Journal of Geodesy (2000) 74: 275–290

Journal of Geodesy

© Springer-Verlag 2000

The multiresolution character of collocation

C. Kotsakis

Department of Geomatics Engineering, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4
e-mail: ckotsaki@ucalgary.ca; Tel.: +1-403-2204113; Fax: +1-403-2841980

Received: 26 January 1999 / Accepted: 16 August 1999

Abstract. An interesting theoretical connection between the statistical (non-stochastic) collocation principle and the multiresolution/wavelet framework of signal approximation is presented. The rapid developments in multiresolution analysis theory over the past few years have provided very useful (theoretical and practical) tools for approximation and spectral studies of irregularly varying signals, thus opening new possibilities for 'non-stationary' gravity field modeling. It is demonstrated that the classic multiresolution formalism according to Mallat's pioneering work lies at the very core of some of the general approximation principles traditionally used in physical geodesy problems. In particular, it is shown that the use of a spatio-statistical (non-probabilistic) minimum mean-square-error criterion for optimal linear estimation of deterministic signals, in conjunction with regularly gridded data, always gives rise to a generalized multiresolution analysis in the Hilbert space $L^2(\mathbf{R})$, under some mild constraints on the spatial covariance function and the power spectrum of the unknown field under consideration. Using the theory and the actual approximation algorithms associated with statistical collocation, a new constructive framework for building generalized multiresolution analyses in $L^2(\mathbf{R})$ is presented, without the need for the usual *dyadic restriction* that exists in classic wavelet theory. The multiresolution and 'non-stationary' aspects of the statistical collocation approximation procedure are also discussed, and finally some conclusions and recommendations for future work are given.

Key words: Multiresolution approximation – Spatio-statistical approximation – Wavelets – Collocation

1 Introduction

There is no doubt that the method of least-squares collocation (LSC) represents one of the major theoretical and practical foundations of modern physical geodesy. Closely related to Bjerhammar's initial idea on discrete underdetermined boundary value problems, collocation has evolved into a powerful linear estimation method for either global or local gravity field modeling. Despite the various different interpretations and their associated mathematical concepts upon which LSC has been based (see e.g. Sanso 1986; Tscherning 1986), a rigorous unified approach that merges both the purely deterministic (Krarup's formulation) and the purely stochastic (Wiener's linear prediction theory) approximation viewpoints behind collocation was achieved by Sanso (1980). Such an approach has eliminated, to some degree, most of the 'pitfalls' associated with each individual original formulation (e.g. reproducing kernel choice problem, non-stochasticity of the actual gravity field); see also Moritz (1980) and Moritz and Sanso (1980). We have therefore been used to considering LSC as a rigorous *statistical* method for gravity field approximation, where the term 'statistical' is used not to describe some underlying stochastic behavior of the actual gravity field, but rather to specify the spatio-statistical nature of the deterministic norm that is used to quantify the approximation error and to optimize the approximation algorithm.

In this paper we will establish an important theoretical connection between this spatio-statistical version of LSC and the fields of multiresolution approximation theory and wavelets. Sanso (1987) had previously discussed the notion of discrete data resolution and how this affects the results of optimal 'operational' estimation methods for continuous gravity field signals. In the lecture notes by Schwarz (1984) the relationship between data resolution and estimated gravity field quantities was also discussed in detail, but it was restricted to a purely Fourier-based analysis setting using the well-known Nyquist principle to connect data resolution and recovered signal information. The developments in multiresolution/wavelet theory over the past few years, on the other hand, have provided us with the necessary mathematical tools to incorporate the data resolution

parameter into our approximation methods in a very attractive way. Wavelet-based approximation models reconstruct the unknown signals according to a *zoom-in/zoom-out* scaling approach, by adding self-similar localized building blocks ('details') up to a maximum scale level dictated from the resolution of the available discrete data. Many different wavelet models have been developed over the last 10 years which can be employed for signal estimation (and other) purposes, all varying in terms of the specific functional waveform that they use to describe signal variations between the data points. The important advantage that multiresolution/wavelet bases have to offer (over the classic Fourier-based spectral techniques) is their convenient localization properties, which makes them a very efficient and useful tool for spectral studies of irregularly varying ('non-stationary') signals. In this way, if we manage to *transform* the original collocation approximation framework into a wavelet-type multiresolution expansion, we will have achieved a major step forward, improving significantly the quality of information that the available data can give us about the behavior of the underlying gravity field.

Few attempts to incorporate the multiresolution theory tools into the optimal approximation principles have been reported in the geodetic literature. Li (1996b) claimed that LSC itself does not take into account the available data resolution, and he applied a combination of LSC-based optimal principles and independent wavelet formalism to connect the two concepts. Following a different approach, Keller (1998) used collocation in its 'Krarup mode' [i.e. minimum-norm approximation in a reproducing kernel (RK) Hilbert space] and applied it to various multiresolution model spaces.

