frac{1}{4\pi R} \int_\sigma f\cdot g\ d\sigma \qquad (3.1)$$

The subset for which $\| f \| < \infty$ will form a separable Hilbert space which we will denote $H^0(\sigma)$. (The integral in eq. (3.1) is to be understood as the limit of the integrals over concentric spheres with radii $R + \varepsilon$, $\varepsilon > 0$ for $\varepsilon \to 0$. This is because we only require the boundary values to be square integrable). ●

Exercise 3.1. The potential of a point mass located at $P_1 = (\varphi_1, \lambda_1, R_1)$ with mass M_1 is at $P = (\varphi, \lambda, r)$ equal to $G \cdot M_1 / L_1$, $L_1 = \sqrt{r^2 - 2rR_1 \cos\psi_1 + R_1^2}$, with ψ_1 being the spherical distance between P and P_1. Use the inner product (3.1) to show that the inner product of two potentials in P_1 and $P_2 = (\varphi_2, \lambda_2, R_2)$, mass M_2, is equal to

$$\left(\frac{GM_1}{L_1}, \frac{GM_2}{L_2}\right)_0 = \sum_{n=0}^{\infty} \frac{GM_1 \cdot GM_2}{(2n+1)R^2} \left(\frac{R_1 R_2}{R^2}\right)^n P_n(\cos\psi_{12}). \quad ●$$

3.2 Representers of functionals

One of the most useful properties of Hilbert spaces (and certain Banach-spaces) is that the continuous linear functionals may be represented "analytically" as the inner product of the function on which the functional is to be applied with a fixed function. This is as pointed out in example 2.2 a well known property in $\mathbf{R}^n$.

A functional, L, is bounded if there exist a constant, M, so that $|L(f)| \leq M \cdot \|f\|$ for all $f \in H$. If $\|f_i - f\| \to 0$ then

$$|L(f_i) - L(f)| = |L(f_i - f)| \leq M \cdot \|f_i - f\|, \tag{3.2}$$

so $|L(f_i) - L(f)| \to 0$, i.e. a bounded functional is continuous. Also a continuous functional is bounded. For these functionals a norm may be introduced being equal to the maximal value M for which eq. (3.2) holds for all $f \in H$, i.e.

$$\|L\|_* = \sup_{f \in H} \frac{|L(f)|}{\|f\|} \qquad (f \neq 0)$$

and then consequently

$$|L(f)| \leq \|f\| \cdot \|L\|_* \tag{3.2a}$$

Example 3.3. The norm of the coordinate functionals. Regard L_i, $L_i(f) = (f, f_i) = a_i$. Then it is obvious that

$$\|L_i\|_* = \sup \frac{|a_i|}{(\Sigma a_i^2)^{\frac{1}{2}}} = 1. \quad ●$$

This introduction of a norm of the bounded linear functionals makes this linear subspace of the space of all linear functionals, a Hilbert space, H^*.

To each $L \in H^*$ there exist a unique element $\ell \in H$ so that

$$L(f) = (\ell, f) \ .$$

The element ℓ is denoted the (Riesz) representer of L. Using eq. (3.2)

$$|L(f)| \leq |(f,\ell)| \leq \|f\| \cdot \|\ell\|$$

i.e. $\|L\|_* \leq \|\ell\|$. Also

$$\|L\|_* = \sup_{f \in H} \frac{|L(f)|}{\|f\|} \geq \frac{|L(\ell)|}{\|\ell\|} = \frac{|(\ell,\ell|}{\|\ell\|} = \|\ell\|,$$

so $\|L\|_* \geq \|\ell\|$ and consequently $\|L\|_* = \|\ell\|$.

The inner product in H^* of two functionals L_1 and L_2 may be defined directly as the inner product of the representers, ℓ_1 and ℓ_2,

$$(L_1, L_2)_* = (\ell_1, \ell_2).$$

3.3 Reproducing kernels

Let us now regard a separable Hilbert space of functions, $f: \Omega \to \mathbb{R}$. If the evaluation functionals $L_p(f) = f(P)$ are bounded, then there exist a reproducing kernel. Let us denote the representer of L_p by $K_p(Q)$. Then

$$(K_p(Q), f(Q)) = f(P) \qquad P,Q \in \Omega \ .$$

If P varies, we have a mapping

$$K(P,Q): \Omega \times \Omega \to \mathbb{R},$$

with the reproducing property

$$(K(P,Q), f(Q)) = f(P). \tag{3.3}$$

It is called a kernel, since it, when used in spaces with inner products like in example 3.2, looks like the kernel of an integral equation.

The reproducing kernel may be expressed explicitly using a known orthonormal base, f_i, $i=1,\ldots,\infty$,

$$K(P,Q) = \sum_{i=1}^{\infty} f_i(P) f_i(Q) \ . \tag{3.4}$$

This is easily seen by inserting this expression in eq. (3.3) and using the expansion of $f = \sum_{i=1}^{\infty} a_i f_i$,

$$(f(P),\ K(P,Q)) = \left(\sum_{i=1}^{\infty} a_i f_i(P),\ \sum_{j=1}^{\infty} f_j(P) f_j(Q) \right) = \sum_{i=1}^{\infty} a_i f_i(Q) = f(Q),$$

where we have used the linearity of the inner product and the orthonormality of the base functions.

