Universität Stuttgart

Schriftenreihe der Institute des Studiengangs Geodäsie und Geoinformatik

Technical Reports
Department of Geodesy and Geoinformatics

C. Kotsakis, M.G. Sideris

The Long Road from Deterministic Collocation to Multiresolution Approximation

ISSN 0933-2839 Report Nr. 1999.5

The Long Road from Deterministic Collocation to Multiresolution Approximation

by

Christopher Kotsakis and Michael G. Sideris

Department of Geomatics Engineering, University of Calgary

2500 University Drive N.W.

Calgary, Alberta, Canada T2N 1N4

July 1999

Table of Contents

Abstract 1

1. **Introduction** 2
    - 1.1 Deterministic Collocation 2
    - 1.2 Stochastic Collocation 3
    - 1.3 Statistical Collocation: A Compromise 4
    - 1.4 Some Important Existing Problems 5
2. **The General Problem of Linear Approximation in Hilbert Spaces** 6
    - 2.1 Inversion Scheme for Case 1 9
    - 2.2 Inversion Scheme for Case 2 9
    - 2.3 Comments 10
3. **Numerical Stability and the Role of Frames** 12
    - 3.1 Frames 13
    - 3.2 Practical Computations Using Frames 15
    - 3.3 Frames and Linear Approximation in Hilbert Spaces 17
    - 3.4 An Example 19
    - 3.5 A Note on Ill-Posed Problems 21
4. **The Hilbert Space Choice Problem** 21
    - 4.1 Data Type and Configuration 23
    - 4.2 Conditions for Stable and Convergent Linear Approximation 23
    - 4.3 Remarks on Previous Methodologies 27
5. **The Connection with the Multiresolution Concepts** 30
    - 5.1 Linear Approximation as a Multiresolution Approximation 31
    - 5.2 Final Remarks 34
6. **Conclusions and Future Work** 36

Acknowledgements 36

References 37

Abstract

The aim of this report is twofold. At first, the general problem of linear approximation in Hilbert spaces is presented. The formulation and the solution of this problem is made in a very detailed way, illustrating important aspects which are often overlooked in many textbooks and monographs found in the geodetic literature. A relatively recent mathematical tool of functional analysis, called a frame, is also used to study the important issue of stability for the solution of the linear approximation problem. The second, and most important, scope of the report is to explore the relation between the choice of the Hilbert space in which we perform the linear approximation, and the stability/convergence properties of the solution for increasing data density. In particular, it will be shown that a constantly stable and convergent (in the sense of infinitely dense data) linear approximation scheme in a Hilbert space H requires the incorporation of a multiresolution structure of subspaces in H, similar to the well-known dyadic subspace sequences in $L^2(\Re)$ that were introduced by Stephane Mallat a decade ago and have created a new exciting alternative approach to wavelet theory.

1. Introduction

Before we study in rigorous mathematical terms the linear approximation problem in Hilbert spaces and its associated stability and convergence issues, a descriptive overview of the general *collocation concept* in Physical Geodesy will be first given in this section. Collocation is widely known as an optimal linear estimation method used for gravity field modelling. Behind this vague definition there are lying two fundamentally different viewpoints, with correspondingly different mathematical/physical concepts and ideas associated with them. Both approaches have certain advantages and drawbacks, and they have been the subject of extensive debate in the geodetic scientific community over the past three decades.

1.1 Deterministic Collocation

The first approach, which will be called deterministic collocation, uses a Hilbert space setting according to which the unknown field under consideration f (e.g., the anomalous potential of the Earth) is modelled as an individual element in a reproducing kernel Hilbert space (RKHS) H. The available discrete observations for the unknown field are also modelled as continuous linear functionals $L_i f$, belonging in the dual Hilbert space H' of H. The optimally approximated field $\hat{f}$ is then defined as the smoothest function, in the topology of H, that satisfies the given functionals. Such a linear approximation method is not of course an exclusive privilege of Physical Geodesy problems, but it has been borrowed from other areas of mathematics where it is often found under the name *minimum-norm interpolation in RKHS* (see, e.g., Davis, 1975). The original idea of using such deterministic Hilbert space methods for gravity field modelling is due to Torben Krarup, who in his famous publication (Krarup, 1969) developed a pioneering framework for solving discrete boundary value problems in Hilbert spaces of harmonic functions outside a certain spherical approximation of the Earth. The reformulation, however, of the gravity field determination problem as an underdetermined *discrete* boundary value problem (bvp), instead of a *continuous* bvp (i.e., Stokes, Molodensky), whose solution should employ only the finite number of the available discrete data in a certain optimal fashion, was made earlier than Krarup's developments by Bjerhammar (1964). The interpolatory character of the discrete bvp, along with the proposed idea by Bjerhammar for the downward reduction of the discrete data on a certain reference sphere, have contributed in many several ways to a clarification of the conceptual foundations in modern Physical Geodesy; see also, Bjerhammar (1973, 1975, 1987). More details on the deterministic aspects of collocation can also be found in Dermanis (1976), Leglemann (1979), Meissl (1976), Moritz (1980), and Tscherning (1986).

Deterministic collocation 'suffers' from two important problems, one of which is the so-called norm choice (or reproducing kernel choice) problem; see, e.g., Dermanis (1977). In order to use the method, and to actually compute an approximation for the unknown field, we need to know the reproducing kernel in the Hilbert space H, which in turn requires the prescription of a specific norm (topology) for H. The choice of such norm $\| \ \|_{\mathrm{H}}$, or reproducing kernel, not only does it affect the physical interpretation of the approximation, but it also influences other important aspects like the stability and convergence properties of the solution algorithm, as well as the admissible spatial configurations of the data needed to obtain such stable and convergent solutions (Rummel et al., 1979). These issues will be further explored later in this report. The second problem in deterministic collocation stems from the lack of efficient measures to evaluate the accuracy of the minimum-norm interpolation. Although there do

exist rigorous bound values for the approximation error $\left\| f - \hat{f} \right\|_{\mathrm{H}}$ that can characterize the overall performance of the linear approximation in a given Hilbert space, their use is of limited practical importance and their actual computation requires the complete knowledge of the unknown field itself (see, e.g., Dermanis, 1976; Tscherning, 1986).

1.2 Stochastic Collocation

The second fundamental approach in linear approximation problems of Physical Geodesy constitutes the stochastic version of collocation. According to this approach the unknown field f is modelled as a zero-mean stochastic process, and the available discrete observations related to the field are considered as zero-mean random variables. The optimally approximated (or more precisely, predicted) field $\hat{f}$ is now defined as the one satisfying the minimum mean square error (MMSE) principle, i.e. $E\left\{(f - \hat{f})^2\right\} = \min$, where E denotes the expectation operator in a probabilistic sense. The final solution is obtained by additionally requiring that $\hat{f}$ should be linearly related to the available discrete data. As in the deterministic collocation case, this stochastic approximation framework has also been borrowed from other areas of mathematics and applied sciences (communication engineering, signal analysis) where it was originally developed. The underlying theory is formally known as the Wiener-Kolmogorov (W-K) linear prediction theory, pointing to the pioneering work of the two scientists back in the 1930s and 1940s. An excellent review paper on many different aspects of W-K theory, with more than 350 relevant references on the subject, is Kailath (1974). The original introduction of the W-K methodology in Physical Geodesy approximation problems should be attributed to the work of Kaula (1959) and Moritz (1962). For a detailed treatment of the stochastic principles in collocation theory and related applications in gravity field modelling see, e.g., Dermanis (1976), Dermanis and Sanso (1997), Bjerhammar (1982), Moritz (1980), Sanso (1986).

For the actual computation of the predicted field $\hat{f}$ (or any other field of interest that is linearly related to the original f) in the W-K framework, it is only required to know the covariance (CV) function $C(P,Q) = E\{f(P)f(Q)\}$ of the unknown field, which describes the stochastic behavior of f between pairs of points in its domain. Often, in practical applications, the unknown field is additionally modelled as a stationary and ergodic stochastic process. The main benefit of such an assumption is that the estimation of the (generally) unknown CV function becomes possible using the available realization of the unknown field (i.e. discrete observations). Furthermore, in the stationary case it is particularly beneficial to transform the approximation problem in the frequency domain using the Fourier transform, since the computational effort of the solution algorithm is significantly reduced due to the de-correlation ('whitening') properties of the Fourier transform over stationary stochastic signals; for more mathematical details, see, e.g., Parzen (1967), Papoulis (1993), and for geodetic applications of frequency domain collocation, see, Eren (1980), Schwarz et al. (1990), Sideris (1995), Sanso and Sideris (1995), and Nash and Jordan (1978).

The main drawback in the stochastic viewpoint of collocation is that it is not physically acceptable, since the gravity field of the Earth is definitely <u>not</u> a stochastic phenomenon as the W-K theory requires. For an amusing, as well as interesting, discussion on this aspect, see Sanso (1978) and Moritz and Sanso (1980). Furthermore, modelling the deterministic gravity

field as a stochastic process creates important problems regarding the physical interpretation of the accuracy measures that we often use to evaluate our linear prediction. The formalism of variance-covariance propagation allows us to easily obtain the variances and covariances of the prediction error, $e(P) = f(P) - \hat{f}(P)$, at the points where the prediction is applied. Such accuracy information has a purely probabilistic nature that is physically meaningless for all of our purposes. Note that all the discussions so far are referred to a noiseless data setting.

On the other hand, the exact equivalence between the final solution algorithms for both the deterministic and stochastic collocation, which occurs when we identify the reproducing kernel used in the former approach with the covariance function employed in the latter, makes it possible to develop an intermediate 'semi-stochastic' viewpoint in collocation that can eliminate, to some degree, most of the above pitfalls in the two individual formulations.

1.3 Statistical Collocation: A Compromise

According to the concept of statistical collocation, instead of using quantities defined through 'experiment repetitions' via the probabilistic expectation operator E, we employ certain *spatial statistics* measures to describe the behavior of the unknown gravity field, as well as the behavior of its prediction error. In this way, the CV function of an unknown signal is now defined in a purely deterministic sense using a spatial averaging operator M. For example, in the one-dimensional case, the CV function is defined as

$$C(P,Q) = M_{\tau}\{ f(P+\tau) f(Q+\tau) \} \tag{1.1}$$

where the subscript τ means that the operator M is applied to all the different spatial locations of the pair $f(P) f(Q)$. A common choice for the operator M, which actually makes also possible the efficient incorporation of Fourier methods in the statistical collocation framework, is the classic integral

$$C(P,Q) = \int f(P+\tau)\, f(Q+\tau)\, d\tau \tag{1.2}$$

By a simple change of variables, it is easily seen that the last equation is reduced to the 'stationary' form

$$C(\xi) = C(|P-Q|) = \int f(P)\, f(P+\xi)\, dP \tag{1.3}$$

For multi-dimensional problems, we should additionally consider more translation (and possibly directional-azimuth) parameters in the definition of the averaging operator M, instead of a single translation parameter τ. In such cases, a customary terminology calls for *homogeneous* CV functions (translation parameters only in M), or *isotropic* CV functions (translation and azimuth parameters in M).

As in the stochastic collocation case, the optimal approximation $\hat{f}$ is again defined as the one satisfying a MMSE principle which is now expressed through the spatial operator M. In the 1D case, the error principle takes the form

$$M_\tau\left\{e^2(P,\tau)\right\}=M_\tau\left\{[f(P)-\hat{f}(P,\tau)]^2\right\}=\min \qquad (1.4)$$

The minimization principle (1.4) takes into account the fact that, if we translate by τ the available spatial configuration of the data points, we will generally obtain some new observation values which will produce a different approximation for the unknown field. In this way, the approximation error becomes also a function of the 'position' τ of the given data point configuration with respect to the reference system used to describe the unknown field. The statistical collocation solution will minimize the mean square approximation error over all possible positions of the given data point geometry. Two more conditions are also taken into account in order to finally compute the solution, which are: (i) linearity of the solution $\hat{f}(P)$ with respect to the available discrete observations, and (ii) translation-invariance of the solution with respect to corresponding changes of the reference system. The last property will ensure that, if $\hat{f}(P)$ is the optimal linear approximation for $f(P)$, then $\hat{f}(P+\tau)$ should be the corresponding optimal linear approximation for $f(P+\tau)$. In the multi-dimensional case, we should also assume approximation invariance under more general affine transformations of the reference system, which may possibly include rotational-invariance. The use of this *spatio-statistical* version of collocation was already proposed in the classic book of Heiskanen and Moritz (1967), but it is actually Sanso (1978) who first formulated a rigorous and complete mathematical setting for the method. A relevant and extensive discussion can also be found in Moritz (1978, 1980).

The final solution formula for the approximated field under the statistical collocation concept is similar to the ones obtained under either the deterministic or the stochastic approach, where instead of a reproducing kernel or a CV function of a stochastic signal, we now use a spatio-statistical CV function defined in a purely deterministic sense by eq.(1.1). In this way, the MMSE principle (1.4) automatically gives also rise to a minimum-norm solution in a Hilbert space with reproducing kernel equal to the deterministic CV function given by eq.(1.1). The norm that corresponds to this 'induced' Hilbert space, however, will not be generally the same with the norm used to quantify the approximation error in the optimal principle (1.4). For an interesting topological paradox that exists during this identification procedure, see Tscherning (1977).

1.4 Some Important Existing Problems

Regardless of which of the previous three approaches we adopt, the linear approximation of an unknown field from discrete data always requires the inversion of an $N\times N$ symmetric matrix, where N is the number of the available observations. Such a numerical task usually creates important problems in terms of the required computational effort and the stability of the solution algorithm. Although the computing time/storage requirements can be significantly reduced via special algorithmic and modelling techniques (see, e.g., Bottoni and Barzaghi, 1993; Sanso and Schuh, 1987) and/or Fourier transform methods (see, e.g., Sideris, 1995), the stability problem is not generally overcome by them. For example, if a very dense data configuration is used relatively to the 'spread' of the selected CV function (or reproducing kernel), then the resulting matrix will be highly ill-conditioned irrespectively of the domain (space or frequency) in which we perform its inversion. In this way, the establishment of convergence properties for the solution as the data density increases becomes also very difficult, if not impossible.