In contrast to Keller's paper, which was considered as a special case of LSC in a priori selected multiresolution Hilbert spaces, the aim of the present study is to show that the use of an optimal mean-square-error (MSE) principle with gridded data (almost) *always* gives rise to a multiresolution framework for signal approximation. Using the theory and the actual approximation algorithms associated with statistical collocation, the present paper will present a new constructive (frequency-domain-based) framework for building generalized multiresolution analyses (and thus wavelet-type bases) in the Hilbert space $L^2(\mathbf{R})$, without the need of the usual *dyadic restriction* that exists in classic wavelet theory. For the sake of simplicity, all the following developments will be restricted to a one-dimensional (1D) setting. The extension into higher dimensions, although not straightforward in some cases, is certainly possible. A brief overview of 1D multiresolution approximation theory and wavelets is given in Sect. 2. The LSC problem is solved in Sect. 3 following Sanso's spatio-statistical spirit (Sanso 1980) and being additionally modified in order to reveal the multiresolution properties of its solution, which are discussed in more detail in Sect. 4. In Sect. 5, the theoretical connection between the LSC solution and multiresolution analysis theory is made, and some comments regarding the 'non-stationarity' aspects of the approximation procedure are given. Finally, some conclusions and recommendations for future work are drawn.

2 Overview of multiresolution theory and wavelets

The concept of *multiresolution theory* for signal approximation is a relatively recent one, originally formulated by Mallat (1989a, b). *Wavelet signal expansions*, on the other hand, existed long before Mallat's developments, with their most common example being the asymptotic approximation of L^2 signals by translates of piecewise-constant base functions (i.e. Haar wavelets). Since there exists a very strong connection between the two concepts, they are usually considered as the two sides of the same coin, although there do exist pathological cases of wavelet expansions which cannot be identified under Mallat's multiresolution framework. In this section we are going to present just a very small sample from these two vast mathematical subjects, restricting ourselves only to what is necessary in order to follow the discussion in the following sections. The interested reader can find a fully comprehensive reference for these subjects, including some of the most recent developments, in the book by Mallat (1998) and in the excellent review paper by Jawerth and Sweldens (1994). Geodetic and geophysical applications of multiresolution/wavelet theory are discussed in the papers by Freeden and Schneider (1998), Kumar and Foufoula-Georgiou (1997), Ballani (1996), Li (1996a) and Keller (1998), among many others.

2.1 Multi-resolution analysis

A *multiresolution analysis* (MRA) in the Hilbert space $L^2(\mathbf{R})$ is defined as an infinite sequence of closed linear Hilbert subspaces $V_j \subset L^2(\mathbf{R})$, having the following five properties:

1. $V_j \subset V_{j+1}, \quad \forall\, j \in \mathbf{Z}$ (1a)
2. $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$ (1b)
3. $f(x) \in V_j \Leftrightarrow f(x + n2^{-j}) \in V_j, \quad \forall\, n \in \mathbf{Z}$ (1c)
4. $\overline{\bigcup_{j=-\infty}^{+\infty} V_j} = L^2(\mathbf{R})$ and $\bigcap_{j=-\infty}^{+\infty} V_j = \{0\}$ (1d)
5. A *scaling function* $\varphi(x) \in V_o$, with a non-vanishing integral, exists such that the family $\varphi(x-n)_{n\in\mathbf{Z}}$ is a Riesz basis of V_o . (1e)

The definition given above is not minimal, in the sense that some of the conditions (1)–(5) can be derived from the remaining ones (Wojtaszczyk 1997). However, it has been customary to use all five properties of an MRA as independent statements. A *Riesz basis* is just a generalization of the notion of an orthonormal basis in Hilbert spaces, corresponding to a set of linearly independent functions that forms a complete 'oblique' and stable system of reconstructing elements. If we have a Riesz basis $\varphi_n(x)_{n\in\mathbf{Z}}$ in a Hilbert space H, then there always exists a unique *biorthonormal* system $\tilde{\varphi}_n(x)_{n\in\mathbf{Z}}$ which also forms a Riesz basis for H. The biorthonormality

property between the two systems can be expressed through the relation

$$\langle \varphi_n(x), \tilde{\varphi}_m(x) \rangle = \delta_{n,m} \tag{2}$$

where $\langle , \rangle$ denotes the inner product in the Hilbert space H, and $\delta_{n,m}$ is the Kronecker delta. More details for Riesz bases can be found in Young (1980), Heil and Walnut (1994) and Wojtaszczyk (1997).