Example 3.4. Regard the 3-dimensional space of example 2.8. The base functions are 1, $\sqrt{3}x$ and $\sqrt{5}(\frac{3}{2}x^2 - \frac{1}{2})$. Hence,

$$K(x,y) = 1 \cdot 1 + 3xy + 5(\tfrac{3}{2}x^2 - \tfrac{1}{2})(\tfrac{3}{2}y^2 - \tfrac{1}{2}) = \tfrac{9}{4} + 3xy - \tfrac{15}{4}(x^2 + y^2) + \tfrac{45}{4} x^2 y^2 . \bullet \tag{3.5}$$

Exercise 3.2. Check that the function $g(x) = -2x+1$ used in example 2.9 is reproduced by the kernel (3.5). ●

Exercise 3.3. Show using eq. (2.20), (3.1) and (3.4) that the Hilbert space of harmonic functions $H_0(\sigma)$ has the reproducing kernel

$$K_0(P,Q) = \sum_{i=0}^{\infty} (2i+1) \frac{R^{2i+2}}{(r r')^{i+1}} \ P_i(\cos\psi), \tag{3.6}$$

and that the reproducing property eq. (3.3) then becomes identical to Poissons integral

$$f(\varphi,\lambda,r) = \frac{1}{4\pi} \int_{-\pi/2}^{\pi/2} \int_{-\pi}^{\pi} \frac{(r^2 - R^2)}{(r^2 + R^2 - 2Rr\cos\phi)^{3/2}} \ f(\varphi',\lambda',R)\cos\varphi' d\lambda' d\varphi' \tag{3.7}$$

with $Q = (\varphi', \lambda', r')$. Show also, that the evaluation functionals at the boundary does not belong to $H_0(\sigma)^*$. ●

Example 3.5. The Krarup kernel. Regard the space of harmonic functions in example 2.1. A Hilbert space is formed by the functions for which the norm

$$\| f \|^2 = \frac{1}{4\pi} \int_{\Omega} \frac{1}{r} (\nabla f)^2 d\Omega \tag{3.8}$$

is finite .

The reproducing kernel of this space is of special interest because it is very simple. In the space the solid spherical harmonics will still be orthogonal, but they must be normalized with respect to the new norm. Since

$$\frac{1}{r}\nabla\bar{V}_{ij} = \begin{Bmatrix} -\frac{(i+1)}{r^2}\bar{V}_{ij} \\ \frac{1}{r^2}\frac{\partial}{\partial\varphi}\bar{V}_{ij} \\ \frac{1}{r^2}\frac{1}{\cos\varphi}\frac{\partial}{\partial\lambda}\bar{V}_{ij} \end{Bmatrix} = \begin{Bmatrix} -\frac{(i+2)}{r}(\frac{1}{r}\bar{V}_{ij}) \\ \frac{1}{r}\frac{\partial}{\partial\varphi}(\frac{1}{r}\bar{V}_{ij}) \\ \frac{1}{r\cos\varphi}\frac{\partial}{\partial\lambda}(\bar{V}_{ij}) \end{Bmatrix} + \begin{Bmatrix} \frac{1}{r^2}\bar{V}_{ij} \\ 0 \\ 0 \end{Bmatrix}$$

we see that

$$\begin{aligned} \|\bar{V}_{ij}\|^2 &= \frac{1}{4\pi}\int\nabla(\tfrac{1}{r}\bar{V}_{ij})\nabla\bar{V}_{ij}d\Omega + \frac{1}{4\pi}\int\frac{1}{r^2}\bar{V}_{ij}(-i-1)\tfrac{1}{r}\bar{V}_{ij}\,d\Omega \\ &= \frac{1}{4\pi}\int_\sigma \tfrac{1}{r}\bar{V}_{ij}\frac{\partial}{\partial r}|_R V_{ij}\,R^2\cos\varphi\,d\lambda\,d\varphi + \int_\Omega (\tfrac{R}{r})^{2i+2}\frac{(-i+1)}{r^3} \\ &\quad \cdot(\bar{P}_{ij}(\sin\varphi)\begin{Bmatrix}\cos j\lambda\\ \sin j\lambda\end{Bmatrix})^2\,r^2 d\lambda\cos\varphi\,d\varphi dr \\ &= i+1+[\tfrac{-1}{2i+2}(\tfrac{R}{r})^{2i+2}(-i-1)]_R^\infty = \frac{2i+1}{2R}\ . \end{aligned}$$

Hence

$$K(P,Q) = \sum_{i=0}^{\infty}\sum_{j=-i}^{i}\frac{2R}{2i+1}\bar{V}_{ij}(P)\bar{V}_{ij}(Q) = 2R\sum_{i=0}^{\infty}(\frac{R^2}{rr'})^{i+1}P_i(\cos\psi)\ .$$

Using $L = (R^2-2rr'\cos\psi + (rr'/R)^2)^{\frac{1}{2}}$

$$K(P,Q) = \frac{2R}{L}\ . \tag{3.9}$$

Note that this kernel may be interpreted as the potential of a point mass located at $Q' = (\varphi',\lambda',\frac{R^2}{r'})$ with mass $2R^2/(r'\cdot G)$,

$$K(P,Q) = (2R^2/r')/|P-Q'|\ . \quad \bullet \tag{3.10}$$

Example 3.6. The Hilbert space of exploration geophysics.

Suppose a geological structure has been divided in n non-overlapping blocks with volume v_k, k=1,...,n. Then regard the linear vector space spanned by the indicator functions I_k, equal to 1 inside the k'th block and zero outside. The functions

span an n-dimensional space. Using the inner product equal to the integral of the product of two functions, we see that the functions are orthogonal, with norm equal to the square-root of the volume. The reproducing kernel is then

$$K(P,Q) = \sum_{k=1}^{n} I_k(P)\, I_k(Q)/v_k \quad \bullet \tag{3.11}$$

Example 3.7. The covariance function of a stohastic process as a reproducing kernel.