Apart from the stability/convergence problem, there exists one more issue of special importance in collocation. The introduction of the deterministic optimal criterion (1.4) by Sanso has been perceived by many geodesists only as an attempt for assigning a non-stochastic interpretation to Moritz's (or Wiener's) prediction formulas, overcoming in this way the non-existence of a stochastic gravity field. Along with this perception, however, it has also remained the false belief that we still need to model the gravity field as a stationary stochastic process because of the 'stationary' form of Sanso's deterministic CV function. This is believed to provide serious limitations in the whole approximation procedure of collocation, since the actual behavior of the gravity field is 'non-stationary'. Besides the fact that the above claim is meaningless since the notion of stationarity is not defined at all for deterministic signals (see also the related discussion in Sanso, 1978), we personally believe that Sanso's formulation is a very powerful and complete tool as it is, without having the need to be considered as a 'supplement' to W-K theory for Physical Geodesy problems. It can actually be shown that the use of 'stationary' CV functions, like (1.3), does not prohibit us from studying the local irregular details of the gravity field individually in different parts of its domain, thus overcoming any 'stationarity-restriction' concerns about its behavior; see Kotsakis (1999a,b).

From the above two important problems in collocation theory (stability/convergence, stationarity), we will deal only with the first one in this report. Its treatment, actually, will be seen to open the way for wavelet-like approximation models, answering thus in an indirect way the 'stationarity' problem as well. Our starting point will be the deterministic viewpoint in collocation, from which throughout the report we will gradually arrive at a multiresolution version, without the need of using the intermediate stochastic 'burden'.

2. The General Problem of Linear Approximation in Hilbert Spaces

An infinite-dimensional, separable Hilbert space H is given. For an unknown function $f \in \mathrm{H}$, there are available 'observations' b_n which have a linear dependence on f, i.e.

$$b_n = < f , g_n > \qquad , \qquad n \in \Im \tag{2.1}$$

where < , > denotes the inner product in the Hilbert space H, and g_n are known elements of the same space H. The index set $\Im$ will be assumed a finite subset of the integers for the moment, i.e. only a finite amount of observations is available. The problem is how to recover the unknown f using the observation values b_n and the 'observational representers' g_n.

Let us denote by V the Hilbert subspace of H, defined as the closed linear span of the finite sequence $\{g_n\}_{n \in \Im}$. Then, the unknown function f can be uniquely decomposed into two orthogonal elements as follows:

$$f = f_{\mathrm{V}} + f_{\mathrm{V}^{\perp}} \tag{2.2}$$

where f_V is the orthogonal projection of f onto V, and $f_{V^\perp}$ is the orthogonal projection of f onto the Hilbert subspace $V^\perp$ (orthogonal complement of V in H, i.e., $V \oplus V^\perp = H$). In this way, the observation equation (2.1) takes the following form:

$$\begin{aligned} b_n &= < f_V + f_{V^\perp} , g_n > \\ &= < f_V , g_n > + < f_{V^\perp} , g_n > \\ &= < f_V , g_n > \end{aligned} \qquad (2.3)$$

It is seen that the available observations (2.1) can only determine partially the unknown field f. In fact, they can determine only its orthogonal projection f_V onto the finite-dimensional subspace V spanned by the known family $\{g_n\}$, and they contain no information at all about the second component $f_{V^\perp}$. Since the observations do not supply any information on the orthogonal complement of f_V, the desired approximation $\hat{f}$ for the unknown field can be expressed in the general linear form

$$\hat{f} = f_V = \sum_{n \in \Im} a_n g_n \qquad (2.4)$$

where a_n are the unknown coefficients with respect to the sequence that generates the actual 'solution space' V. Using (2.4), the observation equations (2.3) yield

$$b_n = < \sum_{k \in \Im} a_k g_k , g_n > = \sum_{k \in \Im} a_k < g_k , g_n > \quad , \quad n \in \Im \qquad (2.5)$$

or, by using matrix notation,

$$\begin{bmatrix} b_1 \\ \vdots \\ b_n \\ \vdots \\ b_N \end{bmatrix} = \begin{bmatrix} < g_1, g_1 > & \cdots & < g_1, g_N > \\ \vdots & < g_n, g_k > & \vdots \\ < g_N , g_1 > & \cdots & < g_N , g_N > \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_k \\ \vdots \\ a_N \end{bmatrix} \qquad (2.6a)$$

$$\mathbf{b} = \mathbf{G}\,\mathbf{a} \qquad (2.6b)$$

where N denotes the finite number of the available observations. It will also be useful to express the original observation equations (2.1) in the following operator form:

$$\mathbf{b} = U\, P_V\, f = U\, f_V \qquad (2.7)$$

where P_V is the orthogonal projector from the Hilbert space H onto its Hilbert subspace V, and U is a linear operator defined as follows:

$$U : \mathrm{V} \to l^2(\Im) \qquad f_{\mathrm{V}} \to < f_{\mathrm{V}}, g_n > \qquad \forall\ n \in \Im \tag{2.8}$$

The role of the operator U is basically to take every function in the solution space V and to compute its inner products with respect to the sequence $\{g_n\}_{n\in\Im}$. Since $\{g_n\}$ is a finite set of functions, the collection of all these inner products, for every $f_{\mathrm{V}} \in \mathrm{V}$, can be considered as an individual element **b** of the Hilbert space $l^2(\Im)$. Also, since the family $\{g_n\}$ spans the whole subspace V, the operator U will be always an injective (one-to-one) operator.

The computation of the recoverable part f_{V} of the unknown field f requires the inversion of the operator U, or equivalently the inversion of the symmetric $N\times N$ matrix **G** in (2.6). In order to perform such an inversion, we will distinguish the following two cases regarding the behavior of the observational representers g_n:

Case 1 : $\{g_n\}_{n\in\Im}$ is a linearly independent set
Case 2 : $\{g_n\}_{n\in\Im}$ is a linearly dependent set

A geometrical illustration of the above two cases is shown in figure 2.1.

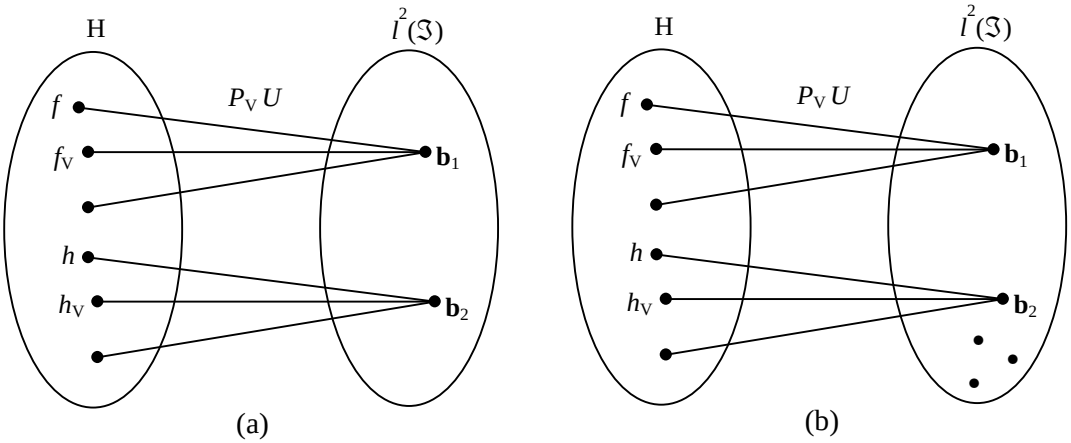

Figure 2.1: Mapping type of the observation equations for *Case 1* (a) and *Case 2* (b)

The only difference between the two cases is that the operator U is surjective (onto) in the first case, whereas in the second it is not. The internal correlations existing between the observation values $< f, g_n >$ when the set $\{g_n\}$ is linearly dependent, make the reproduction of every *N*-tuple of numbers in the observation space impossible. Since we are interested in the inversion of the operator U only, we can also create the following representative diagram:

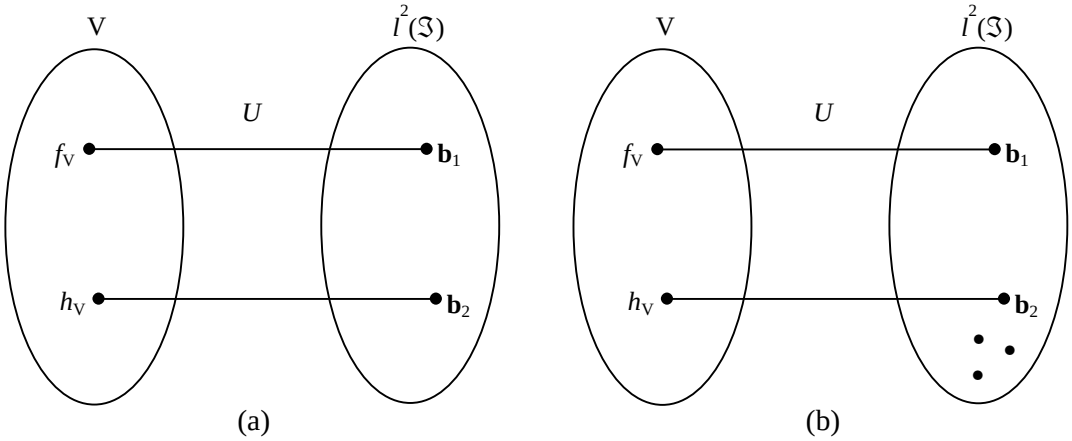

Figure 2.2: Geometrical interpretation of the operator U for *Cases 1* (a) and *2* (b)

For both cases 1 and 2, there will exist a unique *pseudo-inverse* operator $\tilde{U}^{-1}$ which can be defined on the image space **Im***U* of the operator U (i.e., **Im***U* is the actual observation space). In the sequel, the inversion scheme is analyzed for each case separately. The properties of the solution $\hat{f}$ for the unknown field are also discussed and explained in detail.

2.1 Inversion Scheme for Case 1

In ***case 1***, the pseudo-inverse operator $\tilde{U}^{-1}$ corresponds to the usual inverse U^{-1} which is well defined since U is bijective (surjective + injective). The numerical computation of $\tilde{U}^{-1}$ is achieved by simply inverting the matrix **G** in (2.6), from which we finally compute the (unique) values of the unknown coefficients a_n needed in (2.4), i.e.

$$\mathbf{a} = \mathbf{G}^{-1}\,\mathbf{b} \tag{2.9}$$

The system of the observational representers $\{g_n\}_{n\in\Im}$ will provide a basis for the solution space V in this case, and the actual linear approximation (2.4) of the unknown field should be considered as an expansion with respect to this (non-orthogonal in general) basis.

2.2 Inversion Scheme for Case 2

In ***case 2***, the pseudo-inverse operator $\tilde{U}^{-1}$ will not correspond to U^{-1} in the usual sense, because U^{-1} does not exist in this case (U is not surjective). However, since U is always an injective operator, we can uniquely compute one of its infinite left-inverses which will also be identical with its unique pseudo-inverse (see, e.g., Naylor and Sell, 1982; Mallat, 1998). The pseudo-inverse operator of U will have thus the form

$$\tilde{U}^{-1} = (U^{*}U)^{-1}\,U^{*} \tag{2.10}$$

where U^{*} denotes the adjoint of U. An elegant proof for the invertibility of $U^{*}U$ can be found in Mallat (1998, p.130). Since the symmetric matrix **G** is non-invertible in this case, the numerical computation of $\tilde{U}^{-1}$ requires basically a minimum-norm solution for the singular

linear system (2.6). Such a solution is obtained by computing the unique pseudo-inverse (Moore-Penrose inverse) $\mathbf{G}^+$ of the $N\times N$ matrix **G**, i.e.

$$\hat{\mathbf{a}} = \mathbf{G}^+ \mathbf{b} \tag{2.11}$$

The general form of the pseudo-inverse matrix $\mathbf{G}^+$ is given by the following formula (see, e.g., Rao and Mitra, 1971):

$$\mathbf{G}^+ = (\mathbf{G} + \mathbf{E}^{\mathrm{T}}\mathbf{E})^{-1} - \mathbf{E}^{\mathrm{T}}(\mathbf{E}\mathbf{E}^{\mathrm{T}})^{-2}\mathbf{E} \tag{2.12}$$

where **E** is an $N\times N$ matrix, whose rows are some linearly independent solutions of the singular homogeneous system $\mathbf{G}\mathbf{x} = \mathbf{0}$. The matrix **E** is not uniquely defined and it always satisfies the condition $\mathbf{G}^{\mathrm{T}}\mathbf{E} = \mathbf{0}$. A particularly interesting and useful formula, which has extensively been used by various routines for the numerical computation of $\mathbf{G}^+$, is the following (Albert, 1972):

$$\mathbf{G}^+ = \lim_{\lambda\to 0}(\mathbf{G}^{\mathrm{T}}\mathbf{G} + \lambda^2\mathbf{I})^{-1}\mathbf{G}^{\mathrm{T}} = \lim_{\lambda\to 0}\mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{G}^{\mathrm{T}} + \lambda^2\mathbf{I})^{-1} \tag{2.13}$$

The observational representers $\{g_n\}_{n\in\Im}$ in ***case 2*** provide a <u>redundant</u> system of base functions for the solution space V, which is not a basis. The desired solution $\hat{f} = f_{\mathrm{V}}$ can thus be expressed in an infinite number of ways with respect to the family $\{g_n\}$. The use of the pseudo-inverse operator $\tilde{U}^{-1}$ gives, through eq.(2.11), just a single set $\hat{\mathbf{a}}$ among all admissible coefficients **a** satisfying the system (2.6), which has the minimum-norm property, i.e.

$$\|\hat{\mathbf{a}}\|^2 = \hat{\mathbf{a}}^{\mathrm{T}}\hat{\mathbf{a}} \le \|\mathbf{a}\|^2 = \mathbf{a}^{\mathrm{T}}\mathbf{a} \tag{2.14}$$

Finally, it can be easily shown that when the pseudo-inverse operator $\tilde{U}^{-1}$ is applied to elements of $l^2(\Im)$ which do not belong in **Im***U*, then the result will be the zero element of the Hilbert space V, i.e.

$$\forall\, \mathbf{b} \in (\mathbf{Im}U)^{\perp}, \quad \mathbf{Im}U \oplus (\mathbf{Im}U)^{\perp} = l^2(\Im) \qquad \text{then} \qquad \tilde{U}^{-1}\mathbf{b} = 0 \tag{2.15}$$

For more details, see Mallat (1998, p.131).

2.3 Comments

Summarizing the previous presentation for the linear approximation problem in an infinite-dimensional Hilbert space H, using a finite amount of observations of the form (2.1), we have shown that its solution $\hat{f}$ should always be expressed in the form of eq.(2.4), since the available data do not supply any information regarding the part $f_{\mathrm{V}^{\perp}}$ of the unknown field

which is orthogonal to the Hilbert subspace V spanned by the observational representers $\{g_n\}_{n\in\Im}$.