If a Riesz basis (under the usual L^2 inner product) is formed by the translates $\varphi(x-n)_{n\in\mathbf{Z}}$ of a basic scaling function $\varphi(x)$, then the Fourier transform $\Phi(\omega)$ of the scaling function should satisfy the following condition (see e.g. Jawerth and Sweldens 1994):

$$0 < A \leq \sum_k |\Phi(\omega + 2\pi k)|^2 \leq B < +\infty \tag{3}$$

for some strictly positive, finite bounds A and B. If Eq. (3) is true, then the sets $\varphi(x/h-n)_{n\in\mathbf{Z}}$ form Riesz bases for their corresponding linear spans, for every non-zero value of the scaling parameter h (Unser and Daubechies 1997). In this way, the collection of functions $\varphi(2^j x-n)_{n\in\mathbf{Z}}$ will also form a Riesz basis in each corresponding subspace V_j of a dyadic MRA. Condition (3), for the special case where the family $\varphi(x-n)_{n\in\mathbf{Z}}$ is an *orthonormal* set, takes the simple form

$$\sum_k |\Phi(\omega + 2\pi k)|^2 = 1 \tag{4}$$

In every nested subspace V_j of an MRA, an *infinite* number of complete orthonormal sets can be constructed from a given Riesz basis $\varphi(2^j x-n)_{n\in\mathbf{Z}}$, according to the orthonormalization trick given in Young (1980, p. 48); see also Holschneider (1995, p. 187) and Wojtaszczyk (1997, pp. 24–25). These orthonormal sets will, too, be comprised of integer translates of a basic function. For the same MRA in $L^2(\mathbf{R})$, therefore, we can have many different choices for the generating scaling function. Let us denote by $\varphi(x)$ a scaling function which generates a complete orthonormal system for a certain MRA. Each subspace V_j of this MRA is a RK Hilbert space (under the usual L^2 inner product), with its reproducing kernel $k_j(x,y)$ given by

$$k_j(x,y) = 2^j k(2^j x, 2^j y) \tag{5}$$

and

$$k(x,y) = \sum_n \varphi(x-n)\,\varphi(y-n) \tag{6}$$

where $k(x,y)$ is the RK of the 'unit' resolution subspace V_o. For some technical mathematical conditions, see the paper by Walter (1992). It can also be shown that the collection of functions $k(x,n) = k(x-n,0)_{n\in\mathbf{Z}}$ provides an alternative Riesz basis for V_o (Walter 1992). The biorthonormal basis corresponding to $k(x,n)_{n\in\mathbf{Z}}$ has some very special properties, namely being a *sampling basis* for the same subspace V_o. The expansion of an arbitrary signal $f(x) \in V_o$, with respect to such a basis, takes the form of a *sampling theorem* associated with the specific subspace V_o, i.e.

$$f(x) = \sum_n f(n)\, s(x-n) \tag{7}$$

The situation can easily be extended for an arbitrary MRA subspace V_j. Further details on the connection between sampling theorems and L^2 multiresolution theory can be found in Aldroubi and Unser (1992, 1994), Xia and Zhang (1993), Zayed (1993) and Walter (1992). Another excellent reference is also Nashed and Walter (1991), where the notion of sampling theorems is studied in a general, arbitrary Hilbert space setting.

2.2 Multi-resolution approximation through orthogonal projection

The original definition of an MRA, according to Mallat (1989b), differs slightly from the one given in Eq. (1). In Mallat's definition, instead of introducing a priori a scaling function $\varphi(x)$, the central role is played by a sequence of orthogonal projectors P_j associated with a sequence of translation-invariant and dyadically nested subspaces $V_j \subset L^2(\mathbf{R})$, according to properties (1a), (1b) and (1c). These projectors are used to determine the best linear approximation of an arbitrary signal $f(x) \in L^2(\mathbf{R})$ at a specific dyadic resolution level 2^{-j}. The *consistency* of this approximation scheme was also enforced by the fact that P_j should converge to the identity operator as the resolution index j increases. The existence of a scaling function, whose integer translates generate the sequence of the corresponding nested subspaces, can then be proven according to the fundamental theorem given in Mallat (1989b).

Assuming that $\varphi(x)$ is an orthonormal scaling function in some MRA, the orthogonal projection of an arbitrary signal $f(x) \in L^2(\mathbf{R})$ onto a nested subspace V_j will be given by the formulas (see e.g. Mallat 1989a)

$$(P_j f)(x) = \sum_n a(n)\ \varphi(2^j x - n) \tag{8a}$$

$$a(n) = 2^j \int f(x)\ \varphi(2^j x - n)\, \mathrm{d}x \tag{8b}$$

The projection procedure is illustrated in Fig. 1, where the first filter $\Phi(-2^{-j}\omega)$ from Eq. (8b) has a kind of 'anti-aliasing' role for the given dyadic resolution level. In Unser and Daubechies (1997) and Blu and Unser (1999), the above orthogonal projection scheme is analytically described for the general case where a non-orthonormal Riesz basis is used in the multiresolution framework.