A family of vectors $\{L_t,\ t \in T\}$ in a Hilbert space H^* is said to be a representation of the stochastic process $\{X(t),\ t \in T\}$ with covariance kernel cov(s,t) if

$$(L_t,\ L_s)_{H^*} = \mathrm{cov}(s,t) = E(X_s X_t).$$

Let now T be a "nice" subset of $\mathbb{R}^n$. Then by the so-called Mercer's theorem, there exist a sequence of normalized eigenfunctions g_i,

$$\int_T \mathrm{cov}(s,t) g_i(s) ds = \lambda_i g_i(t),$$

$$\int_T g_n(t) g_m(t)\, dt = \delta_{nm},$$

so that

$$\mathrm{cov}(s,t) = \sum_{n=1}^{\infty} \lambda_n g_n(s) g_n(t), \quad \lambda_n \geq 0. \tag{3.12}$$

Then let H be the space of all functions spanned by the set $g_n/\sqrt{\lambda_n}$, which we define as being orthonormal. Hereby the inner product of H is fixed, and the space consist of all the functions for which the square-sum of the coordinates (i.e. the norm in H) is finite. The dual space H^* will then be a representation of the process, and cov(s,t) the reproducing kernel. In this manner may all stochastic processes (random functions), be regarded as spanning a Hilbert space with reproducing kernel. The converse is also true . ●

Exercise 3.4. The reproducing kernel eq. (3.11) may be interpreted as a covariance function for the related density structure. What is the covariance of two gravity anomalies in two points P and Q outside the structure ? ●

Exercise 3.5. Show that the reproducing kernel for the subset of the space in example 2.1 equipped with the $H^0(\Omega)$ norm (multiplied by $\frac{1}{4\pi R^3}$), but with the subspace of degree zero excluded is

$$K_0(P,Q) = \sum_{i=1}^{\infty} (2i-1)(2i+1)\left(\frac{R^2}{rr'}\right)^{i+1} P_i(\cos\psi) \; . \bullet \tag{3.13}$$

Like in the Sobolev space $H^1(\Omega)$ and in the space of example 3.6 we may introduce a norm involving the integral of the product of the first order derivatives. This is called the Dirichlet inner product,

$$(f,g)_2 = \int \nabla f \cdot \nabla g \, d\Omega. \tag{3.14}$$

In this case the reproducing kernel is the sum of the Green's and Neumann's functions for Ω , see Garabedian (1964, section 7.3),

$$K_1(P,Q) = G(P,Q) + N(P,Q).$$

Exercise 3.6. Use Green's identity, and the fact that the normal derivative, $\frac{\partial N}{\partial n}$, and G are zero at the boundary to show that the reproducing property is equivalent to the properties that the two kernels solve the boundary value problems where either $\frac{\partial f}{\partial n}$ or f are known. ●

Exercise 3.7. Show that the reproducing kernel for the functions harmonic outside a sphere with radius R, equipped with the Dirichlet inner product (multiplied by $\frac{1}{4\pi R^2}$), is

$$K_2(P,Q) = \frac{1}{R} \sum_{i=0}^{\infty} \frac{2i+1}{i+1} \left(\frac{R^2}{rr'}\right)^{i+1} P_i(\cos\psi) \; . \tag{3.15}$$

Use the result in example 3.6 and

$$\int^{\infty} \frac{ds}{L} = -s \, \ell n\left(\frac{1-\cos\psi}{L+s+\cos\psi}\right), \qquad s = \frac{R^2}{rr'}$$

to construct a closed expression for the kernel . ●

A Hilbert space may also be constructed based on a norm on the space of bounded (density) functions, with the same support Ω_0 as the Earth's density distribution. Since the purpose of using such a space must be that we subsequently should be able to evaluate (compute) density estimates, then the evaluation functionals should be bounded, i.e. the space should have a reproducing kernel. This may be achieved like in example 3.6 by working in a finite dimensional space, but also spaces of functions which fulfil an elliptic (and certain parabolic) partial differential equations can be used. Here the most simple example is the space of harmonic density functions, cf. example 2.10. Let us use the inner product $H^0(\Omega_0)$,

$$(\rho_1,\rho_2)_D = \int_{\Omega_o} \rho_1(P)\rho_2(P)\, d\Omega_o . \tag{3.16}$$

The value of the potential T_1 generated by ρ_1 in a point Q is the value of the (Newton) functional, N_Q, applied on ρ_1, which may be expressed very simply by the inner product

$$N_Q(\rho_1) = T_1(Q) = G \int \frac{\rho_1(P)}{|P-Q|}\, d\Omega_o . \tag{3.17}$$

Note, that the harmonic function $k(P,Q) = G/|P-Q|$, (with singularity in Q outside Ω_o) is the Riesz representer of N_Q in $H^o(\Omega_o)$.

If Q varies, we have an operator N from $H^o(\Omega_o)$ to the set of functions harmonic in Ω . This will be a linear vector space, and we may introduce an inner product by $(T_1,T_2)_H = (\rho_1,\rho_2)_D$. Since

$$T(P) = (\rho(Q),\, k(P,Q))_D = (N(\rho),\, N(k(P,Q)))_H = (T(Q),\, N_P(k(S,Q)))_H ,$$

we see that (cf. Krarup (1978))

$$N_P(k(S,Q)) = G^2 \int_{\Omega_o} \frac{1}{|P-S|\,|Q-S|}\, d\Omega_o = K_D(P,Q) \tag{3.18}$$

is the reproducing kernel in H.