More importantly, the general solution (2.4) obeys the *minimum norm principle* (deterministic collocation) in the sense that

$$\left\|\hat{f}\right\|^2 \leq \|f\|^2 \tag{2.16}$$

for every other function $f \in \mathrm{H}$ that satisfies the given observation equations. In the last equation the symbol $\|\ \|$ denotes the norm in the Hilbert space H. This property is trivial to prove by simply taking into account the decomposition formula (2.2) and the generalized Pythagorean theorem in Hilbert spaces. In fact, if we had started the development of our solution procedure by imposing a-priori the following optimality criterion for the estimated field $\hat{f}$:

$$\left\|\hat{f}\right\|^2 = \min \tag{2.17}$$

then we would have derived the exact same system of equations as in (2.6). The unique solution $\hat{f} \in \mathrm{H}$ that satisfies the minimum norm principle (2.17) and the observation equations (2.1) will always belong to the Hilbert subspace $\mathrm{V} \subset \mathrm{H}$, and it will thus have the general expression of eq.(2.4); see, e.g., Tscherning (1986). However, when the observational representers are linearly dependent (***case 2***) the expansion coefficients a_n of this expression are not uniquely defined and an additional minimum-norm solution for them should be computed according to (2.11). It should be emphasized that the *uniqueness of the linear approximation* $\hat{f} = P_{\mathrm{V}} f$ is always assured, regardless of the linear dependence/independence of the finite set of the observational representers $\{g_n\}$.

It is quite interesting to realize that the minimum-norm principle (2.17), which has extensively been used in the geodetic practice up to date, is more of a *by-product* of the original linear Hilbertian setting for the approximation problem, rather than an 'arbitrary' estimation criterion of maximum smoothness for the unknown field in a given topology. Also, in the classic geodetic literature the problem of the functional approximation of gravity field signals in Hilbert spaces is exclusively treated for the ***case 1*** only; see, e.g., Tscherning (1986), Moritz (1980), Meissl (1976). The ***case 2***, where the observational representers $\{g_n\}$ are linearly dependent, can also be easily included in the Hilbertian approximation framework by simply computing the pseudo-inverse $\mathbf{G}^{+}$ of the singular matrix **G**, as it was explained in the previous paragraphs. Numerical stability problems that may occur during the computation of the matrices $\mathbf{G}^{-1}$ or $\mathbf{G}^{+}$, and other related issues are discussed in the next section.

The previous presentation of the linear approximation problem in Hilbert spaces reveals in an interesting way one aspect of the '*modelling choice problem*', which emerges when we use the deterministic collocation concept in gravity field approximation. The starting point in such cases is just a given *N*-tuple of real numbers $\{b_n\}_{n\in\Im}$, corresponding to the discrete

observations obtained from the unknown field f. As a second step, a certain modelling choice for f will enable us to construct observation equations of the form (2.1), which are then solved according to the previous methodology. In order for this methodology to work, however, we have to ensure that the given *N*-tuple of the observations belongs to the image space $\mathbf{Im}U \subseteq l^2(\Im)$ of the operator U, which in turn depends solely on our modelling choice (i.e., the form of the inner product < , > and the observational representers g_n). Assuming that the observations b_n correspond to point values of the unknown field itself, we can then express the observational representers in the following general form:

$$g_n(\mathrm{P}) = K(\mathrm{P}, \mathrm{Q}_n) \tag{2.18}$$

where $K(\mathrm{P}, \mathrm{Q})$ is the reproducing kernel of the Hilbert space H chosen to model the unknown field, and $\{\mathrm{Q}_n\}_{n\in\Im}$ is the given data point configuration. In this way, our modelling is exclusively based on the selection of a single positive-definite function $K(\mathrm{P}, \mathrm{Q})$. If the set $\{K(\mathrm{P}, \mathrm{Q}_n)\}_{n\in\Im}$ is linearly independent, then our given observation sequence will certainly belong in the space $\mathbf{Im}U \equiv l^2(\Im)$. If we repeat, however, the same modelling choice using denser data point configurations $\{\mathrm{Q}_n\}_{n\in\Im'}$, we should expect that the observational representers $\{K(\mathrm{P}, \mathrm{Q}_n)\}_{n\in\Im'}$ start to become linearly dependent. In order for the denser observations to remain in $\mathbf{Im}U \subset l^2(\Im')$, the unknown field must exhibit a certain 'correlated' behavior induced by our modelling choice $K(\mathrm{P}, \mathrm{Q})$. In simple words, the unknown field f *has to* belong in the Hilbert space with the specific reproducing kernel $K(\mathrm{P}, \mathrm{Q})$. For a fixed model, the level of the imposed correlation will increase as the data density (resolution) increases; i.e., the $\mathbf{Im}U$ space will 'shrink'. In this way, an essential factor in the modelling procedure becomes the maximum spatial resolution of data points $\{\mathrm{Q}_n\}$, above which the behavior of the unknown field starts to exhibit an increasing correlation in its values. The answer to this question (if any) will still leave, though, a relative freedom in choosing a specific model $K(\mathrm{P}, \mathrm{Q})$ for 'measuring' this correlation, whose unique selection requires the incorporation of additional criteria.

3. Numerical Stability and the Role of Frames

An issue of special importance in the framework of linear approximation in Hilbert spaces is that of *stable signal representations*. The importance of numerical stability is not restricted only on practical issues (e.g., computation of the matrices $\mathbf{G}^{-1}$ or $\mathbf{G}^{+}$), but it affects theoretical aspects of the approximation problem as well (e.g., convergence conditions). In our present Hilbertian approximation setting, the notion of stability is exclusively related to the continuity of the pseudo-inverse linear operator $\tilde{U}^{-1}$, which is applied to the original data $\mathbf{b} \in l^2(\Im)$ for the (partial) recovery of the unknown field f, i.e.

$$\hat{f} \;=\; f_{\mathrm{V}} \;=\; \tilde{U}^{-1}\mathbf{b} \tag{3.1}$$

Note that there is no need for the observational representers to be linearly independent in order to have a stable approximation scheme, since the continuity of the operator $\tilde{U}^{-1}$ does not depend on the linear dependence/independence of the family $\{g_n\}_{n\in\Im}$, as it will be explained in the sequel.

3.1 Frames

The most appropriate mathematical tool for studying the behavior of the linear operator $\tilde{U}^{-1}$ is the concept of a *frame*. In simple words, a frame in a Hilbert space V is a family of elements $\phi_j \in \mathrm{V}$, such that

(i) every $f \in \mathrm{V}$ is uniquely determined by its projections $< f, \phi_j >$

(ii) the reconstruction of f from the values $< f, \phi_j >$ is a stable algorithmic process

The rigorous mathematical definition of a frame $\{\phi_j\}$ is based on the following double inequality:

$$A\|f\|^2 \le \sum_j \left| < f, \phi_j > \right|^2 \le B\|f\|^2 \qquad (3.2)$$

which should be satisfied for every $f \in \mathrm{V}$. Note that (3.2) provides a simple generalization of the well known Plancherel-Parseval relation for orthonormal bases. The values *A* and *B* are some strictly positive constants, independent of f, and they are called *frame bounds*. Frame theory was originally developed by Duffin and Schaeffer (1952) in the context of nonharmonic Fourier series for the reconstruction of band-limited signals from their irregularly spaced samples. Some very good introductory references on this subject include Young (1980), Heil and Walnut (1994), and Strohmer (1995). For more advanced details on frame theory and related applications, Daubechies et al. (1986), and Daubechies (1990) should be consulted.

Associated with every frame $\{\phi_j\}$ in a Hilbert space V, there is a corresponding *frame operator* F which is defined in a similar way as the operator U in (2.8), i.e.

$$F : \mathrm{V} \to l^2(\mathrm{Z}) \qquad f \to < f, \phi_j > \qquad \forall\ j \in \mathrm{Z} \qquad (3.3)$$

where it is assumed for generality that the number of the frame components ϕ_j is infinite (i.e., V is an infinite-dimensional Hilbert space). The definition in (3.2) ensures that the linear frame operator F is always: (i) injective, and (ii) bounded (continuous). Furthermore, eq.(3.2) is a necessary and sufficient condition guaranteeing that F is invertible on its image with a bounded (continuous) inverse; for a proof, see, e.g., Mallat (1998, p.129). If the frame components $\{\phi_j\}$ are all linearly independent elements of V, then the frame operator F is in addition surjective and its image space **Im***F* is the whole Hilbert space $l^2(\mathrm{Z})$. In this special case the frame is called *Riesz (stable) basis*.

A frame thus always defines a unique, complete and stable discrete representation for continuous signals in a Hilbert space, which may also be redundant. When the frame components are normalized, $\|\phi_j\| = 1$, this redundancy is measured by the frame bounds *A* and *B*. If the normalized $\{\phi_j\}$ are linearly independent, then it can be proved that $A \leq 1 \leq B$. The frame is an orthonormal basis if and only if $A = B = 1$. If $A > 1$, then the frame is overcomplete (redundant) and the value of *A* can be interpreted as a minimum redundancy factor.

The reconstruction of an arbitrary $f \in \mathrm{V}$ from its frame coefficients $< f, \phi_j >$ is computed with the help of a *dual frame* $\{\tilde{\phi}_j\}$, according to the equation (see, e.g., Daubechies et al., 1986)

$$f = \sum_j < f, \phi_j > \tilde{\phi}_j = \sum_j < f, \tilde{\phi}_j > \phi_j \tag{3.4}$$

The family $\{\tilde{\phi}_j\}$ is also a frame for the Hilbert space V, having frame bounds 1/*A* and 1/*B*, and being uniquely defined by the operator formula

$$\tilde{\phi}_j = (F^* F)^{-1} \phi_j \tag{3.5}$$

where F^* is the adjoint of F. A proof for the invertibility of the operator $F^* F$ can be found in Mallat (1998). The reconstruction equation (3.4) can also be written in the operator form

$$f = (F^* F)^{-1} F^* Ff = \tilde{F}^{-1} Ff \tag{3.6}$$

It can easily be shown that the pseudo-inverse operator $\tilde{F}^{-1} = (F^* F)^{-1} F^*$ corresponds to the usual inverse of F, when the latter is restricted on its image space $\mathbf{Im}F \subseteq l^2(\mathrm{Z})$. Furthermore, $\tilde{F}^{-1}$ is always bounded (continuous) and it has the minimum sup norm among all possible generalized inverses of F (for more details and proofs see, e.g., Daubechies et al., 1986; Mallat, 1998). In simple words, therefore, the fundamental property of a redundant frame $\{\phi_j\}$ is to provide a unique and stable expansion for an arbitrary signal $f \in \mathrm{V}$

$$f = \sum_j a_j \phi_j \tag{3.7}$$

using, among all infinite possibilities for the sets of coefficients a_j, the one that has the smallest $l^2(\mathrm{Z})$ norm.

Note that, in general, the components $\{\phi_j\}$ will not be a basis in the technical sense, although their closed linear span is all of V. This is so because $\{\phi_j\}$ need not be independent. On the other hand, a basis for V that always assumes the expansion form (3.7) is not necessarily

stable, and therefore not a frame. A number of simple examples can be used to clarify the situation. Let $\{e_j\}_{j=1}^{\infty}$ be an *orthonormal* basis for an infinite-dimensional Hilbert space V. Then

1. $\{e_j\}_{j=1}^{\infty}$ is a frame for V, with bounds $A = B = 1$
2. $\{e_1, e_1, e_2, e_2, e_3, e_3, ...\}$ is a frame for V, with bounds $A = B = 2$
3. $\{2e_1, e_2, e_3, e_4, ...\}$ is a frame for V, with bounds $A = 1$ and $B = 4$
4. $\{e_1, \frac{e_2}{2}, \frac{e_3}{3}, \frac{e_4}{4}, ...\}$ is a complete orthogonal basis for V, but not a frame

In infinite-dimensional Hilbert spaces, therefore, a family of vectors may be complete and not yield a stable signal representation. For more examples and explanations, see Strohmer (1995). Also, one can easily verify that a finite set of N vectors $\{\varphi_j\}_{1 \le j \le N}$ is <u>always</u> a frame for the Hilbert space generated by their closed linear span, regardless of their linear independence or not (Mallat, 1998). This last fact will be used later on in section 3.3.

3.2 Practical Computations Using Frames

In order to recover a signal f from its frame coefficients $< f, \phi_j >$ we should first analytically pre-compute the dual frame components $\tilde{\phi}_j$ according to (3.5), and then we can reconstruct f using the sum

$$f = \sum_j < f, \phi_j > \tilde{\phi}_j \tag{3.8}$$

Fortunately, the above computational scheme is greatly simplified when the frame bounds A and B are equal (*tight frames*), in which case we have the simple relation (Daubechies et al., 1986)

$$\tilde{\phi}_j = \frac{1}{A} \phi_j \tag{3.9}$$

Within the context of computationally efficient tight frames, the major mathematical developments that took place during the last decade have resulted in two main frame families for various types of Hilbert spaces (including L^2 and Sobolev spaces). These two families are widely known as: (i) Weyl-Heisenberg or Gabor frames, and (ii) wavelet or affine frames. The first type basically corresponds to a discretized, windowed Fourier transform for the signal f, with the individual frame components generated by a simple translation and modulation of a basic window function w. For more details, see Daubechies et al. (1986), and Benedetto and Walnut (1994). In the one-dimensional case, the Gabor frames have the general form

$$\phi_{n,k} = w(x - nu_o) e^{ik\xi_o x} \tag{3.10}$$

where now the frame components depend on two discrete indices (n,k). The basic sampling intervals u_o and ξ_o are adjusted to the space-frequency spread of $w(x)$.