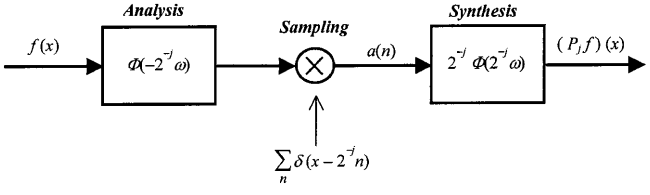

Fig. 1. Orthogonal projection onto a subspace V_j of an MRA (use of orthonormal scaling function)

2.3 Wavelets

Associated with every MRA in $L^2(\mathbf{R})$ is a corresponding *wavelet basis*, which provides the means to connect consistently and efficiently signal information from different dyadic resolution levels, according to a *zoom-in/zoom-out* approach. If we denote by W_j the orthogonal complement of the linear subspace V_j in V_{j+1} (i.e. $V_j \oplus W_j = V_{j+1}$), then there exists a basic *mother wavelet* function $\psi(x)$ such that the family $\psi(2^j x - n)_{n\in\mathbf{Z}}$ will provides a Riesz basis for every W_j. The Hilbert space $W_j \subset L^2(\mathbf{R})$ contains basically the signal 'local details' needed to go from the dyadic scale level 2^j to the next upper scale level 2^{j+1} within the specific MRA $\{V_j\}$. Furthermore, the collection of all these Riesz wavelet bases (from all the different detail subspaces W_j) will form a single Riesz basis for the whole Hilbert space $L^2(\mathbf{R})$ (see e.g. Jawerth and Sweldens 1994).

In this way, the study of a signal at a specific dyadic resolution level 2^{-j_k} can be considerably enriched by computing its wavelet coefficients at coarser resolution values $2^{-j} > 2^{-j_k}$, i.e.

$$f(x) = \sum_n a(n)\varphi(2^{j_k}x - n) = \sum_{j=-\infty}^{j_k-1}\sum_n b(n,j)\psi(2^j x - n) \quad \forall f \in V_{j_k} \tag{9}$$

The wavelet spectrum $b(n,j)$ can be used for a *spatially localized* analysis of the signal behavior, providing in this way a very useful tool (over classical Fourier-based methods) for spectral studies of irregularly varying fields. Detailed algorithms for the computation of the wavelet coefficients $b(n,j)$ from the scaling coefficients $a(n)$ can be found in many places in the wavelet literature and they will not be given here (see e.g. Mallat 1998). These algorithms cover all possible cases, from the simplest one where the translates of both the scaling function $\varphi(x)$ and the mother wavelet $\psi(x)$ provide orthonormal bases for their corresponding spaces, to the most complicated case where the translates of $\varphi(x)$ and/or $\psi(x)$ create just general non-orthogonal Riesz bases.

3 Optimal linear approximation and data resolution

In this section, the optimal linear approximation problem for an unknown *deterministic* field, $g \in L^2(\mathbf{R})$, will be solved in such a way that the immediate connection between the approximated field $\hat{g}$ and the available data resolution will explicitly appear in the solution formulas. In particular, the final optimal estimate $\hat{g}$ will be seen to depend only on a basic kernel $\varphi \in L^2(\mathbf{R})$, which is scaled accordingly to 'match' the given data resolution level. We will assume that the available discrete data represent noiseless point values $g(nh)$ of the unknown field itself, taken on a uniform grid with known resolution level h. The field will be considered as 1D for simplicity. The multi-dimensional case [i.e. when the unknown field belongs in the $L^2(\mathbf{R}^2)$ or in the $L^2(\mathbf{R}^3)$ Hilbert space] is just a straightforward extension of the following derivations.

3.1 General formulation

Since we are seeking a *linear* approximation, the recovered signal $\hat{g}(x)$ will have the general form

$$\hat{g}(x) = \sum_n g(nh)\,\varphi_{n,h}(x) \tag{10}$$

where $\varphi_{n,h}(x)$ is a family of unknown base functions which should be optimally selected to approximate $g(x)$. The dependence (if any) of these base functions on the data resolution is introduced through the use of the subscript h. If we further impose the condition of *translation-invariance* for the estimated field $\hat{g}$ with respect to the reference system used to describe the position of the data points (in the multi-dimensional case this becomes invariance under more general transformations of the reference system), then the family $\varphi_{n,h}(x)$ should be generated from a single kernel $\varphi_h(x)$, such that

$$\varphi_{n,h}(x) = \varphi_h(x - nh) \tag{11}$$

and Eq. (10) becomes

$$\hat{g}(x) = \sum_n g(nh)\,\varphi_h(x - nh) \tag{12}$$

The above approximation formula can now be illustrated in terms of the linear filtering procedure shown in Fig. 2. Applying the Fourier transform to the previous convolution equation, we obtain