Example 3.8. If Ω_o is a sphere with radius R, K_D may be computed explicitly.

$$\int \frac{1}{|P-S|}\,\frac{1}{|S-Q|}\, d\Omega_o =$$

$$\int_\varphi \int_\lambda \int_o^R \sum_{n=0}^{\infty} \frac{r_s^{\,n}}{r_P^{\,n+1}}\, P_n(\cos\psi_{PS}) \sum_{i=0}^{\infty} \frac{r_s^{\,i}}{r_Q^{\,i+1}}\, P_i(\cos\psi_{QS})\, r_s^{\,2} \cos\varphi_s\, d\varphi_s\, d\lambda_s\, dr_s$$

$$= \sum_{j=0}^{\infty} \frac{R^{2j+3}}{2j+3}\, \frac{4\pi}{(r_P r_Q)^{j+1}}\, \frac{1}{2j+1}\, P_j(\cos\psi_{PQ})$$

$$= \sum_{j=0}^{\infty} \frac{4\pi R}{(2j+3)(2j+1)} \left(\frac{R^2}{r_P r_Q}\right)^{j+1} P_j(\cos\psi_{PQ}) = K_{\Omega_o}(P,Q) \tag{3.19}$$

Note that the evaluation functionals at the boundary have finite norm . ●

We have now seen two examples of Hilbert spaces, where the set of harmonicity is the actual set of harmonicity for T, and where the inner product is computed by integration over the actual set over harmonicity. Hence, T may be supposed to be an

element of such a space. This is important if one tries to assure the convergence of an approximation algorithm, where more and more base functions (or data) are used. On the other hand, T may be approximated arbitrarily well using functions which are not even harmonic, or which belongs to spaces having functions with sets of harmonicity larger than the set outside the solid earth, cf. the Runge-Krarup theorem, (Krarup, 1969).

We will now consider an important space, where the inner product in some way has been selected so that it suits T optimally, but as we will see does not contain T itself. The idea is to use the metric used in practice to evaluate the quality of an approximation, namely the integral of the square of the difference between the true values and the estimated values. The basic operator, is the mean value operator

$$M(f(P)) = \frac{1}{A} \int f(P)\, d\sigma \tag{3.20}$$

where A is the area or volume over which the integral is calculated. In statistics, the mean value is calculated as the integral with respect to a probability measure, but the actual estimate is the usual mean value of the observations (assuming they all have the same weight). When computing the so-called empirical covariances, we will use the same mean value operator, but introduce (in some cases) a further mean over all directions, so that the mean is formed over all point pairs with the same spherical distance and having the same distance from the origin. (See Goad et al. (1985) and Tscherning (1985, section 3.2) for a discussion of the estimation and modelling of empirical covariance functions.) Let us evaluate this covariance function on the sphere with radius R. Then

$$C(P,Q) = M(T(P) \cdot T(Q)) =$$

$$\frac{1}{8\pi^2} \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \int_{\alpha=0}^{2\pi} T(\varphi,\lambda,R)T(\varphi',\lambda',R)\cos\varphi \, d\varphi d\lambda \, d\alpha \, . \tag{3.21}$$

If we work in sperical approximation, we may suppose that T can be developed in normalized solid spherical harmonics,

$$T(\varphi,\lambda,r) = \frac{GM}{R} \sum_{i=2}^{\infty} \sum_{j=-i}^{i} \bar{a}_{ij} \bar{V}_{ij} (\varphi,\lambda,r) \, . \tag{3.22}$$

Then as shown in Heiskanen and Moritz (1967, section 7.3) we have

$$C(P,Q) = C(\phi_{PQ}) = \sum_{i=2}^{\infty} \sigma_i P_i(\cos\phi_{PQ}), \sigma_i = (\frac{GM}{R})^2 \sum_{j=-1}^{i} \bar{a}_{ij}^2 \, , \tag{3.23}$$

where σ_i are the so-called (potential) degree-variances.

The general expression for r or r' ≠ R contains a factor $(\frac{R^2}{rr'})^{i+1}$. It is well known that the degree-variances tend to zero for $i \to \infty$ somewhat faster than i^{-3}, see (Tscherning, and Rapp 1974), Forsberg (1984).

The covariance function may as explained in example 3.8 (in a slightly different context) be interpreted as a reproducing kernel, where the norm has been fixed, by defining which vectors form an orthogonal system of eigenfunctions. For the function (3.23) the eigenfunctions are the solid spherical harmonics, and the eigenvalues are $(2i+1)\sigma_i$ (where we suppose $\sigma_i > 0$, i=1). The new base vectors are then

$$V^*_{ij} = (\sigma_i/(2i+1))^{\frac{1}{2}} \bar{V}_{ij}$$

and

$$T(P) = \sum_{i=2}^{\infty} \sum_{j=-i}^{i} \frac{\bar{a}_{ij}(2i+1)^{\frac{1}{2}}}{\sigma_i^{\frac{1}{2}}} V^*_{ij}(P)$$

Hence

$$\| T \|^2_C = \sum_{i=2}^{\infty} \frac{2i+1}{\sigma_i} \sum_{j=-i}^{i} \bar{a}^2_{ij} = \sum_{i=2}^{\infty} 2i+1 = \infty ,$$

which shows that T does not belong to the space. (This paradox has been solved in F. Sansò: Statistical Methods in Physical Geodesy, published also in these Lecture Not

3.4 Representers of functionals in a reproducing kernel Hilbert space

Using the reproducing kernel, the representer of a linear functional is easily obtained,

$$L(f) = L(f(P), K(P,Q)) = (f(P), L(K(P,\cdot)),$$

i.e. the representer is $L(K(P,\cdot))$, which we in the following will denote K(P,L). If two functionals are applied on K(P,Q) we will use

$$L_1 L_2(K(P,Q)) = K(L_1,L_2) .$$

This is equal to the inner product of the two functionals, because it is defined as the inner product of their representers:

$$(L_1,L_2)_* = (K(P,L_1), K(P,L_2)) = L_1(K(P,Q), K(P,L_2)) = K(L_1,L_2) \qquad (3.24)$$

And then

$$\| L \|^2_* = K(L,L).$$

We have earlier shown how a covariance function may be interpreted as a reproducing kernel in a Hilbert space, and also conversely explained how a reproducing

kernel may be regarded as a covariance function of the evaluation functionals, again being equal to the inner product of the two functionals. From eq.(3.24) we may conclude, that also the inner product of two arbitrary functionals (in H^*) may be interpreted as covariances, obtained by applying the linear functionals on the basic covariance function (reproducing kernel). In this manner we can find the covariance function of the gravity anomaly functionals, by applying eq. (2.6) and (3.23):

$$C(\Delta g_P, \Delta g_Q) = \sum_{i=2}^{\infty} \sigma_i \frac{(i-1)^2}{R^2} \left(\frac{R^2}{rr'}\right)^{i+2} P_i(\cos\psi), \tag{3.25}$$

However, let us look at some simple examples.