The second family of tight frames (wavelet frames) was essentially developed over the last few years in order to overcome certain limitations for the resolution properties of the space-frequency Gabor spectrum $\left| < f, \phi_{n,k} > \right|^2$. In this case, instead of a modulation, the basic window function $w(x)$ is dilated and translated to generate the various frame components, i.e.

$$\phi_{n,k} = a^{-k/2} \; w\left(\frac{x - n u_o a^k}{a^k}\right) \tag{3.11}$$

where again the parameters u_o and a depend on the space-frequency localization of the selected window $w(x)$. More details on wavelet frames can be found in Daubechies et al. (1986), Daubechies (1990, 1992), and Benedetto (1994). Note that the generating window function $w(x)$, as well as the discretization parameters (u_o, ξ_o, a), should satisfy certain conditions in order for (3.2) to be valid with $A = B$, and they cannot be arbitrarily selected for either the Gabor or the wavelet frame case. In Daubechies et al. (1986) and Daubechies (1990, 1992), sufficient conditions are given in order to build window functions $w(x)$ that generate either a Gabor or a wavelet tight frame in L^2 Hilbert spaces. Furthermore, these conditions allow us to construct generating functions $w(x)$ which are: (i) smooth (differentiable), and (ii) well localized in both space and frequency domain sense. The first property is important when we try to study the Gabor or wavelet spectrum $\left| < f, \phi_{n,k} > \right|^2$ for signal analysis purposes, since the use of a discontinuous $w(x)$ may introduce artificial high-frequency content in the spectrum which can 'distort' the image of the analyzed signal. As for the second property, the localization properties of the frame components $\{\phi_{n,k}\}$ allow us to get a closer look at the local details of the signal, avoiding a simple and vague 'global averaging' that it is usually obtained when we use non-localized reconstruction components (e.g., infinite sinusoids in classic Fourier signal analysis).

It should be mentioned that Gabor and wavelet frames $\{\phi_{n,k}\}$ have also been constructed for the general, 'non-tight' case. In the Gabor case, we are furthermore faced with the pleasant result that the dual frame $\{\tilde{\phi}_{n,k}\}$ will too be comprised of translations and modulations of a 'dual window' function $\tilde{w}$ (Mallat, 1998; p.143). However, the same does not apply in general for the wavelet frame case; for more details, see Daubechies (1990, 1992).

In certain applications, the frame vectors ϕ_j may depend on the specific analyzed signal f, and they cannot be arbitrarily chosen from a pre-determined Gabor or wavelet frame 'dictionary'. A classic example is the frames associated with irregular sampling problems (Duffin and Schaeffer, 1952; Feichtinger and Grochenig, 1994), where every frame component depends on the specific location of each sample value $< f, \phi_j >$. If the sampling grid varies from signal to signal this modifies the frame vectors, and it is then highly inefficient to compute the dual frame for each new signal. A detailed overview of various

numerically efficient, iterative algorithms that can be applied in such cases is given in Mallat (1998, p.135).

One of the main practical advantages of frame decompositions is that their redundancy becomes very useful in reducing possible additive noise in the frame coefficients $< f, \phi_j >$, and it significantly increases the robustness of the signal reconstruction, as compared to 'minimal basis' reconstruction schemes. These issues have been studied in detail mainly in the signal and image processing community, where frames are often used for high precision analog-to-digital conversion based on oversampling. For more details, see, e.g., Benedetto (1998), and Cvetkovic and Vetterli (1998).

3.3 Frames and Linear Approximation in Hilbert Spaces

After the preceding brief overview on the theory of frames, let us now try to reveal their relevance with respect to the original approximation problem of section 2. Since a finite set of *N* vectors is always a frame for the space generated by their linear combinations, the set of the observational representers $\{g_n\}_{n\in\Im}$ will constitute also a frame for the Hilbert subspace $\mathrm{V} \subset \mathrm{H}$ (solution space). The general solution for the unknown field, according to eq.(2.4), will thus be just an expansion with respect to the 'observational frame' $\{g_n\}$. Furthermore, the unknown coefficients $\{a_n\}$ of this expansion, obtained from the solution of the linear system (2.6) by either (2.9) or (2.11), will correspond to the optimal/stable ones implied by frame theory. In order to better see this, let us denote by **g** the *N*×1 vector of the observational representers, i.e.

$$\mathbf{g} = \left[g_1 \ldots g_j \ldots g_N \right]^{\mathrm{T}} \tag{3.12}$$

Each component g_j of the observational frame $\{g_n\}$ can be expressed with respect to the unique dual frame $\{\tilde{g}_n\}$ according to the equation (see section 3.1)

$$g_j = \sum_{n\in\Im} < g_j, g_n > \tilde{g}_n \qquad \forall\, j \in \Im \tag{3.13}$$

or, by using matrix notation

$$\mathbf{g}^{\mathrm{T}} = \tilde{\mathbf{g}}^{\mathrm{T}} \mathbf{G} \tag{3.14}$$

where

$$\tilde{\mathbf{g}} = \left[\tilde{g}_1 \ldots \tilde{g}_j \ldots \tilde{g}_N \right]^{\mathrm{T}} \tag{3.15}$$

and **G** is the symmetric matrix given in eq.(2.6). Also, the general solution in eq.(2.4) can be written in the equivalent vector form

$$\hat{f} = f_{\mathrm{V}} = \mathbf{g}^{\mathrm{T}} \mathbf{a} \tag{3.16}$$

which, using eq.(3.14), takes the form

$$\hat{f} \;=\; f_{\mathrm{V}} \;=\; \tilde{\mathbf{g}}^{\mathrm{T}}\mathbf{G}\mathbf{a} \tag{3.17}$$

Now, if we substitute the product **Ga** according to eq.(2.6), we finally get

$$\begin{aligned}\hat{f} \;=\; f_{\mathrm{V}} \;=\; \tilde{\mathbf{g}}^{\mathrm{T}}\mathbf{b} \;&=\; \sum_{n\in\Im} < f, g_n > \tilde{g}_n \\ &=\; \sum_{n\in\Im} < f_{\mathrm{V}}, g_n > \tilde{g}_n \\ &=\; \sum_{n\in\Im} < f_{\mathrm{V}}, \tilde{g}_n > g_n\end{aligned} \tag{3.18}$$

Thus, the linear approximation $\hat{f}$ of the unknown field is reduced to a simple *frame expansion* for its recoverable part f_{V} in the finite-dimensional Hilbert subspace $\mathrm{V} \subset \mathrm{H}$. The linear operator U, see eq.(2.8), will be a *frame operator* for the subspace V and its pseudo-inverse $\tilde{U}^{-1}$, see eqs.(2.10) and (3.1), is thus always continuous (stable). There are two different but equivalent ways to view the numerical computation of the approximation $\hat{f}$, which both require the calculation of the $N\times N$ pseudo-inverse matrix $\mathbf{G}^{+}$. In particular, we can either compute the projections (frame coefficients) of f with respect to the dual frame $\{\tilde{g}_n\}$ of the observational representers $\{g_n\}$, according to the scheme

$$\mathbf{G}^{+}\mathbf{b} \quad\rightarrow\quad < f, \tilde{g}_j > \;=\; < f_{\mathrm{V}}, \tilde{g}_j > \tag{3.19a}$$

or we can compute the dual frame components $\tilde{g}_j$ using the formula

$$\mathbf{g}^{\mathrm{T}}\mathbf{G}^{+} \quad\rightarrow\quad \tilde{g}_j \tag{3.19b}$$

In either case, the final solution will always be given by the general vector form

$$\hat{f} \;=\; \mathbf{g}^{\mathrm{T}}\mathbf{G}^{+}\mathbf{b} \tag{3.19c}$$

Of course, when $\{g_n\}$ are linearly independent then the matrix **G** will be invertible, and the last three equations are simplified accordingly by using $\mathbf{G}^{+} = \mathbf{G}^{-1}$. Furthermore, in this case the sets $\{g_n\}$ and $\{\tilde{g}_n\}$ will constitute a pair of biorthonormal Riesz bases for the solution space V (see, e.g., Young, 1980).

The use of frame theory shows us that the stability of the linear approximation problem in an arbitrary Hilbert space H is always guaranteed, as long as we deal with a finite amount of data

$< f, g_n >$. Although this result may seem adequate for real-life applications, it actually opens new interesting problems and questions related to the convergence of the approximation $\hat{f}$ towards the unknown field f, as well as to the original setup of the approximation problem (e.g., selection of the space H and its associated inner product, type of the observational representers g_n, invariance properties of the solution, etc.). In particular, the convergence problem should not be viewed just as a single 'advanced theoretical detail', since the highly increasing flow of various observations for the Earth's gravity field that takes place today, makes convergence considerations especially important in both theoretical and practical/computational sense.

Regardless of the choice of the Hilbert space H in which the linear approximation takes place, a necessary condition for convergence is that the closed linear span V of the observational representers $\{g_n\}$ tends to H, as the amount of the discrete data b_n increases. Even if the last condition is satisfied, the structure of H and/or the structure of the 'observational modelling' (i.e., form of the various g_n's) may be such that the observational frame $\{g_n\}$ becomes less and less stable as the data density increases. Although the system $\{g_n\}$ would be constantly a frame for every finite amount of observations, its bound values *A* and *B* may converge rapidly towards 0 and ∞ respectively, for high data density. This can cause serious numerical problems in the computation of the matrices $\mathbf{G}^{+}$ or $\mathbf{G}^{-1}$, making the whole approximation procedure highly ill-conditioned. More importantly, in the limit where we consider infinitely dense data, the frame bounds can either reach the values $A = 0$ and/or $B = \infty$, which makes the recovery of f from the infinite observations $< f, g_n >$ a completely unstable process. This does not mean, however, that f will not be uniquely defined from its 'very dense values' $< f, g_n >$. It just means that the approximation algorithm employing the available discrete data (see, eq.(3.19c)) cannot be reduced to a well-defined signal description when the data become infinitely dense, and convergence therefore cannot be established. An illuminating example is given below.

3.4 An Example

A very classic example of the above situation can be found in Mallat (1998, p.144). One other, more technical example arises when we consider the linear approximation of a *bandlimited* signal from its regularly gridded samples. The general Hilbert space H is now a specific *Paley-Wiener* (PW) *space* of bandlimited functions. Without loss of generality, let us assume that this PW space corresponds to functions bandlimited inside the interval $[-\pi,\pi]$. In this case, the sampled values $f(n\Delta x)$ of the unknown field $f(x) \in$ PW take the inner product form

$$b_n = f(n\Delta x) = < f(x), \mathrm{sinc}(x - n\Delta x) > \qquad (3.20)$$

where Δx denotes the available grid spacing, and the inner product $< , >$ corresponds to the standard L^2 definition. For simplicity, we will assume that the sampling interval has the dyadic form $\Delta x = 2^k$, where *k* is an arbitrary integer, which nevertheless does not destroy the generality of the following results.

When $k > 0$, the system of the observational representers, $s = \{\text{sinc}(x - n2^k)\}_{n \in \mathrm{Z}}$, is always a tight frame for the linear subspace V which is spanned by the family *s*. Note that the set *s* cannot cover the whole original Hilbert space PW, since that would require $k \leq 0$. Its frame bounds are $A = B = 1$, and thus *s* provides a complete orthonormal set for the subspace V. In such case, no stability problems will arise under the general approximation procedure of section 2.

For $k = 0$, the observational representers $\{\text{sinc}(x - n)\}_{n \in \mathrm{Z}}$ will again provide a tight frame with $A = B = 1$, but now they can span the whole Hilbert space PW. The approximation procedure of section 2 would be reduced to a simple, errorless reconstruction formula for the unknown field (Shannon's sampling theorem). However, when the integer sampling parameter *k* starts to decreases ($k < 0$), then the system $\{\text{sinc}(x - n2^k)\}_{n \in \mathrm{Z}}$ becomes an overcomplete tight frame for the whole PW space, with frame bounds $A = B = 2^{-k}$. Decreasing the value of *k* it basically means that we obtain denser and denser samples each time. In order to verify the value 2^{-k} for the two equal frame bounds, we shall use the fundamental relation that defines every frame in a Hilbert space, see eq.(3.2). For every function $f(x) \in \text{PW}$, we will have

$$\sum_n \left| < f(x), \text{sinc}(x - n2^k) > \right|^2 = \sum_n \left| f(n2^k) \right|^2 \tag{3.21a}$$

Note that we now consider only negative values for the integer parameter *k*. Let us select, for example, $k = -2$. In this way, eq.(3.21a) yields

$$\begin{aligned} \sum_n \left| f(\tfrac{n}{4}) \right|^2 = & \left(\ldots + f^2(-2) + f^2(-1) + f^2(0) + f^2(1) + f^2(2) + \ldots \right) + \\ & \left(\ldots + f^2(\tfrac{1}{4} - 2) + f^2(\tfrac{1}{4} - 1) + f^2(\tfrac{1}{4}) + f^2(\tfrac{1}{4} + 1) + f^2(\tfrac{1}{4} + 2) + \ldots \right) + \\ & \left(\ldots + f^2(\tfrac{2}{4} - 2) + f^2(\tfrac{2}{4} - 1) + f^2(\tfrac{2}{4}) + f^2(\tfrac{2}{4} + 1) + f^2(\tfrac{2}{4} + 2) + \ldots \right) + \\ & \left(\ldots + f^2(\tfrac{3}{4} - 2) + f^2(\tfrac{3}{4} - 1) + f^2(\tfrac{3}{4}) + f^2(\tfrac{3}{4} + 1) + f^2(\tfrac{3}{4} + 2) + \ldots \right) \\ & = \|f\|^2 + \|f\|^2 + \|f\|^2 + \|f\|^2 = 2^2 \|f\|^2 \end{aligned} \tag{3.21b}$$

which is easily verified from the translation-invariant properties of the L^2 norm and the orthonormality of the set $\{\text{sinc}(x - n)\}_{n \in \mathrm{Z}}$. We can easily generalize the final result of eq.(3.21b) for any other negative integer value of *k*, i.e.

$$\forall\, f(x) \in \text{PW}, \quad k < 0, \qquad \sum_n \left| < f(x), \text{sinc}(x - n2^k) > \right|^2 = 2^{-k} \|f\|^2 = A \|f\|^2 \tag{3.22}$$

In the case of infinitely dense data ($k \to -\infty$) the frame bounds $A = B \to +\infty$, and the reconstruction algorithm will completely fail to give a well-defined description of the unknown field. Following the linear approximation methodology of section 2, there is thus no way to establish convergence (in the sense of infinitely dense samples) in a Paley-Wiener space of bandlimited functions.

One could naturally claim that there is no need to consider infinitely dense samples in this specific example, since bandlimited signals can always be reconstructed without any error using a finite sampling interval that is less than, or equal to, the Nyquist limit (in the above example the Nyquist limit is $\Delta x = 1$). In any case, the pedagogical value of the above example always remains and it lies on the fact that a well-conditioned and convergent linear approximation scheme, that would use any possible data point configuration, is not always compatible with the original setting of our approximation problem (selection of H, type of the observational representers).