$$\hat{G}(\omega) = \bar{G}_h(\omega)\,\Phi_h(\omega) \tag{13}$$

where $\hat{G}(\omega)$ and $\Phi_h(\omega)$ are the Fourier transforms of the approximated signal and the basic "interpolating" kernel $\varphi_h(x)$, respectively. The term $\bar{G}_h(\omega)$ corresponds to the *periodic* Fourier transform of the generalized function

$$\bar{g}_h(x) = g(x)\sum_n \delta(x - nh) = \sum_n g(nh)\,\delta(x - nh) \tag{14a}$$

and it has the form (Oppenheim and Schafer 1989)

$$\bar{G}_h(\omega) = \frac{1}{h}\sum_k G\left(\omega + \frac{2\pi k}{h}\right) = \sum_n g(nh)\mathrm{e}^{-i\omega nh} \tag{14b}$$

with $G(\omega)$ being the Fourier transform of the true unknown signal $g(x)$ and $\delta(x)$ the classic Dirac delta function.

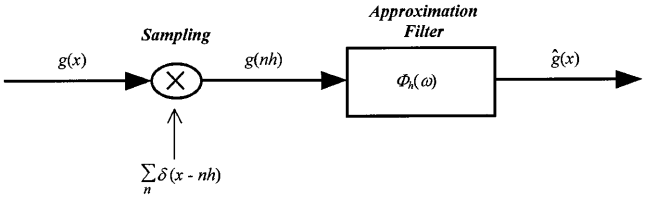

Fig. 2. Filtering configuration of linear, translation-invariant signal approximation using discrete samples

Note that the previous frequency domain formulas imply that we have sampled the unknown signal $g(x) \in L^2(\mathbf{R})$ over its *entire* (finite or infinite) support. If the available data grid $g(nh)$ covers only some limited part of the signal's support, then the previous Fourier transform formalism is certainly not valid and a rectangular window function should be additionally incorporated. In order to avoid such complications, we will assume that the unknown field $g(x)$ covers only the region inside the given data grid boundaries. Although such an assumption may be unacceptable for applications involving temporal signals with finite data grids (where predictions into the future may be required), it nevertheless provides a very reasonable framework for local approximation studies in spatial fields. It should also be emphasized that, even though $g(x)$ is assumed zero outside the data grid boundaries, its approximation $\hat{g}(x)$ by Eq. (12) may exhibit a non-zero pattern outside the data grid. Of course, the theoretical case of infinitely extended 1D data grids is still embedded in all the previous equations.

Another, more technical, condition that should also be imposed in order for the previous frequency domain framework to be rigorously correct, is to assume that the available data sequence $g(nh)$ is always 'measurable', in the following sense:

$$\sum_n |g(nh)| < \infty \tag{15}$$

Indeed, under such a condition the periodic Fourier transform $\bar{G}_h(\omega)$ in Eq. (14b) will always converge uniformly to a finite, continuous function of ω (see e.g. Oppenheim and Schafer 1989, p. 47).

3.2 A spatio-statistical optimal principle

The approximation error, in both the space and the frequency domain, for the given data configuration $g(nh)$ is

$$e(x) = g(x) - \hat{g}(x), \quad E(\omega) = G(\omega) - \hat{G}(\omega) \tag{16}$$

and its power spectrum can easily be derived by taking Eq. (13) into account, i.e.

$$\begin{aligned} |E(\omega)|^2 &= E(\omega)E^*(\omega) \\ &= G(\omega)G^*(\omega) - \Phi_h^*(\omega)G(\omega)\bar{G}_h^*(\omega) \\ &\quad - \Phi_h(\omega)\bar{G}_h(\omega)G^*(\omega) \\ &\quad + \Phi_h(\omega)\Phi_h^*(\omega)\bar{G}_h(\omega)\bar{G}_h^*(\omega) \end{aligned} \tag{17}$$

where the asterisk * denotes complex conjugation.

The sampled sequence $g(nh)$, however, is not the only possible information that we could have extracted from the unknown signal at the given resolution level h. If we shift the sampler (or impulse train) $\sum_n \delta(x - nh)$ by an amount x_o, an infinite number of *different* data sequences can be obtained, which all represent different sampling schemes for the same unknown signal at the same resolution. This situation is illustrated in Fig. 3,

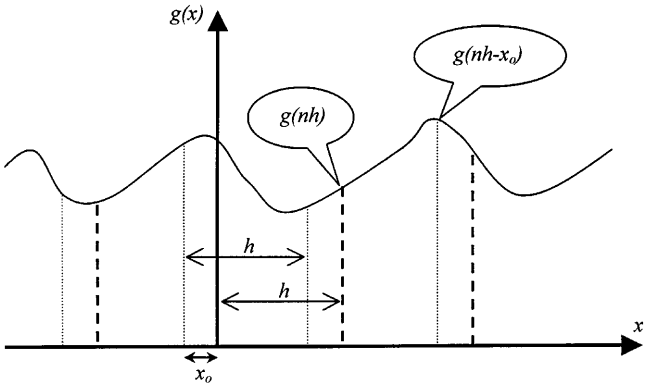

Fig. 3. Different sampling configurations at a given resolution level h

from which we can see that (at a specific resolution value h) all the possible sampled sequences of $g(x)$ can be described by the general form $g(nh - x_o)$, where the *sampling phase* parameter x_o varies in the range $-h/2 \le x_o \le h/2$.