Example 3.9. Let us find the representers of some simple functionals in the space with the reproducing kernel given in example 3.4

(a) $L(f) = f(\frac{1}{2}) \Rightarrow \ell(x) = K(x,\frac{1}{2}) = -\frac{15}{16}x^2 + \frac{3}{2}x + \frac{21}{16}$

(b) $L(f) = f(0) \Rightarrow \ell(x) = K(x,0) = -\frac{15}{4}x^2 + \frac{9}{4}$

(c) $L(f) = \frac{\partial f}{\partial x}\big|_0 \Rightarrow \ell(x) = \frac{\partial}{\partial x}\big|_0 K(x,y) = 3x$ ●

Example 3.10. A subspace of the space described in example 2.1 has reproducing kernel

$$K(P,Q) = \sum_{i=2}^{\infty} \frac{A}{(i-1)^2} \left(\frac{R^2}{rr'}\right)^{i+1} P_i(\cos\psi), \tag{3.26}$$

where A is in units of m^4/s^4. Having selected the kernel on this specific form makes the inner product of the gravity anomaly functionals very simple:

$$K(\Delta g_P, \Delta g_Q) = \left((-\frac{\partial}{\partial r}\big|_P - \frac{2}{r}L_P), (-\frac{\partial}{\partial r}\big|_Q - \frac{2}{r'}L_Q)\right)_*$$

$$= \sum_{i=2}^{\infty} \frac{A}{rr'}\left(\frac{R^2}{rr'}\right)^{i+1} P_i(\cos\psi) = \frac{A\,R^2/(rr')^2}{(1-2\frac{R^2}{rr'}\cos\psi + (\frac{R^2}{rr'})^2)^{\frac{1}{2}}} - \frac{AR^2}{(rr')^2}$$

$$- \frac{AR^4}{(rr')^3}\cos\psi \,. \tag{3.27}$$

Let us evaluate the covariance for $r = r' = \sqrt{2}R$. Then

$$K(\Delta g_P,\Delta g_Q) = \frac{A/(4R^2)}{(5/4-\cos\psi)} - \frac{A}{4R^2} - \frac{A}{8R^2}\cos\psi \quad .$$

For $P = Q$ we get

$$\| \frac{\partial}{\partial r} - \frac{2}{r}L_P \|_*^2 = \frac{A}{4R^2}(2-1-\tfrac{1}{2}) = \frac{A}{8R^2} \quad .$$

Note, that if the summation started from i=0, then $K(\Delta g_P,\Delta g_Q)$ would have been positive everywhere. By subtracting two terms, the function will get two zero-points, i.e. for two values of the spherical distance will the gravity anomaly functionals be independent.

The "trick" of selecting a model for the degree-variances (3.23) which makes the derived inner products for the most commonly used functional very simple has been use in Tscherning and Rapp (1974), where

$$\sigma_i \simeq \frac{A}{(i-1)(i-2)(i+B)} \quad , \quad B \in Z_+ \quad . \tag{3.28}$$

The reason for not using the denominator $(i-1)^2$ like in eq. (3.27), is that a closed expression can be found for a reproducing kernel using (3.28), while no closed expression is known for eq. (3.26).●

We conclude this chapter by pointing out how the norms and degree-variances of the associated reproducing kernels are related. In general the degree variances behave like i^{-2k+2} or i^{-2k+1} if we use a $H^k(\Omega)$, $H^k(\sigma)$-norm, respectively. This is important, if we want to select a space where the evaluation functionals, or gravity anomaly functionals have finite norm at the boundary. However by selecting a Hilbert space of functions harmonic in a set just slightly larger than the set outside the Earth, makes all functionals of geodetic interest nice and continuous. We only miss the possibility to estimate the variance of density which correspond to a certain estimated variance of T(P) in the set between the Earths surface and the new set of harmonicity. And this may unfortunately be exactly where we would like to know something about the density variation.

1. Introduction

Many methods have been proposed (and a few used in practice) for gravity field approximation:

-series expansion in orthogonal functions,
-linear combinations of potentials of point masses, multipoles, or mass lines,
-linear combinations of harmonic splines, kernel functions or finite elements,
-least squares or minimum norm collocation.

The methods have in common that the approximation to the gravity potential is a linear combination of a number of base functions. They differ by their choice of base functions and by the method used to determine the coefficients of the linear combination.

Some of the methods are based on clear mathematical principles, even admitting proofs of convergence for the situation where the number of base functions increase. But in practice they do not always give good results. Conversely, methods have been proposed based on mainly heuristic arguments, which in practice works very well, and for wich proofs of convergence are not known.

Functional analysis is the branch of mathematics, which provides us with the tool to analyse, and hopefully better understand, the functioning of the different methods. The main propose of these lecture notes is therefore to give an introduction to the basic concepts in functional analysis. However, large parts of basic functional analysis is now well known to geodesists and geophysists, at least as it is presented in basic mathematical texts. Also potential theory, and physical geodesy is well known but I feel that the close connection between functional analysis and physical geodesy maybe is not valued well enough.

As in my lecture notes prepared for earlier summerschools (Tscherning, 1975, 1978a,1985) (from which much material presented here has been taken) I will skip nearly all proofs and concentrate on examples. Some of these are the solutions to excercises contained in earlier lecture notes. This might interest the reader, who has not been able to verify his solutions to these excercises.