3.5 A Note on Ill-Posed Problems

From the discussion given in the previous sections it seems that an arbitrary selection of the Hilbert space H, in which we choose to put our unknown signal, is not enough to ensure a well-conditioned linear approximation scheme for any data configuration, and most importantly it does not provide convergence in the case of infinitely dense data. Although the selected space H may be perfectly suitable to describe some physical system, its 'structure' could pose certain limitations regarding the admissible data configurations that can be used for a reasonably stable approximation procedure. Generalizing the notion of an ill-posed problem according to Hadamard (1923), we can say that some additional type of ill-posedness may exist in the original approximation problem of section 2, which is now related to the stability condition of the solution *for different data configurations*, as well as to its convergence properties for infinitely dense data.

The general solution procedure that was followed in section 2 corresponds basically to a simple Tikhonov-type regularization, which is needed to overcome the non-uniqueness aspect of the original underdetermined approximation problem of eq.(2.1). The formalism of frame theory ensured also that this regularization scheme will always lead to stable solutions for a given configuration of finite data. But when it comes to considering different possible data configurations with increasing density and related convergence issues, an additional regularization of the initial approximation problem may be feasible. Such specific regularization procedures will be discussed later on in this report. For some general alternatives, see Nashed (1976).

4. The Hilbert Space Choice Problem

In this section we will study more closely the general *modelling problem* for linear approximation in Hilbert spaces, which was briefly discussed in the previous section. Note that the modelling issues now do not include the problems mentioned at the end of section 2.3, i.e., it is assumed a-priori that the unknown field does indeed belong to the selected Hilbert space. In particular, we just try to establish a set of conditions satisfied by the 'global' Hilbert space H enclosing all of our unknown signals, that are sufficient to ensure certain

desirable properties for the solution of the linear approximation problem in both theoretical and practical sense. Such properties are:

(i) convergence of the linear approximation $\hat{f}$ in the case of infinitely dense data;

(ii) stable algorithms for the numerical computation of $\hat{f}$; and

(iii) the level of stability should be independent of the available data configuration, i.e., the condition number of the pseudo-inverse operator $\tilde{U}^{-1}$ should not worsen as the data density increases.

One additional important property is the invariance of the linear approximation $\hat{f}$ with respect to arbitrary affine transformations. Essentially, this will guarantee the independence of the final estimated field from the reference system used to describe the position of the data points. For 1D approximation problems, these affine transformations will correspond to simple *translations* of the reference system with respect to the data point configuration, whereas for higher dimensions we should additionally include possible *rotations* of the reference system. In order to achieve such affine-invariance properties for the linear approximation problem, the norm $\| \ \|$ in the global Hilbert space H should satisfy a corresponding invariance condition*

$$\forall\ f \in \mathrm{H}, \quad \| f(\mathrm{Q}) \|^2 = \| f(afn\mathrm{Q}) \|^2 \tag{4.1}$$

where Q is an arbitrary point in the domain of f, and *afn* denotes a general affine transformation. In the one-dimensional case, the last equation will take the form

$$\forall\ f \in \mathrm{H}, \quad \| f(x) \|^2 = \| f(x-\tau) \|^2 \tag{4.2}$$

with τ being an arbitrary real number.

It would seem only natural to require that the linear approximation scheme should also satisfy some kind of *scale-invariance*. For this type of invariance, a reasonable condition on the norm of H could be, e.g., for the 1D case,

$$\forall\ f \in \mathrm{H}, \quad \| f(x) \|^2 = \frac{1}{|a|} \left\| f\left(\frac{x}{a}\right) \right\|^2 \tag{4.3}$$

where a is some non-zero scaling factor. In the following, therefore, we will assume that the norm $\| \ \|$ in the Hilbert space H satisfies both invariance conditions (4.2) and (4.3). The derivation of their multi-dimensional counterparts is straightforward and it will not be given here.

* Actually, such a norm condition is needed to ensure the invariance of the *approximation error* $\|e\| = \|f - \hat{f}\|$ under arbitrary affine transformations of the reference system.

4.1 Data Type and Configuration

In the two previous sections, 2 and 3, the available data of the unknown field were assumed to have the inner product form $b_n = < f, g_n >$, where g_n are known elements of the global Hilbert space H. This is the classic modelling approach for the observation equations when we consider linear approximation problems in Hilbert spaces (Dermanis and Sanso, 1997). The study of the structure of the space H, that will ensure the solution properties (i)-(iii), requires a more specific physical meaning for both b_n and g_n. In this way, expressions like 'infinitely dense data', 'increasing data density', or 'data point configuration', could be unambiguously formulated in proper mathematical terms.

The most straightforward case arises when the data b_n represent point values of the unknown field f itself. In such cases the observational representers $\{g_n\}$ should correspond to the actual *reproducing kernel (r.k)* that is associated with the global Hilbert space H (see, e.g., Tscherning, 1986; Moritz, 1980). A point value of the unknown field, at a point Q, can always be expressed in the linear form

$$f(\mathrm{Q}) = < f(\mathrm{P}), K(\mathrm{P}, \mathrm{Q}) >_{\mathrm{P}} \qquad (4.4)$$

where $K(\mathrm{P}, \mathrm{Q})$ is the reproducing kernel of the Hilbert space H, and the subscript P means that the inner product $< , >$ is calculated using the point P as the independent variable. According to the dimensionality of the problem, the points P and Q will depend on one, two, or more coordinates. In order to keep the notation simple, we will work in the sequel for the 1D case only. The extension into higher dimensions does not produce any conceptual complications and it will not restrict the generality of the subsequent results.

The consideration of different configurations for the point values $f(\mathrm{Q}_n)$ is also greatly simplified if we work with regularly gridded data, i.e., $f(\mathrm{Q}_n) = f(n\Delta x)$. In this way, cases of increasing data density, as well as limiting cases of infinitely dense data, can be simply considered by changing the value of the single sampling parameter Δx. Hence, in the following we will deal with discrete data of the unknown field $f \in \mathrm{H}$ that assume the form

$$b_n = f(n\Delta x) = < f(x), K(x, n\Delta x) > \qquad (4.5)$$

where the range (n) of the available values may be finite or infinite.

4.2 Conditions for Stable and Convergent Linear Approximation

Let us assume that we have a specific configuration of gridded point values $f(n\Delta x)$ for an unknown field f belonging in a general Hilbert space H with reproducing kernel $K(x, y)$. Following section 2, the solution space $\mathrm{V}_{\Delta x} \subset \mathrm{H}$ of the linear approximation problem will be the closed linear span of $K(x, n\Delta x)_{n \in \mathrm{Z}}$. Since the solution space is a closed linear subspace of the global RKHS H, it will also be a RKHS itself with its reproducing kernel being generally different from the global kernel $K(x, y)$; see, e.g., Aronszajn (1950). In order to

have a stable approximation scheme (*for the given data configuration* Δx), we know that the following inequality should be satisfied for every function $g(x) \in \mathrm{V}_{\Delta x}$ (see, section 3):

$$A\|g\|^2 \le \sum_{n\in \mathrm{Z}} |< g(x), K(x, n\Delta x) >|^2 \le B\|g\|^2 \qquad (4.6)$$

where $0 < A, B < +\infty$; i.e., the translates of the reproducing kernel $K(x, y)$ should constitute a frame for the space $\mathrm{V}_{\Delta x}$ generated by their linear combinations. We can further simplify (4.6) using the reproducing properties of $K(x, y)$. In this way, for stable approximation, all the functions $g(x)$ of the solution space $\mathrm{V}_{\Delta x}$ should satisfy the relation

$$A\|g\|^2 \le \sum_{n\in \mathrm{Z}} |g(n\Delta x)|^2 \le B\|g\|^2 \qquad \forall \;\; g(x) \in \mathrm{V}_{\Delta x} \subset \mathrm{H} \qquad (4.7)$$

The left part of the inequality (4.7) means that the functions of the solution space that take zero values at the data points, should all be identically equal to the zero function. As a result, $\mathrm{V}_{\Delta x}$ should not contain signals like the one shown in figure 4.1.

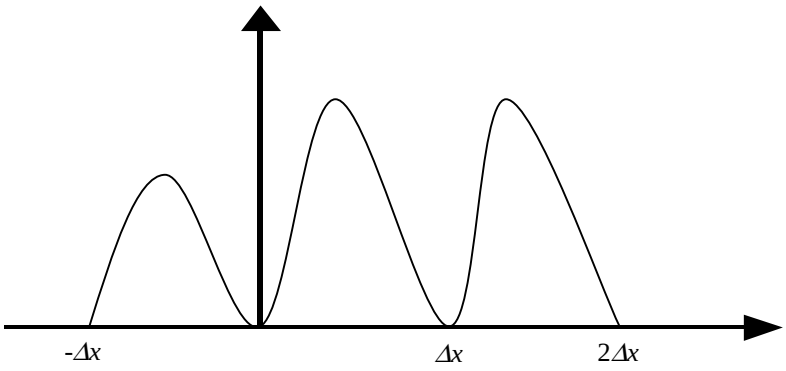

Figure 4.1: Non-acceptable function for the solution space $\mathrm{V}_{\Delta x}$ in the case of stable approximation

As the data configuration changes (varying sampling interval Δx), the solution space $\mathrm{V}_{\Delta x}$ changes accordingly. In order to ensure that the stability of the linear approximation not only remains, but also it does not become worse, the finite frame bounds A and B in (4.7) must be independent of the sampling interval Δx. Let us denote by $\mathrm{V}_{\Delta x'}$ the solution space corresponding to a new data configuration, using sampling interval $\Delta x' \neq \Delta x$. To keep the same level of numerical stability in the linear approximation, we should require that all functions $h(x)$ of the new solution space satisfy the inequality

$$A\|h\|^2 \le \sum_{n\in \mathrm{Z}} |h(n\Delta x')|^2 \le B\|h\|^2 \qquad \forall \;\; h(x) \in \mathrm{V}_{\Delta x'} \subset \mathrm{H} \qquad (4.8)$$

How is it possible to satisfy both (4.7) and (4.8) in the same global Hilbert space H, for every value of the new sampling interval $\Delta x'$? An easy and straightforward way is to require that

the two different solution spaces, $V_{\Delta x}$ and $V_{\Delta x'}$, are related through a simple *isometric scaling*, i.e.

$$\forall\, h(x) \in V_{\Delta x'},\ \exists\ g(x) \in V_{\Delta x} :\quad h(x) = \left(\frac{\Delta x}{\Delta x'}\right)^{1/2} g\left(\frac{\Delta x}{\Delta x'}\, x\right) \tag{4.9}$$

Indeed, if equations (4.7) and (4.9) hold true, then equation (4.8) will also be satisfied by taking into account the isometric scaling property of the norm in the Hilbert space H; see, eq.(4.3).

A non-varying stability level for the linear approximation problem would require, therefore, that all the different solution spaces $V_{\Delta x}$ (for different data configurations Δx) are just scaled versions of each other, according to the isometric transformation (4.9). Each of these solution subspaces is a reproducing kernel Hilbert space itself, with its reproducing kernel $K_{\Delta x}(x, y)$ depending of course on the sampling interval Δx. The above scaling condition between the various spaces can be equivalently expressed as a scaling condition between their corresponding reproducing kernels. In this way, the r.k of every solution space $V_{\Delta x}$ should have the following form for non-varying stability in the linear approximation:

$$K_{\Delta x}(x, y) = \frac{1}{\Delta x} K_o\left(\frac{x}{\Delta x}, \frac{y}{\Delta x}\right) \tag{4.10}$$

where $K_o(x, y)$ is the r.k corresponding to the solution space for a normalized data sampling interval $\Delta x = 1$. According to the scale-invariance property of the norm $\|\ \|$ in the Hilbert space H, imposed by eq.(4.3), all the different $K_{\Delta x}(x, y)$ will satisfy also the equation

$$\left\| K_{\Delta x}(x, y) \right\| = \text{constant} \tag{4.11}$$

where the constant at the right hand side is independent of the sampling interval Δx. The norm operator in (4.11) should be applied twice to the reproducing kernel, first considered as a function of x only, and then as a function of y.

The fulfillment of the third property for the linear approximation problem (i.e., convergence in the case of infinitely dense data) requires of course that

$$\lim_{\Delta x \to 0} V_{\Delta x} = H \tag{4.12}$$

which can also be written in terms of the reproducing kernels of the corresponding spaces as

$$\lim_{\Delta x \to 0} K_{\Delta x}(x, y) = K(x, y) \tag{4.13}$$

Using eq.(4.10), the final condition (4.13) takes the form

$$K(x, y) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} K_o\left(\frac{x}{\Delta x}, \frac{y}{\Delta x}\right) \tag{4.14}$$

We have finally derived a sufficient condition for the global Hilbert space H, in order for the linear approximation problem to satisfy the three fundamental properties stated in page 39. In doing so, we needed also to impose a certain scale-invariance property for the norm in H. Note that other types of scale-invariant norms, not necessarily the same as in (4.3), may be introduced and used for the modelling of the Hilbert space H. In such a case, we should modify accordingly the isometric scaling (4.9) for the solution spaces, as well as the condition (4.10) for their reproducing kernels. An analogous change will also occur in the final condition (4.14).

Equation (4.14) gives the general form of the r.k that an 'optimal' Hilbert space should have. The term 'optimal' is meant in the sense of the three basic properties (i)-(iii) (see, page 39), as well as with respect to a specific scale-invariance condition imposed for $\left\| \hat{f} \right\|$; see eq.(4.3). It is seen, however, that this optimal reproducing kernel $K(x, y)$ will not be a proper function in the usual sense. As a matter of fact, $K(x, y)$ will have all the characteristics of the classic Dirac's *delta distribution* $\delta(x-y)$. This type of 'function' is a rather complicated notion, defined as a limiting process on well-behaved kernels (e.g., Gaussian, Dirichlet, Fejer, etc.), and being particularly useful for its reproducing properties under the standard L^2 inner product, i.e.

$$\forall\ f(x) \in L^2(\Re) \qquad \int_{-\infty}^{+\infty} f(x)\,\delta(x-y)\,dx = f(y) \tag{4.15}$$

For more details, see, e.g., Gelfand and Shilov (1964), Debnath and Mikusinski (1999), and Walter (1992a). Our 'delta' kernel $K(x, y)$, defined by the combination of equations (4.10), (4.11), and (4.14), will even be more general than the classic $\delta(x-y)$, in the sense that its reproducing properties should hold under an arbitrary inner product definition*. Hence, the $L^2(\Re)$ space can be considered just as a single member of a more extended family of Hilbert spaces, whose r.k's can take the form of the limiting process described above. The norm in every space H of this family will satisfy the isometric condition (4.3), and each such H will make a possible candidate for modelling our unknown signals in order to achieve constantly stable and convergent linear approximations. More details and explanations will be given in section 5.