In accordance with the translation-invariance condition for the approximation framework, the general linear equation for the approximated signal from an *arbitrary* sampled sequence at the resolution level h will have the form

$$\hat{g}(x, x_o) = \sum_n g(nh - x_o)\varphi_h(x + x_o - nh) \tag{18}$$

The Fourier transform of Eq. (18), considered as a function of x only, yields

$$\hat{G}(\omega, x_o) = \frac{1}{h}\Phi_h(\omega)\sum_k G\left(\omega + \frac{2\pi k}{h}\right)\mathrm{e}^{-i\frac{2\pi k}{h}x_o} \tag{19}$$

where it is again assumed that all possible sampled sequences $g(nh - x_o)$ of the unknown field $g(x)$ are always measurable in the sense of Eq. (15). Thus, for each different sampling phase value x_o we will have a correspondingly different approximation error $e(x, x_o)$, i.e.

$$e(x, x_o) = g(x) - \hat{g}(x, x_o) \tag{20a}$$

whose Fourier transform is

$$E(\omega, x_o) = G(\omega) - \frac{1}{h}\Phi_h(\omega)\sum_k G\left(\omega + \frac{2\pi k}{h}\right)\mathrm{e}^{-i\frac{2\pi k}{h}x_o} \tag{20b}$$

The optimal criterion for choosing the best estimation kernel $\Phi_h(\omega)$ will be

$$P_e(\omega) = \frac{1}{h}\int_{-h/2}^{h/2} |E(\omega, x_o)|^2 \, \mathrm{d}x_o = \min \tag{21}$$

Equation (21) represents a *minimum mean-square-error* (MMSE) principle, expressed in the frequency domain. The above quantity $P_e(\omega)$ is nothing other than the *mean*

error power spectrum. Note that the term 'mean' is not used in a probabilistic sense (as in the classic Wiener linear prediction theory), but rather it has a *spatio-statistical* meaning. In other words, the optimization of the linear approximation algorithm does not employ the classic expectation operator considering different 'experiment repetitions', but it is based on the average error over all possible sampling configurations for the *given* data resolution level h. In Appendix A it is proven that

$$\int_{-h/2}^{h/2} |E(\omega, x_o)|^2 \, \mathrm{d}x_o = hC(\omega) - \Phi_h^*(\omega)C(\omega) - \Phi_h(\omega)C(\omega) + \Phi_h(\omega)\Phi_h^*(\omega)\bar{C}_h(\omega) \tag{22}$$

where $C(\omega)$ is the Fourier transform of the spatial covariance (CV) function $c(x)$ of the unknown deterministic signal $g(x)$. This spatial CV function has the usual 'stationary' form

$$c(x) = \int g(y)g(y+x)\mathrm{d}y \overset{\Im}{\longleftrightarrow} C(\omega) = G(\omega)G^*(\omega) = |G(\omega)|^2 \tag{23}$$

where the symbol $\Im$ in the last equation denotes a Fourier transform pair. The term $C(\omega)$ is thus just the usual signal power spectrum, and the term $\bar{C}_h(\omega)$ denotes its following periodization (see Appendix A):

$$\bar{C}_h(\omega) = \frac{1}{h}\sum_k C\left(\omega + \frac{2\pi k}{h}\right) \tag{24}$$

Using Eqs. (21) and (22), we can finally obtain the optimal approximation filter as follows:

$$\Phi_h(\omega) = \frac{C(\omega)}{\bar{C}_h(\omega)} = h\frac{C(\omega)}{\sum_k C\left(\omega + \frac{2\pi k}{h}\right)} \tag{25}$$

For justification of the mathematical procedure that leads to the above result, see Bendat and Piersol [1986, Sect. 6.1.4, Eqs. (6.55)–(6.57)], or Sideris [1995, Eqs. (11)–(13)]. The corresponding optimal space domain kernel $\varphi_h(x)$ can be now expressed through the scaling relationship

$$\varphi_h(x) = \varphi\left(\frac{x}{h}\right) \tag{26}$$

where the generating scaling function $\varphi(x)$ is defined in the frequency domain as follows:

$$\varphi(x) \overset{\Im}{\longleftrightarrow} \Phi(\omega) = \frac{C\left(\frac{\omega}{h}\right)}{\sum_k C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right)} \tag{27}$$