In the following we will mainly deal with objects which are functions defined in a subset of a three-dimensional enclidian space R^3, which we will denote Ω. The boundary will be denoted σ or ω. Functions will denoted f,g,h or l and a,b,c are reserved for real or complex numbers. Points in R^n will be denoted x and y with coordinates $\{x_i\}$ and $\{y_i\}$. A fat "period" ● will denote end of example.

2. Linear vector spaces, dual spaces, norms and inner products

2.1 Linear vector spaces.

A linear vector space over the real (or complex) numbers is a set of elements (or vectors) f,g,h,.....,for which it is possible to add two elements or to multiply an element with a scalar. There must also exist a unique zero element, 0, and every element must have an inverse, -f with $f+(-f)=0$. The following rules must hold: $f+g=g+f$, $f+(g+h)=(f+g)+h$, $0+f=f$, $a(f+g)=af+ag$, $a(bf)=(ab)f$, $(a+b)f=af+bf$, and $1\cdot(f)=f$, where 1 as written is the real number one.

The vectors which form a linear combination,

$$f=\sum_{i=1}^{n} a_i f_i, \qquad a_i \in R$$

are called linear dependent if there exist constants $\{a_i\}\in R^n$, so that $f=0$. Otherwise they are called independent. A vector space is said to have dimension n, if there exist n linearly independent vectors, while any set of n+1 vectors are dependent. If no n exist, the space is infinite dimensional.

A set of elements f_1, f_2,...., is a basis for the vector space, if they are independent and every element can be expressed uniquely as a linear combination of the elements.

Well known examples of finite dimensional spaces are the real enclidian space R^n and polynomials p(x) of maximal degree n. In general functions $f:\Omega\subseteq R^n\rightarrow R$ form a linear vector space since we are able to add functions and multiply functions by a constant. The space of functions with derivatives continuous up to order m is denoted $C^m(\Omega)$.

Example 2.1 · The harmonic functions outside a sphere i R^3.
Let Ω be the open set in R^3 outside a sphere with center of the origin and radius R, and consider the vector space of functions harmonic in Ω and regular at infinity,

$$\Delta f = \sum_{i=1}^{3} \frac{\partial^2 f}{\partial x_i^2} = 0 , \qquad \lim_{|x|\to\infty} |f(x)|x|| < \infty .$$

A finite dimensional sub space of this vector space is formed by talking all linear combinations of the solid spherical harmonics,

$$V_{ij}(P) = \left(\frac{R}{r}\right)^{i+1} P_{ij}(\sin\varphi) \begin{Bmatrix} \cos j\lambda \\ \sin|j|\lambda \end{Bmatrix} \begin{matrix} 0 \le j < i \\ -i \le j < 0, \end{matrix} \qquad (2.1)$$

where $0 \le i \le n$. The point P has spherical coordinates φ (latitude), λ (longitude) and r(distance from the origin), and $P_{ij}(\sin\varphi)$ are the associated Legendre functions of degree i and order $|j|$. The dimension of the subspace is $(n+1)^2$. ●

In general, linear vector spaces may be formed by taking all (finite) linear combinations of given elements. This is called the space spanned by the elements,

$$\text{span}\{f_i,\ i \in I\},$$

where I is an index-set. Note , that span $\{V_{ij},\ i = 0, \ldots,\infty ,\ |j| \le i\}$, does not contain all functions harmonic in Ω.

2.2 Dual spaces.

In practice we encounter, not functions, but measurements. Forgetting measurement errors, we generally regard the measurements as values of mappings from a function into the real numbers, for example the value of one of its derivatives in a point. An important class of functionals are the linear functionals, which we will denote by L, sometimes with a subscript affixed. The functional just mentioned is an example of a linear functional. In general they must fulfil

$$L(af+bg) = aL(f)+bL(g).$$

For linear vector spaces of functions the most important linear functional is the evaluation functional. $L_P(f) = f(P)$. It should be obvious that the set of linear functionals associated with a vector space, H, will form a new linear vector space H^* , the dual space. (We will later restrict H^* to the space only containing the bounded functionals).

Example 2.2 Linear functionals in R^n.
In R^n the linear functionals are the linear mappings

$$L(x) = \sum_{i=1}^{n} a_i x_i$$

Note, that also $\{a_i\}$ is an element of R^n, and that all linear functionals have a unique representation as the scalar product of a fixed vector $\{a_i\}$ with the variable vector x. We shall see later that this nice property is valid in a large class of function spaces. ●

Example 2.3 The coordinate functionals for functions harmonic outside or inside a sphere. It is well known that functions harmonic outside a sphere and regular at ∞ permit a representation through a convergent series

$$f(P) = \sum_{i=0}^{\infty} \sum_{j=-i}^{i} a_{ij} V_{ij}(P). \tag{2.2}$$

The functions harmonic inside a sphere and regular at 0 have a similar expansion with $V^o_{ij}(P)$ substituted for $V_{ij}(P)$, where $V^o_{ij}(P)$ are the inner solid spherical harmonics,

$$V^o_{ij}(P) = \frac{r^{2i+1}}{R^{2i+1}} V_{ij}(P) \tag{2.3}$$

The coordinate functionals $L_{ij}(f) = a_{ij}$ are evaluated by

$$L_{ij}(f) = \frac{1}{4\pi R} \int_\sigma f(P) V_{ij}(P) d\sigma = a_{ij}, \tag{2.4}$$

where σ is the surface of the sphere. (The integral should be taken as the limit over concentric spheres with radius converging towards R from above or from below.)●