* The only restriction on the inner product $< f, g >$ in the Hilbert space H is that a scale-invariance condition of the form (4.3) has to be satisfied by the induced norm $\left\| f \right\|^2 = < f, f >$.

4.3 Remarks on Previous Methodologies

The issues related to the stability and the convergence of the linear approximation problem in Hilbert spaces have been extensively discussed in the geodetic literature. Their importance regarding the optimal estimation of the anomalous potential of the Earth from discrete measurements was first pointed out by Eeg and Krarup (1973). The following quotation is taken from the concluding remarks in their classic publication:

> *" No doubt the less satisfactory point in integrated geodesy is that of the choice of the norm; the result is only defined when the norm has been chosen; it depends on the norm; and the choice of the norm is to some extend arbitrary.*
> *… we could ask: under which conditions will the solution in integrated geodesy converge to the correct result independently of the choice of the norm when the observations are correct and their number increases without limit? This should give conditions about the class of norms from which the choice may be made and about the nature and distribution of the observations. In order to be meaningful, this question must be modified so as to demand not only convergence but also stability, i.e., the results must depend continuously on the observation data."*

Numerous theoretical studies exist for the convergence problem in collocation; see, e.g., Barzaghi and Sanso (1986), Moritz (1976), Sanso and Tscherning (1980), Tscherning (1978). A common characteristic in all these studies is that they rely upon rather strong assumptions, which restrict considerably the validity of the convergence results. Furthermore, these assumptions create new and even more difficult questions that need to be answered for a complete and rigorous treatment of the problem. In particular, the first standard assumption in these convergence studies is that the observational representers in the Hilbert space, within which the convergence problem is studied, are always linearly independent so that the matrix **G** in eq.(2.6) is always non-singular. Such an assumption is supposed to hold even for very dense data point configurations (i.e., ε-net configurations; see Moritz, 1976), which are considered sufficient to ensure the full recovery of the unknown field. In this spirit there is also the underlying assumption that the observational representers, corresponding to such sufficiently dense data point geometries, provide a complete system of independent base functions for the whole Hilbert space H within which we model our unknown signals (Tscherning, 1978).

The above, briefly described, general setting for the convergence issue in collocation implies basically that there should exist a close connection between the adopted reproducing kernel and the overall *resolution properties* of the signals that belong in the corresponding Hilbert space. In other words, for a given Hilbert space H with reproducing kernel $K(\mathrm{P},\mathrm{Q})$, we should be able to identify a point distribution $\{\mathrm{Q}_i\}$ such that the functions $K(\mathrm{P},\mathrm{Q}_i)$ constitute a complete system of independent base elements (i.e., basis) for H. Every other data point configuration $\{\mathrm{Q}_i'\}$ that violates the linear independence condition for $K(\mathrm{P},\mathrm{Q}_i')$ is considered unacceptable and it cannot be incorporated in the approximation framework. In this way, apart from the unpleasant result that we cannot establish some kind of convergence for infinitely dense data, we are additionally faced with a restriction problem for the admissible spatial configurations of data points. But is it reasonable to use a reproducing kernel based on pre-determined 'sampling scenarios' for the unknown field, excluding other possible data configuration schemes with varying spatial density that may arise in practice? Furthermore, even if the answer is yes, how are we going to a-priori select/determine the resolution properties of the unknown signals that should be implicitly associated with the form of the reproducing kernel (i.e., spatial configuration of the 'ideal', sufficiently dense point distribution $\{\mathrm{Q}_i\}$, actual form of the corresponding complete system $K(\mathrm{P},\mathrm{Q}_i)$)?

In essence, the methodology that has already been applied for establishing the convergence properties of collocation in a Hilbert space H, leads back to its original major problem: the reproducing kernel choice problem. The reproducing kernel in H is supposed to dictate a certain convergence scheme, in the sense that (Moritz, 1976)

$$\forall\, f(\mathrm{P}) \in \mathrm{H} \qquad f(\mathrm{P}) = \lim_{N\to\infty} \sum_{i=1}^{N} a_i\, K(\mathrm{P},\mathrm{Q}_i) \tag{4.16}$$

where the coefficients a_i are obtained from the solution of the non-singular system

$$\mathbf{G}\,\mathbf{a} = \mathbf{b}$$

$$\mathbf{a}[i] = a_i \qquad \mathbf{G}[i,j] = K(\mathrm{Q}_i,\mathrm{Q}_j) \qquad \mathbf{b}[i] = < f(\mathrm{P}), K(\mathrm{P},\mathrm{Q}_i) > = f(\mathrm{Q}_i) \tag{4.17}$$

Thus, the critical issue that will control the convergence (and the stability) of the linear approximation, becomes the spatial point distribution $\{\mathrm{Q}_i\}$ for which (4.16) and (4.17) are satisfied. It is a rather challenging and (to our knowledge) relatively unexplored mathematical problem to try finding, for a given reproducing kernel $K(\mathrm{P},\mathrm{Q})$, its associated point distribution $\{\mathrm{Q}_i\}$ that can provide completeness and linear independence in the corresponding Hilbert space H; for some special cases, see Nashed and Walter (1991). It is also worth mentioning that the convergence setting, expressed through the last two equations, corresponds clearly to the search of a *sampling theorem* in the underlying Hilbert space, where the sampling functions will be given by the general formula:

$$S(\mathrm{P},\mathrm{Q}_j) = \lim_{N\to\infty} \sum_{i=1}^{N} K(\mathrm{P},\mathrm{Q}_i)\,[\mathbf{G}(i,j)]^{-1} \tag{4.18}$$

It is intuitively obvious that the use of a smooth reproducing kernel, with a sparse $\{\mathrm{Q}_i\}$ point distribution associated with it, will limit our modelling capabilities for the gravity field to correspondingly smooth signals, which may not be compatible with dense data sets (see also the remarks in section 2.3, pp.21-22). On the other hand, if we want to increase the resolution content of the signals in the adopted Hilbert space H, we have to choose a rougher reproducing kernel $K(\mathrm{P},\mathrm{Q})$ with a denser 'optimal' $\{\mathrm{Q}_i\}$ point distribution. It is less obvious, however, that the use of 'higher-resolution' reproducing kernels in conjunction with sparse data point configurations (relatively to the spread of $K(\mathrm{P},\mathrm{Q})$ or the density of $\{\mathrm{Q}_i\}$), can result in minimum-norm interpolations that may be completely erroneous.

An example of such situation is illustrated in figure 4.2, where 13 point gravity anomaly values, with a normalized sampling interval $\Delta x = 1$, are used to determine linear approximations of the underlying gravity anomaly signal using two different reproducing kernels.

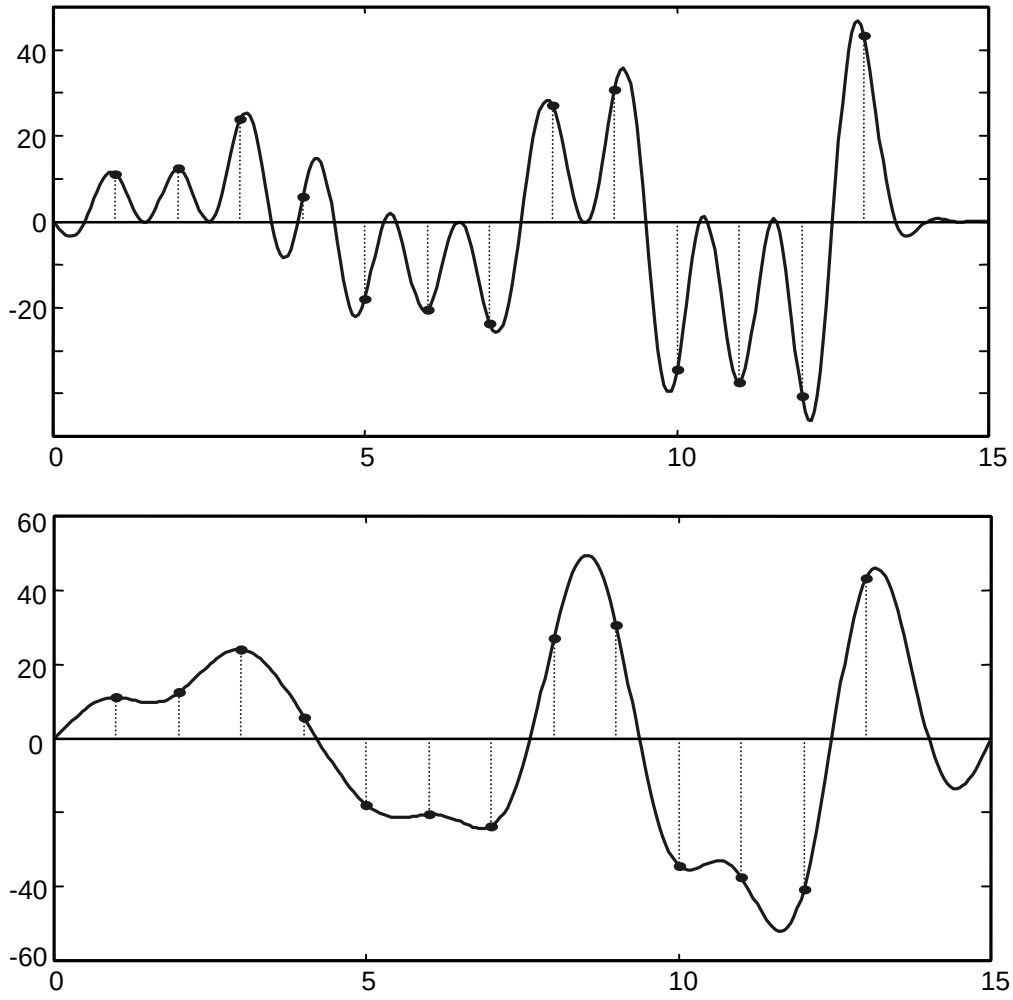

Figure 4.2: Minimum-norm collocation using a smooth r.k (lower graph) and a higher-resolution r.k (upper graph). The dots represent the data points.

In the one case (upper graph) the used reproducing kernel corresponds to the Hilbert space of bandlimited functions inside the interval $(-2\pi,2\pi)$, whereas in the second case (lower graph) the minimum-norm interpolation takes place in the Hilbert space of bandlimited functions inside the interval $(-\pi,\pi)$. Although the first RKHS is capable of providing signals with higher resolution than the second RKHS, its use with a sparse data point distribution will produce a minimum-norm approximation for the unknown field that contains artificial details which cannot be justified from the available data, as it is seen from the irregular signal variations occurring between the data points. On the other hand, the use of a smoother RKHS in the second case results to a more reasonable approximation scheme, which 'merges' the spatial resolution of the available data points with the resolution of the signals in the RKHS in a certain optimal fashion (i.e., in this specific example we have just an application of Shannon's sampling theorem).

In every convergence problem in approximation methods the central role is always played by the notion of a *correct* or *true* solution for the unknown field we try to approximate. Unlike the framework and the philosophy of pure mathematics, in physical applied sciences (like Physical Geodesy) the correct solution is just a model obtained by simplifying in some way

physical reality. In Sanso's (1987) words: '*this is what we can do at most*'. When we consider the deterministic collocation concept for our approximation purposes, the modelling takes the form of a specific Hilbert function space H with an associated reproducing kernel. As it was explained in the previous paragraphs, however, there will exist a certain trade-off among the overall signal details that this model can provide, the quality of the approximation for sparse data point configurations (relatively to the implied model resolution), and the existence of a solution for data sets denser than the implied model resolution (see also, Rummel et al., 1979). This would suggest that we should change our modelling choice for every new data point configuration in order to construct good and reasonable signal approximations that balance between the data and model resolution. But in this way the convergence problem becomes meaningless since the global RKHS H will be different for every data set!

5. The Connection with the Multiresolution Concepts

The final result of section 4.2 may seem a bit puzzling, and even contradictory with the original setting of the problems discussed therein. In essence, we have shown that in order to achieve

- a constantly stable linear approximation for increasing regular data $f(n\Delta x)$ density, and
- convergence for infinitely dense data,

then the Hilbert space, within which we model our unknown signals $f(x)$, should not possess a reproducing kernel! That is because $K(x, y)$, as determined by the limiting procedure in section 4.2, could not definitely belong in any reasonable function space of 'natural signals' capable of describing physical phenomena like Newtonian gravity fields, and therefore it is not a reproducing kernel in the strict sense. On the other hand, this peculiar behavior of the optimal $K(x, y)$ is a perfectly acceptable mathematical structure, that lies in the vast field of *generalized functions* and *distributions* (Schwartz, 1951; Sobolev, 1964). Just like we use the classic Dirac distribution $\delta(x-y)$ as a reproducing kernel for L^2 signals, in the same spirit we can also accept the kernel $K(x, y)$ given by (4.14) as a generalized r.k for a specific class of Hilbert spaces (including L^2) with scale-invariant norms according to eq.(4.3).

What is the important characteristic of all these Hilbert spaces admitting the generalized form (4.14) as their reproducing kernel[*]? Basically, a Hilbert space H with such a reproducing kernel should be interpreted as an *infinite-resolution* function space. Regardless of its topological structure (i.e., the specific type of the scale-invariant norm $\|\ \|$), the value of an arbitrary signal at a point Q cannot be generally determined by its behavior in the neighborhood of Q. That is because no well-behaved kernel $K(x, y) \in \mathrm{H}$ exists which, when applied to an arbitrary $f(x) \in \mathrm{H}$, could recover the exact same $f(x)$. However, this does not mean that H will not contain smooth functions, whose value at a specific point could be predicted arbitrarily well from set of values at a dense network of adjacent points.

[*] Note that the term 'reproducing kernel' is now used in a non-rigorous sense.