The above result can be easily verified by taking into account the fundamental scaling property of the Fourier transform. Finally, if we combine Eqs. (12) and (26), the optimal translation-invariant linear estimation formula for an unknown deterministic field $g(x)$ according to the MMSE principle of Eq. (21), using its discrete samples on a uniform grid with resolution level h, will have the wavelet-like form

$$\hat{g}(x) = \sum_n g(nh)\varphi\left(\frac{x}{h} - n\right) \tag{28}$$

It is worth mentioning that the basic reconstructing kernel $\varphi(x)$ will always be a *symmetric* function, since its Fourier transform in Eq. (27) is always real-valued [i.e. the signal power spectrum $C(\omega)$ is always a real-valued function].

3.3 Comments

The approximation of unknown deterministic functions in terms of convolution-based linear models of the form of Eq. (28) is very common in many signal processing applications in the context of classical interpolation, quasi-interpolation, and multi-scale approximation through projections into multiresolution subspaces (see e.g. Aldroubi and Unser 1994; Unser and Daubechies 1997; Blu and Unser 1999). In such cases, however, the selection of the kernel $\varphi(x)$ is usually made a priori (e.g. sinc-based interpolation, polynomial spline interpolation, etc.), and its performance is evaluated according to an assumed behavior for the unknown signal (e.g. band-limitedness, spectrum decay rate, smoothness, etc.) and/or certain theoretical error bounds which depend on the form of the used kernel (i.e. Strang–Fix conditions); for more details, see Unser and Daubechies (1997). In the present paper, on the other hand, we have a priori introduced a spatio-statistical error power spectrum as a specific accuracy measure for the linear approximation algorithm, which is then optimized in order to choose the 'best' approximation kernel $\varphi(x)$ for the specific unknown signal $g(x)$. The translation-invariance condition, which was also imposed in the estimation procedure, makes this optimal kernel to depend only on the 'stationary' spatial CV function of the unknown field under consideration, according to Eq. (27). The additional dependence of $\varphi(x)$ on the data resolution level h, as it is evident from Eq. (27), will be discussed in detail in the next section.

In our derivations we never required that the optimally approximated signal should reproduce the available noiseless data, i.e. $\hat{g}(nh) = g(nh)$. However, this will always be satisfied since the optimal kernel $\varphi(x)$, defined by Eq. (27), is a *cardinal* (sampling) function. This simply means that

$$\varphi(n) = \begin{cases} 1, & n = 0 \\ 0, & n = \pm 1, \pm 2, \pm 3, \ldots \end{cases} \tag{29a}$$

Indeed, using Eq. (27) we easily see that the Fourier transform $\Phi(\omega)$ of the optimal approximation kernel satisfies the relation

$$\sum_n \Phi(\omega + 2\pi n) = \frac{\sum_n C\left(\frac{\omega + 2\pi n}{h}\right)}{\sum_k C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right)} = 1 \tag{29b}$$

which assures, through the well-known *Poisson summation formula*, that the corresponding space domain function is cardinal. Some mild technical conditions on the signal power spectrum $C(\omega)$, required to ensure the validity of Eq. (29b), will be discussed in the next sections.

An interesting similarity exists between the derived optimal approximation filter in Eq. (25) and the classic *Wiener filter* for noisy, stationary random signals. According to Wiener's linear prediction theory, the optimal filter is defined as the ratio between the power spectral densities (PSDs) of the noiseless stochastic signal and the noisy input signal; see Sideris (1995). This is very similar to Eq. (25), where the numerator $C(\omega)$ is the Fourier transform of the spatial CV function of the true deterministic signal $g(x)$, and the denominator $\bar{C}_h(\omega)$ can be identified as the Fourier transform of the CV function of the 'noisy' input signal $\bar{g}_h(x)$ [see Eq. (14a)]. In our case, the noise takes the form of the lost signal information due to the discretization of $g(x)$ (aliasing error), shown in Fig. 2.

It should be noted that, in contrast to Wiener filtering theory, no stochastic concepts are used in the present paper for the linear approximation problem. The term *covariance function*, which has been used throughout this section, should be understood in a purely deterministic spatial sense [Eq. (23)] and not in any stochastic context under some stationarity and ergodicity assumption. This is especially important in view of the 'stationarity restriction' problem which is believed to exist in the statistical collocation framework. Our formulation can be considered as 'stationary' only in the sense that we use a 1D covariance function for 1D signals, which results solely from the logical requirement of having a translation-invariant approximation scheme (i.e. independent of the origin of the reference system used to describe the position of the data points). See also the related discussion given in Sanso (1980). This, however, does not mean that the approximated/unknown signals have (or should have) a uniform behavior across their domain, and it certainly does not exclude us from obtaining *localized* information for this varying behavior.