Example 2.4 The basic measurement functionals.
The basic measurements functionals of physical geodesy are (with W denoting the gravitational potential and T the anomalous potential):

$$\text{Gravity: } g(P) = \left(\sum_{i=1}^{3} \left(\frac{\partial W}{\partial x_i}\right)^2 \right)^{\frac{1}{2}} \tag{2.5}$$

$$\text{Gravity anomaly and disturbance: } \Delta g(P) = -\frac{\partial T}{\partial r}\Big|_P - \frac{2}{r} T(P) \tag{2.6}$$

$$\delta g(P) = -\frac{\partial T}{\partial r}\Big|_P \tag{2.7}$$

Deflections of the vertical (meridian and prime vertical components):

$$\xi(P) = -\frac{1}{r\gamma} \frac{\partial T}{\partial \varphi}\Big|_P, \quad \eta(P) = -\frac{1}{r\gamma} \frac{\partial T}{\partial \lambda} \frac{1}{\cos\gamma} \tag{2.8}$$

$$\text{Gradiometry: linear combinations of } T_{ij}(P) = \frac{\partial^2 T}{\partial x_i \partial x_j}\Big|_P \tag{2.9}$$

$$\text{Height anomaly: } \zeta(P) = L_P(T)/\gamma = T(P)/\gamma, \tag{2.10}$$

where γ is normal gravity. (The indication of the point of evaluation may in the following be dropped.) The functionals (2.6)-(2.10) are linear(ized), while (2.5) is non-linear.

Example 2.5 The linear space spanned by a random function.
In this example one should keep in mind a set of (harmonic) functions all with a certain probality being equal to the anomalous gravity potential of the Earth. However, the basic probability space to be studied consist of the (linear) functionals giving for a fixed point P the values of the anomalous potential or the gravity anomaly in this point.

In general the starting point is a probability space (H, A, P), where H is an arbitrary set (generally denoted Ω in statistical literature), A is a Boolean σ-algebra of subsets of H and P is a probability measure. The stochastic or random variables, X, are mappings from H to $\mathbf{R}$. If these mappings are parameterized through an index set, T, a stochastic process $\{X_t,\ t \in T\}$ is defined, provided the joint probabilities are defined, cf. Parzen (1959). The index set is frequently the real line (t is time), but here we should think of T as being the set R^3.

The stochastic process is called a random function provided the second order moments are finite, i.e. $\int_H X_t^2\, dP < \infty$ for all t. In this case we will define a linear vector space as the one spanned by the random function,

$$L\{X_t,\ t \in T\} = \text{span}\{X_t,\ t \in T\} \tag{2.11}$$

If H is the linear vector space of example 2.1, a stochastic process is formed by the evaluation functionals $L_P(f) = f(P)$, with Ω as the index set. The probability measure may be defined by requiring all L_P, $P \in \Omega$ to be normally distributed with zero mean and covariance

$$\int_H L_P \cdot L_Q\, dP = \text{cov}(P, Q), \tag{2.12}$$

where cov(P,Q) is a positive definite function, $\Omega x \Omega \to R$.

Note, that the linear vector space (2.11) will not automatically be equal to H^*. We would also like that this space should contain the functionals, which give the value of the Laplace operator in a point, and that these stochastic variables should be equal to zero with probility one. This will be achieved, if cov(P,Q) is a harmonic function in each of its variables, and if we add to the space (2.11) all limits of sequences of stochastic variables which have finite variance. ●

2.3 Inner product and normed space.

Example 2.5 was in fact an example of what we now will define, namely a linear vector space with an inner product. In general it is a mapping $(\cdot,\cdot)$: $H \times H \to R$, which is symmetric, linear, homogeneous and positive, i.e. $(f,g) = (g,f)$, $(f_1 + f_2, g) = (f_1, g) + (f_2, g)$, $(af,g) = a(f,g)$ and $(f,f) \geq 0$, with zero only occuring for $f \equiv 0$. The most simple example is the inner product in R^n, $(x,y) = \sum_{i=1}^{n} x_i y_i$.

From the inner product a norm may be defined, $\|f\| = (f, f)^{\frac{1}{2}}$, which also will be positive, definite and homogeneous. The norm will furthermore fulfil the triangle inequality,

$$\|f+g\| \leq \|f\| + \|g\| \tag{2.13}$$

A norm need not to be an offspring of an inner product. The most simple example of a norm, which is not related to an inner product is the maximum norm,

$$\|f\| = \max |f(x)| \ , x \in \Omega, \ f \in C^0(\Omega).$$

Note that a normed vector space is also a metric space, since we to all elements f, g may associate their distance, $d(f,g) = \|f-g\|$. ●

Example 2.6 The Sobolev norms. Regard the linear vector space $C^m(\Omega), \Omega \subseteq R^n$. Define

$$D^{\alpha} f = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \cdots\cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f, \quad \sum_{i=1}^{n} \alpha_i = |\alpha| \leq s. \tag{2.14}$$

Then inner products may be defined for positive integer values of s on subsets of $C^m(\Omega)$, $m \geq s$ by

$$(f,g)_s = \sum_{|\alpha| \leq s} \int D^{\alpha} f \cdot D^{\alpha} g \, d\Omega \tag{2.15}$$

Norms may be defined as the p'th root of the integral of $(D^{\alpha} f)^p$. Then the subsets we will consider are these for which the corresponding norm is finite. The corresponding complete spaces (see next capter) are denoted Sobolev or $H^{s,p}$-spaces. We will here only consider spaces with p = 2, so we will drop the second superscript.

Norms may also be defined for non-integer (and even negative) values of s. Non-integer values are defined using Fourier-transforms of the functions, while norms corresponding to negative values of s are defined using

$$\|f\|_{-s} = \max_{g \in H^s} \frac{\int f \cdot g \, d\Omega}{\|g\|_s} \tag{2.16}$$

The Sobolev spaces are used in modern theory of boundary value problems, because the spaces are used to characterize for which spaces boundary value problems have a unique solution. They are also much used in practice, because solutions to for example the Laplace-equation, which are elements of $C^{\infty}(\Omega)$, may be approximated arbitrarity well using elements of the H^S-spaces, see e.g. Aubin (1972). ●

In a linear vector space with an inner product the Schwarz inequality is valid,

$$|(f,g)|^2 \leq \|f\|^2 \cdot \|g\|^2 \tag{2.17}$$

The equality sign is only valid when f and g are linearily dependent. This enables us to define an angle θ between two non-zero vectors f,g, because

$$-1 \leq \cos\theta = \frac{(f,g)}{\|f\| \cdot \|g\|} \leq 1 \tag{2.18}$$

With $\theta = 0$, f, g are parallel and when $\theta = 90^{o}$, the two elements are orthogonal, i.e. $(f,g) = 0$.