These infinite-resolution spaces provide an ideal 'environment' to model signals with highly irregular patterns, where abrupt changes may be expected between nearby points. In view of the rapid increase in gravity field data resolution that takes place today, the recovery of such irregularities in the gravity field signals not only does it seem possible from an observational point of view, but it is actually extremely useful in certain types of applications (e.g., geophysical inverse problems, geodynamical studies, etc.). Thus, it would seem only natural to use a Hilbert space setting that is actually capable of providing signals with such erratic behavior. Extreme cases of signals with singularities should not also be excluded from our approximation framework, since such situations may very well appear in the gravity field (see, e.g., Bochio, 1981, 1983), as well as in other signal processing applications (see, e.g., Mallat and Hwang, 1992).
A very important problem, however, still remains. Since the optimal (with respect to all these issues discussed so far) Hilbert space should not possess a proper reproducing kernel, how are we going to actually construct the symmetric matrix **G** and compute the linear approximation $\hat{f}$ according to eq.(3.19c)? In such cases the observational representers

$$g_n(x) = K(x, n\Delta x) \tag{5.1}$$

are not usual functions, and linear expressions of the form (3.19c) cannot be properly defined and used for the numerical computation of the approximated field. The same situation also appears when the available observations are not necessarily point values of the unknown field, but other linear functionals as well. The solution to this problem is discussed below.

5.1 Linear Approximation as a Multiresolution Approximation

It would be useful, at first, to understand in simple terms why the minimum-norm solution of the linear approximation problem would fail to give a reasonable answer, when the selected Hilbert space H has the infinite-resolution characteristics discussed above. In this case the *evaluation functional* in H is not bounded (continuous), which means that there is no $C > 0$ such that the following inequality:

$$\left| f(\mathrm{P}) \right| \le C \left\| f \right\|_{\mathrm{H}} \tag{5.2}$$

will hold for every $f \in \mathrm{H}$ and for every point P in the domain of f (see, e.g., Moritz, 1980). The invalidity of (5.2) ensures that there will exist functions in H with $\left\| f \right\|_{\mathrm{H}} = 0$ but there may be points in their domain for which $f(\mathrm{P}) \neq 0$. In other words, isolated singularities are acceptable in infinite-resolution Hilbert spaces, and they are topologically equal to the zero function. Therefore, the minimum-norm function that assumes given values at a discrete network of data points will be the zero function, and that makes minimum-norm solutions theoretically and computationally useless. As it was explained in the previous section, knowledge of f at a set of discrete points cannot generally provide any information regarding its behavior in the neighborhood of these points, and this fact is immediately reflected to the inadequacy of (3.19c) to yield a proper numerical result. A more rigorous mathematical treatment requires the incorporation of concepts from *measure theory* (see, e.g., Halmos, 1991) which will be avoided here.

The above situation should not be perceived as a 'deficiency' of the infinite-resolution Hilbert space H, but it is merely a modelling requirement. In order to compute a reasonable linear approximation for an unknown field from its discrete values, we have to develop a different methodology which will not employ the generalized reproducing kernel of H, and which will satisfy the crucial properties (i)-(iii) stated in page 39. Such a methodology will now be briefly described.

The final result of any linear approximation procedure in a Hilbert space H is always a bijective linear mapping T between an observation space Λ and a solution space $V \subset H$. The observation space contains different data sets (**b**) corresponding to different signals from the global space H. All these data sets are associated with a specific data point configuration, which in this report has been always assumed as a sequence of gridded sampled values with a certain resolution level Δx. In order to have a *stable* bijective mapping T, the following boundeness property for the norm in Λ should be satisfied:

$$A\|\mathbf{b}\|_{\Lambda} \;\le\; \left\|\hat{f}\right\|_{H} \;\le\; B\|\mathbf{b}\|_{\Lambda} \qquad \forall\, \hat{f} \in V\ ,\ \forall\, \mathbf{b} \in \Lambda \tag{5.3}$$

where **b** denotes the data set obtained from an unknown field $f \in H$, $\hat{f}$ denotes the linear approximation of f computed via the linear mapping T, and A, B are positive constants. The formula (5.3) is just an alternative definition for an *isomorphic* mapping between the solution space V and the observation space Λ (Holschneider, 1995; p.183).

For all practical purposes it is convenient to choose the norm in Λ as the standard l^2 norm for discrete sequences, and thus the observation space can be always viewed as a space of square-summable sequences. If we further accept the fact that the observations in the various data sets can produce *any* square-summable sequence (i.e., remember that the data sets are obtained from infinite-resolution signals in general, and no a-priori smoothing restrictions are imposed), then the observation space can be regarded as the whole $l^2(Z)$ Hilbert space. This of course does not exclude cases of finite sets of sampled data. The situation is illustrated in figure 5.1.

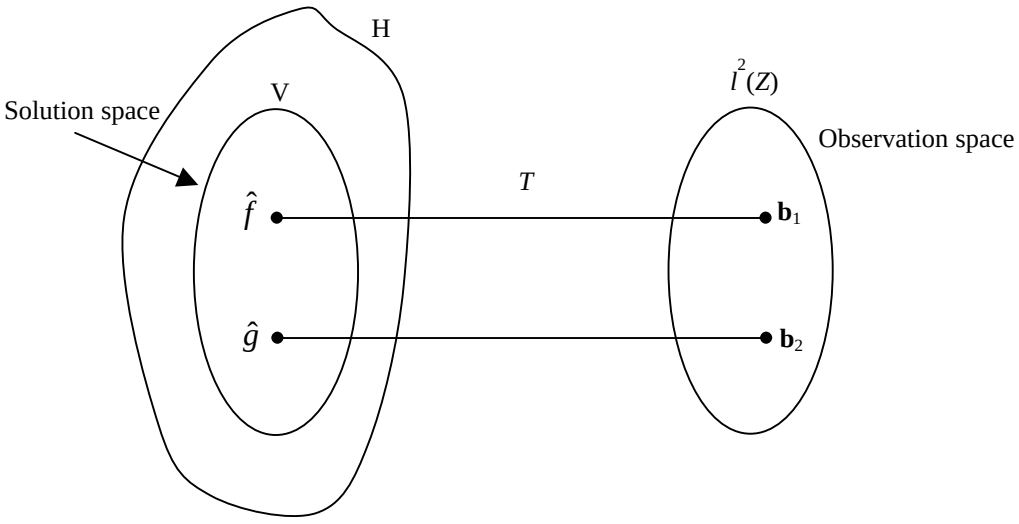

Figure 5.1: Linear Approximation as an isomorphic mapping T

So far we have identified a stable linear approximation procedure, in an infinite-resolution Hilbert space H, as an isomorphic linear mapping T between a solution space $\mathrm{V} \subset \mathrm{H}$ and the observation space $l^2(Z)$ which contains regularly sampled values of the unknown signals in H. If we also require that T is a translation-invariant approximation operator, then V has to be generated by a family of linearly independent functions of the form $\varphi_1(x-n\Delta x)_{n\in Z}$. The proof is very simple and it is omitted (see, e.g., Mallat, 1989; Walter, 1992b). The solution space V (i.e., its generating function $\varphi_1(x)$), as well as the properties of the isomorphism T, will depend on some optimal criteria that we have chosen a-priori for our approximation procedure.

After having established the structure of a single solution space V, corresponding to a specific data resolution level Δx, let us now see what additional properties we should require when the data density changes. For a new resolution level $\Delta x'$, we should again be able to obtain a similar isomorphic mapping T' between a new solution space $\mathrm{V}' \subset \mathrm{H}$ and the observation space $l^2(Z)$. Note that the observation space contains again all square-summable sequences, obtained though with a different sampling rate from the unknown signals. In order to maintain the same stability level in the new isomorphism T' (which should also obey the same optimal criteria as T), equation (5.3) must now be satisfied for every $\hat{f} \in \mathrm{V}'$. It is trivial to prove that this will be true, under the scale-invariance condition (4.3) for the norm in the global Hilbert space H, when the two solution spaces (V and V') are related through the isometric scaling of eq.(4.9). If we combine this last property with the requirement of translation–invariance for the new isomorphism T', then the new solution space V' should too be generated by a family of linearly independent functions of the form $\varphi_2(x-n\Delta x')_{n\in Z}$, such that:

$$\varphi_2(x) = (\frac{\Delta x}{\Delta x'})^{1/2}\, \varphi_1(\frac{\Delta x}{\Delta x'} x) \tag{5.4}$$

In essence, the linear approximation procedure should employ the same type of isomorphic mapping at different scales, depending on the spatial resolution Δx of the available sampled data. Thus the underlying approximation model of the procedure described above can be expressed in the general linear form

$$\hat{f}(x) = \sum_n a_n\, \varphi(\frac{x}{\Delta x} - n) \tag{5.5}$$

where the unknown coefficients a_n are related in a stable and unique way to the available discrete data $f(n\Delta x)$ of the unknown field. The computation of a_n from the data, as well as the choice of the approximating kernel $\varphi(x)$, should be based on certain optimality principles imposed for the approximation procedure. In any case, the set of functions $\varphi(\frac{x}{\Delta x} - n)_{n\in Z}$ should always constitute a Riesz (stable) basis for their closed linear span, which is nothing else than the solution space of the linear approximation problem for the specific data resolution level Δx. Some interesting details regarding convolution-based approximation models, like in eq.(5.5), can be found in Unser and Daubechies (1997), and Blu and Unser (1999a,b).

It is also natural to impose a certain 'causality' principle with respect to the quality of signal information that we extract from the available data through the isomorphic approximation mapping. Such a principle can be stated as follows:

$$\mathrm{V} \subset \mathrm{V}' \quad \text{for} \quad \Delta x > \Delta x' \tag{5.6}$$

where V and V' denote the solution spaces for data resolution levels Δx and $\Delta x'$, respectively. Finally the convergence issue for infinitely dense data requires that

$$\lim_{\Delta x \to 0} \mathrm{V} = \mathrm{H} \tag{5.7}$$

According to the original concept of Multiresolution Analysis (MRA) that was introduced by Mallat (1989), we can conclude from the previous discussions that a (constantly stable, convergent, and translation-invariant) linear approximation scheme in an infinite-resolution Hilbert space H requires, basically, the introduction of an MRA subspace structure in H. The solution, at each data resolution level Δx, will then be given by an isomorphic mapping of the form (5.5), where the kernel $\varphi(x)$ is the one that generates the MRA sequence of subspaces. In a way, we have 're-invented' Mallat's MRA concept as a necessary regularization tool for linear approximation problems in infinite-resolution Hilbert spaces.

5.2 Final Remarks

In section 4.2 it was shown that a constantly stable and convergent (in the sense of infinitely dense data) linear approximation scheme in a Hilbert space H is generally incompatible with the existence of a reproducing kernel for H. In the methodology that was used, the only element that can be considered as being arbitrary is the assumption that the norm in the global Hilbert space H satisfies a scale-invariance condition; see, eq.(4.3). Under this assumption, the reproducing kernel of H should have a 'generalized' form similar to the Dirac distribution, in order to maintain stability and convergence. In this way, the Hilbert space H is equipped with certain infinite-resolution characteristics which were discussed in the previous sections. This, however, prohibit us from using the well-established methodology of minimum-norm collocation to compute linear approximations for unknown functions.

Nevertheless, a useful alternative is provided by considering a structured sequence of multiresolution Hilbert subspaces within the global space H, through which we can construct our affine-invariant linear approximation via isomorphic mappings from the observation space $l^2(Z)$. This sequence will depend on a generating scaling function $\varphi(x)$, and also on a specific rule according to which the data resolution level Δx_j changes from one solution space V_j to the next V_{j+1}. Note that we haven't even established yet that such a structure generally exists. From Mallat's developments, we definitely know that it will exist when we identify the infinite-resolution space H as the $L^2(\Re)$ Hilbert space, and when the data resolution level has the dyadic form $\Delta x_j = 2^{-j}$; multi-dimensional extensions are also possible in this case.

The properties of convergence and unperturbed stability, however, are not the only benefits obtained from using this general methodology. What is furthermore achieved is a kind of 'consistency' between the *data resolution* and *model resolution*. The importance and implications of this concept within the linear Hilbertian approximation framework have already been discussed in section 4.3. We should repeat here that by model resolution we mean the existence in a RKHS of a point distribution $\{Q_i\}$ such that the family of reproducing kernel functions $K(P,Q_i)$ provides a stable (Riesz) basis for this space. Equivalently, we can also define model resolution in a more 'engineering' fashion, as the ability of a Hilbert space to reproduce its elements through a certain *sampling theorem* embedded on it. For some interesting aspects on the connection between sampling theorems and RKHS, see Nashed and Walter (1991). In contrast to the notion of data resolution which always has an 'absolute' meaning (i.e., the sampling interval Δx between the data points), the model resolution is more of a 'relative' term since its definition depends not only on the topology of $\{Q_i\}$, but also on the actual form of $K(P,Q_i)$ or the form of the implied sampling functions in the corresponding RKHS. Thus, for a specific grid of data values there may exist many different *data referencing models* (i.e. RKHSs) that can match the available spatial grid resolution by having the same underlying sampling point resolution $\{Q_i\}$, but they will all differ in the specific functional forms $K(P,Q_i)$ they use to describe the signal variations between these data points. Note that the notion of *model resolution*, with respect to linear approximation problems in Hilbert spaces, does not have yet an established rigorous definition in the mathematical and/or geodetic literature, and all the previous postulates represent just the personal view of the authors who have been inspired mainly from the various theoretical and practical problems existing in deterministic collocation.

In our current, MRA-type approximation framework it may be more reasonable to distinguish between *model* and *modelling* resolution. As modelling resolution we should consider the model resolution implied in each of the solution subspaces V_j. The problem of the existence of an appropriate reproducing kernel $K_j(P,Q)$ in each $V_j \subset H$, which will control the modelling resolution at every data resolution level Δx_j, will not be considered here. On the other hand, the general model resolution should be identified as the resolution of the global Hilbert space H enclosing all of our unknown signals, which is infinite (i.e., H does not posses a proper reproducing kernel). The modelling resolution should be always adapted to the resolution of the available discrete data, and that is exactly one of the roles of the isomorphic mapping during the linear approximation procedure.

We should mention that the interplay between data resolution and linear approximation of gravity field signals has already been discussed by Sanso (1987), in an attempt to establish the equivalence between the 'model approach' and the 'operational approach' in Physical Geodesy. Also, K-P. Schwarz in his classic lecture notes (Schwarz, 1984) studied in detail the relationship between various data types resolution and estimated gravity field quantities, but his whole analysis was restricted only in a purely Fourier-based setting, using the well-known Nyquist criterion as the optimal rule to connect data and model resolution.

There still exist various open problems that remain to be solved. The development of actual algorithms for the numerical computation of the isomorphic mapping T is one of them. Such algorithms should be based on specific criteria regarding the properties of the linear approximation (e.g., the data values in the observation space should be reproduced by the

approximated field in the solution space). Also, the problem of selecting the generating scaling function $\varphi(x)$ for the multiresolution subspace sequence needs to be rigorously explored. In a way, this problem replaces the classic norm choice problem in deterministic collocation.