Example 2.7 Correlation as angle between vectors.
In the space (2.11) the covariance expresses an inner product,

$$E(x,y) = (x,y) = \int_H x \cdot y \, dP$$

$$\|x\| = E(x^2)^{\frac{1}{2}}$$

and the correlation is cosine of the angle

$$\cos(\theta) = E(x,y)/(E(x^2) \cdot E(y^2))^{\frac{1}{2}} \quad ●$$

From a set of linearly independent elements, a new set of orthogonal elements may be constructed using the so-called Gram-Schmidt orthonomalization procedure.

From the set $\{f_i, i = 1,.....,n\}$ we generate a new set $\{g_i, i = 1,..., n\}$ by

$$g_1 = f_1 / \|f_1\|,$$

and for k = 2,....., n

$$g_k = (f_k - \sum_{i=1}^{k-1} (f_k, g_i) g_i) / \|f_k - \sum_{i=1}^{k-1} (f_k, g_i) g_i\| .$$

In fact in each step we first subtract from f_k its projection on the space spanned by the k-1 first vectors and then divide by the norm of this vector. (Hereby we do not only arrive to an orthogonal set, but all vectors have norm equal to 1).

Example 2.8 The first Legendre polynomials.
Regard the 3-dimensional vector-space spanned by the polynomials $f_0=1$, $f_1=x$, $f_2=x^2$

on the interval from -1 to 1, with inner product

$$(f,g) = \tfrac{1}{2}\int_{-1}^{1} f(x)g(x)\,dx$$

Using the Gram-Schmidt procedure we get

$$\| f_0 \|^2 = \tfrac{1}{2}\int_{-1}^{1} 1\cdot dx = 1, \text{ i.e. } g_0 = 1.$$

Since $(g_0, f_1) = 0$ and $\|f_1\|^2 = \frac{1}{2}\int_{-1}^{1} x^2dx = 1/3$, then

$$g_1(x) = \sqrt{3}\cdot x$$

From

$$(g_0, f_2) = \tfrac{1}{2}\int_{-1}^{1} x^2\,dx = 1/3, (g_1, f_2) = \tfrac{1}{2}\int_{-1}^{1} x^2 x\sqrt{3}\,dx = 0$$

and

$$\| f_2 - 1/3\cdot 1 \|^2 = \tfrac{1}{2}\int_{-1}^{1} (x^2 - 1/3)^2dx = (2/(3\sqrt{5}))^2$$

we get

$$g_2(x) = (x^2 - 1/3)/(2/(3\sqrt{5})) = \sqrt{5}(3/2\,x^2 - 1/2).$$

A continuation of the process with x^3, x^4, etc. will produce all the normalized Legendre polynomials, $\bar{P}_n(x) = \sqrt{2n+1}\ P_n(x)$. ●

Let then $g_0, g_1 \ldots\ldots$, be a finite or infinite sequence of orthonormal elements. The series

$$\tilde{f} = \sum_{n=1}^{\infty} (f, g_n)g_n = \sum_{n=1}^{\infty} a_n\, g_n \tag{2.19}$$

is called the orthogonal expansion or Fourier expansion of f with the coefficients (or coordinates) $a_n = (f, g_n)$.

Remember that $\tilde{f}(P)$ not necessarily is equal to $f(P)$. It is well known that the harmonic functions of example 2.1 with boundary values in $H^0(\sigma)$ may be expressed as a series in (normalized) spherical harmonics,

$$\left.\begin{aligned} &\bar{V}_{ij}(P) = \left(2(2i+1)\,\frac{(i-j)!}{(i+j)!}\right)^{\frac{1}{2}} V_{ij}(P) \qquad j \neq 0 \\ \text{and} \\ &\bar{V}_{io}(P) = (2i+1)^{\frac{1}{2}}\, V_{io}(P). \end{aligned}\right\} \tag{2.20}$$

Example 2.9 The function $g(x) = -2x+1$ is an element of the space of example 2.8. The Fourier-expansion of g is $g(x) = 1 \cdot g_0(x) - \frac{2}{\sqrt{3}} g_1(x)$. The norm of $g(x)$ is

$$\|g(x)\|^2 = \tfrac{1}{2}\int_{-1}^{1}(-2x+1)^2 dx = \frac{7}{3} = 1 + (2/\sqrt{3})^2. \bullet$$

Example 2.10 Use of the orthogonality concept to characterize the set of density distributions with zero exterior potential.

We regard as possible density distributions the set of all square-integrable functions in Ω_0. Let $H(\Omega_0)$ be the set of all harmonic functions regular in Ω_0, zero outside Ω_0. Regard then the functions which are equal to the Laplace operator applied on a two times differentiable function, which is zero in a set enclosed in Ω_0. Then $f = \Delta(g)$. This function will generate g as its potential, i.e. have potential zero outside Ω_0, and it will be orthogonal on all functions harmonic in Ω_0. Then suppose $\Delta(h) = 0$ in Ω_0, and using Greens first identity:

$$\int h \cdot f \, d\Omega_0 = \int h\Delta(g) d\Omega_0 = \int \Delta(h) g \, d\Omega_0 + \int_\omega h \frac{\partial g}{\partial n} d\omega - \int_\omega g \frac{\partial h}{\partial n} d\omega = 0.$$

Hence, f and h are orthogonal. ●