6. Conclusions and Future Work

The central idea of this report was to demonstrate, using appropriate mathematical formulation, that the multiresolution/wavelet theory lies behind some of the most important problems encountered in linear approximation methods for Physical Geodesy applications. In particular, it has been shown that the MRA concept can be actually used as a 'regularization tool' for linear approximation problems in infinite-resolution Hilbert spaces, in order to ensure unperturbed stability for increasing data resolution and convergence in the sense of infinitely dense data. At the same time, the standard formulation of minimum-norm collocation was proven to be insufficient to accommodate these two crucial properties. Some other important issues, like the consistency between the data and model resolution, have also been discussed throughout the report.

Although the present work had a purely theoretical character, all the problems discussed herein are of great importance in practical applications, especially considering the constantly increasing size and resolution of the available gravity field databases. In this way, the reformulation of the classic functional approximation methods in gravity field modelling, in terms of multiresolution approximation in infinite-resolution Hilbert spaces, seems a natural step. A good starting point will be to consider the well-established multiresolution approximation theory in L^2 Hilbert spaces, which is actually a very convenient choice with respect to the use of the always useful frequency-domain formalism.

Numerous theoretical and practical issues, however, need to be resolved in order for the MRA theory to become a standard tool in gravity field modelling applications. The problem of selecting an optimal scaling function for the approximation of an unknown 1D field has already been discussed in Kotsakis (1999a,b) using a statistical collocation approach. The extension of these developments in multi-dimensional cases remains to be made. Also, the study of the approximation error within a multiresolution framework is a very interesting and important topic that should be explored. The inclusion of non-stationary random noise in the available discrete data, along with the development of optimal noise filtering schemes, are some additional important problems requiring detailed investigation. Considerably more difficult are finally the problems of combining heterogeneous types of regularly/irregularly distributed data in the multiresolution approximation framework, as well as the problem of a consistent 'merging' between global and local multiresolution approximation schemes in a remove-restore fashion.

Acknowledgements

Financial support for this research was provided by an NSERC research grant and an Alexander von Humboldt Research Fellowship awarded to the second author.

References

Albert, A. (1972) *Regression and the Moore-Penrose PseudoInverse.* Academic Press.

Aronszajn, N. (1950) Theory of Reproducing Kernels. *Transactions of the American Mathematical Society*, vol. 68, pp. 337-404.

Barzaghi, R. and Sanso, F. (1986) New Results on Convergence Problem in Collocation Theory. *Proceedings of the I Hotine-Marussi Symposium on Mathematical Geodesy*, Rome, Italy, June 3-6, 1985, pp. 417-457.

Benedetto, J.J. (1994) Frame decompositions, sampling, and uncertainty principle inequalities. In: Benedetto, J.J. and Frazier, M.W. (eds.) *Wavelets: Mathematics and Applications*, CRC Press, Inc.

Benedetto, J.J. (1998) Noise reduction in Terms of the Theory of Frames. In: Zeevi, Y. and Coifman, R. (eds.) *Signal and Image Representation in Combined Spaces*, Academic Press.

Benedetto, J.J. and Walnut, D.F. (1994): Gabor Frames for L^2 and related spaces. In: Benedetto, J.J. and Frazier, M.W. (eds.) *Wavelets: Mathematics and Applications*, CRC Press, Inc.

Bjerhammar, A. (1964) A new theory of gravimetric geodesy. *Trans. Roy. Inst. Tech.*, 243, Stockholm, Sweden.

Bjerhammar, A. (1973) On the Discrete Boundary Value Problem in Physical Geodesy. *Reports of the Royal Institute of Technology*, Stockholm, Sweden.

Bjerhammar, A. (1975) Discrete approaches to the solution of the boundary value problem of physical geodesy. *Boll. di Geod. Sci. Affi.*, vol. 34, no. 2, pp. 185-240.

Bjerhammar, A. (1982) On the Foundation of Collocation in Physical Geodesy. *Bull. Geod.*, vol. 56, pp. 312-328.

Bjerhammar, A. (1987) Discrete Physical Geodesy. *Dept. of Geodetic Science and Surveying, Ohio State University, Report No.380*, Columbus, Ohio.

Blu, T. and Unser, M. (1999a) Quantitative Fourier Analysis of Approximation Techniques: Part I – Interpolators and Projectors. *IEEE Transactions on Signal Processing*. (to appear)

Blu, T. and Unser, M. (1999b) Quantitative Fourier Analysis of Approximation Techniques: Part II – Wavelets. *IEEE Transactions on Signal Processing.*(to appear)

Bocchio, F. (1983) On a Gravity Field with Geodetic Singularities. *Proceedings of the 8th Symposium on Mathematical Geodesy (5th Hotine Symposium)*, Como, September 7-9, 1981, pp. 417-428.

Bocchio, F. (1981) Geodetic Singularities in the gravity field of a non-homogeneous planet. *Geophys. J. Royal Astr. Soc.*, vol. 68, pp. 417-428.

Bottoni, G.P. and Barzaghi, R. (1993) Fast collocation. *Bull. Geod.*, 67, pp. 119-126.

Cvetkovic, Z. and Vetterli, M. (1998) Overcomplete Expansions and Robustness. In: Zeevi, Y. and Coifman, R. (eds.) *Signal and Image Representation in Combined Spaces*, Academic Press.

Daubechies, I. (1990) The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961-1005.

Daubechies, I. (1992) *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA.

Daubechies, I., Grossmann, A. and Meyer, Y. (1986) Painless nonorthogonal expansions. *Journal of Mathematical Physics*, vol. 27, no. 5, pp. 1271-1283.

Davis, P.J. (1975) *Interpolation and Approximation.* Dover Publications Inc., New York.

Debnath, L. and Mikusinski, P. (1999) *Introduction to Hilbert Spaces with Applications.* Academic Press.

Dermanis, A. (1976) Probabilistic and Deterministic Aspects of Linear Estimation in Geodesy. *Dept. of Geodetic Science and Surveying, Ohio State University, Report No.244,* Columbus, Ohio.

Dermanis, A. (1977) Geodetic Linear Estimation Techniques and the Norm Choice Problem. *Manuscr. Geod.*, 2, pp. 15-97.

Dermanis, A. and Sanso, F. (1997) *Statistical Foundations of Geomatics.* Lecture Notes. Dept. of Geomatics Engineering, University of Calgary.

Duffin, R.J. and Schaeffer, A.C. (1952) A class of nonharmonic Fourier series. *Transactions of the American Mathematical Society*, vol. 72, pp. 341-366.

Eeg, J. and Krarup, T. (1973) Integrated Geodesy. *Internal Report No.7, Danish Geodetic Institute*, Copenhagen.

Eren, K. (1980) Spectral Analysis of GEOS-3 Altimeter Data and Frequency-Domain Collocation. *Dept. of Geodetic Science and Surveying, Ohio State University, Report No.297,* Columbus, Ohio.

Feichtinger, H.G. and Grochenig, K. (1994) Theory and practice of irregular sampling. In: Benedetto, J.J. and Frazier, M.W. (eds.) *Wavelets: Mathematics and Applications*, CRC Press, Inc.

Gelfand, I.M. and Shilov, G.E. (1964) *Generalized Functions*, Vol. 1 and 2, Academic Press, New York.

Hadamard, J. (1923) *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. Yale University Press, New Haven.

Halmos, P.R. (1991) *Measure Theory*. Graduate Texts in Mathematics Series, vol.18, Springer-Verlag.

Heil, C.E. and Walnut, D.F. (1994) Continuous and Discrete Wavelet Transforms. *Siam Review*, vol. 31, no. 4, pp. 628-666.

Heiskanen, W.A. and Moritz, H. (1967) *Physical Geodesy*. W.H. Freeman, San Francisco.

Holschneider, M. (1995) *Wavelets, An Analysis Tool.* Oxford University Press Inc., New York.

Kailath, T. (1974) A View of Three Decades of Linear Filtering Theory. *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 146-181.

Kaula, W. (1959) Statistical and Harmonic Analysis of Gravity. *J. Geoph. Res.*, vol. 64, pp. 2401-2421.

Kotsakis, C. (1999a) The Multiresolution Character of Collocation. *Journal of Geodesy*. (submitted)

Kotsakis, C. (1999b) Wavelets and Collocation: An Interesting Similarity. Paper presented at the *22nd IUGG General Assembly*, Birmingham, U.K, July 18-30, 1999.

Krarup, T. (1969) A Contribution to the Mathematical Foundations of Physical Geodesy. *Report of the Danish Geodetic Institute*, no.44, Copenhagen.

Leglemann, D. (1979) Analytical Collocation with Kernel Functions. *Bull. Geod.*, vol. 53, pp. 273-289.

Mallat, S.G. (1989) Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\mathbf{R})$. *Transactions of the American Mathematical Society*, vol. 315, no. 1, pp. 69-87.

Mallat, S. and Hwang, W.L. (1992) Singularity Detection and Processing with Wavelets. *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617-643.

Mallat, S.G. (1998) *A Wavelet Tour of Signal Processing.* Academic Press, New York.

Meissl, P. (1976) Hilbert Spaces and their Applications to Geodetic Least-Squares Problems. *Boll. di Geod. e Sci. Affi.*, vol.XXXV, no. 1, pp. 181-210.

Moritz, H. (1962) Interpolation and Prediction of Gravity and their Accuracy. *Institute of Geodesy, Photogrammetry and Cartography, Ohio State University, Report No.24,* Columbus, Ohio.

Moritz, H. (1976) Integral Formulas and Collocation. *Manuscr. Geod.*, vol. 1, pp. 1-40.
Moritz, H. (1978) Statistical Foundations of Collocation. *Dept. of Geodetic Science and Surveying, Ohio State University, Report No.272,* Columbus, Ohio.
Moritz, H. (1980) *Advanced Physical Geodesy.* Herbert Wichmann Verlag, Karlsruhe.
Moritz, H. and Sanso, F. (1980) A Dialogue on Collocation. *Boll. di Geod. e Sci. Affi.*, vol. 39, no. 1, pp. 49-51.
Nash, R.A. and Jordan, S.K. (1978) Statistical Geodesy: An Engineering Perspective. *Proceedings of the IEEE,* vol. 66, no. 5, pp. 532-550.
Nashed, M.Z. (1976) Aspects of Generalized Inverses in Analysis and Regularization. In: Nashed, M.Z. (ed.) *Generalized Inverses and Applications,* New York.
Nashed, M.Z. and Walter, G.G. (1991) General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces. *Mathematics of Control, Signals and Systems,* 4, pp. 363-390.
Naylor, A.W. and Sell, G.R. (1982) *Linear Operator Theory in Engineering and Science.* Springer Verlag.
Papoulis, A. (1991) *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc.
Parzen, E. (1967) *Time Series Analysis Papers.* Holden-Day.
Rao, C.R. and Mitra, S.K. (1971) *Generalized Inverses of Matrices and their Applications.* Wiley.
Rummel, R., Schwarz, K.P. and Gerstl, M. (1979) Least-Squares Collocation and Regularization. *Bull. Geod.*, 53, pp. 343-361.
Sanso, F. (1978) The Minimum Mean Square Estimation Error Principle in Physical Geodesy (Stochastic and Non-Stochastic Interpretation). Paper presented at the *7th Symposium on Mathematical Geodesy (4th Hotine Symposium),* Assisi, Italy, June 8-10.
Sanso, F. (1986) Statistical Methods in Physical Geodesy. In: Sunkel, H. (ed.) *Mathematical and Numerical Techniques in Physical Geodesy,* Springer Verlag.
Sanso, F. (1987) Talk on the Theoretical Foundations of Physical Geodesy. Paper presented at the *IUGG 19th General Assembly,* Vancouver B.C., Canada, August 9-22.
Sanso, F. and Tscherning, C.C. (1980) Notes on Convergence in Collocation Theory. *Boll. di Geod. e Sci. Affi.*, vol. 39, no.3, pp. 123-144.
Sanso, F, and Schuh, W.D. (1987) Finite Covariance Functions. *Bull. Geod.*, 61, pp. 331-347.
Sanso, F. and Sideris, M.G. (1995) On the Similarities and Differences between Systems Theory and Least-Squares Collocation in Physical Geodesy. Paper presented at the *XXI IUGG General Assembly*, Boulder, Colorado, July 2-14, 1995.
Schwartz, L. (1951) *Th□orie des distributions.* Vol. I and II, Herman and Cie, Paris.
Schwarz, K.P., Sideris, M.G. and Forsberg, R. (1990) The use of FFT techniques in Physical Geodesy. *Geophysical Journal International,* vol. 100, pp. 485-514.
Sideris, M.G. (1995) On the use of heterogeneous noisy data in spectral gravity field modeling methods. *Journal of Geodesy,* 70, pp. 470-479.
Sobolev, S.L. (1964) *Partial Differential Equations of Mathematical Physics.* Pergamon Press, London.
Strohmer, T. (1995) *Irregular Sampling, Frames and Pseudoinverse.* Ph.D. Dissertation, Dept. of Mathematics, University of Vienna.
Tscherning C.C. (1977) A note on the choice of the norm when using collocation for the computations of approximations to the anomalous potential. *Bull. Geod.*, 51, pp. 137-147.
Tscherning C.C. (1978) On the Convergence of Least Squares Collocation. *Boll. di Geod. e Sci. Affi.*, vol. XXXVII, no. 2-3, pp. 507-517.
Tscherning, C.C. (1986) Functional Methods for Gravity Field Approximation. In: Sunkel, H. (ed.) *Mathematical and Numerical Techniques in Physical Geodesy,* Springer Verlag.

Unser, M. and Daubechies I. (1997) On the Approximation Power of Convolution-Based Least Squares versus Interpolation. *IEEE Transactions on Signal Processing,* vol. 47, no. 7, pp. 1697-1711.

Walter, G.G. (1992a) Approximation of the Delta Function by Wavelets. *Journal of Approximation Theory*, vol. 71, pp. 329-343.

Walter, G.G. (1992b) A Sampling Theorem for Wavelet Subspaces. *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 881-884.

Walter, G.G. (1994) *Wavelets and Other Orthogonal Systems with Applications.* CRC Press, Inc.

Young, R.M. (1980) *An Introduction to Non-Harmonic Fourier Series.* Academic Press, New York.