4 Statistical collocation as a multiresolution approximation

The final result of Sect. 3.2 is quite general and it did not involve any special concepts from Mallat's multiresolution theory. The fact that the statistical collocation framework actually leads to a *scale-invariant* signal approximation scheme (i.e. independent of the scale of the reference system used to describe the position of the gridded data points), similar to the one encountered in wavelet approximation theory, is rather remarkable. However, there is a significant difference between the optimal collocation model of Eq. (28) and the classic wavelet-based approximation methodology, due to the fact that the optimal kernel $\varphi(x)$ associated with the collocation case is now changing for every different data resolution level h, according to the frequency domain form in Eq. (27).

The most appropriate way to describe the behavior of the signal approximation model of Eq. (28), with the associated kernel $\varphi(x)$ defined by the optimal frequency domain form in Eq. (27), is to characterize it as: (1) *translation-invariant*, (2) *scale-invariant* and (3) *data-resolution-dependent*. Regardless of the origin and the scale of the reference system used to describe the physical/spatial position of a *given* set of gridded data points, the approximated field according to the statistical collocation algorithm will always have the same form/shape. Let us briefly demonstrate the scale-invariance aspect of the collocation algorithm (a similar methodology can also be employed for the translation-invariance aspect). If we use a new reference system $x' = x/a$ to describe the original unknown field $g(x)$ and the position of its point data values $g(nh)$, then we basically want now to approximate a new unknown field $g'(x) = g(ax)$ using its point data values $g'(nh') = g'(nh/a) = g(nh)$. The application of the basic estimation formula of Eq. (28) yields

$$\begin{aligned}\hat{g}'(x) &= \sum_n g'(nh')\varphi\left(\frac{x}{h'} - n\right) \\ &= \sum_n g'\left(n\frac{h}{a}\right)\varphi\left(\frac{x}{h/a} - n\right) \\ &= \sum_n g(nh)\varphi\left(\frac{ax}{h} - n\right) = \hat{g}(ax)\end{aligned} \tag{30}$$

which demonstrates the scale-invariance property of the spatio-statistical collocation. Note that the sampling resolution of the unknown deterministic field g is the same for both reference systems x' and x (i.e. we use the *same* point data values each time). The above situation of scale-invariant signal approximation, for a certain data resolution level h, is illustrated in an abstract way in Fig. 4.

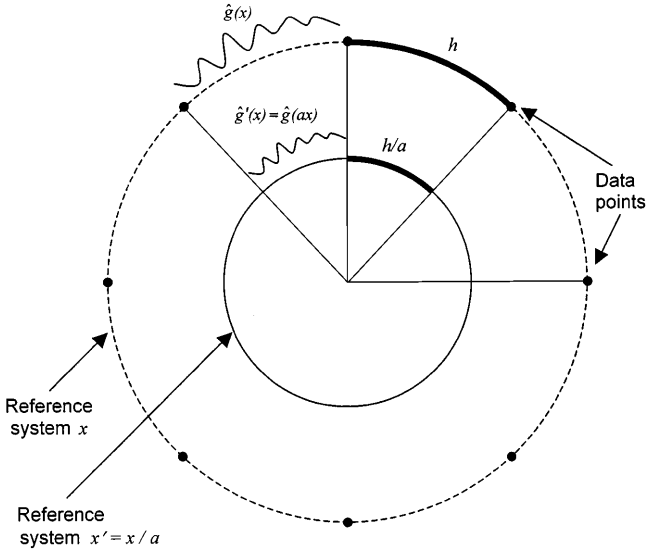

Fig. 4. Scale-invariant signal approximation at a certain data resolution level h (the value of the scaling parameter a is assumed $a > 1$)

The optimal kernel $\varphi(x)$ in the statistical collocation model of Eq. (28) is appropriately *scaled* (shrunk or expanded) in order to 'match' the resolution level h of the given data grid $g(nh)$, as this is expressed in the scale of the used reference system. The final approximated field $\hat{g}(x)$ is then formed by adding translates of the scaled optimal kernel $\varphi(x/h)$, which are centered at all data points. Although such a linear approximation scheme obeys very closely the classic multiresolution/wavelet spirit, it cannot really be identified as such since the actual form of $\varphi(x)$ is also a function of the data resolution h itself. On the other hand, the standard wavelet approximation theory requires the use of a *fixed* scaling kernel $\varphi(x)$, which is just tuned in the desired resolution level of the signal approximation by proper dyadic scalings (see Sect. 2).

In order to better understand the above essential difference, we should express the optimal kernel $\varphi(x)$ associated with the statistical collocation procedure in the following parameterized form [see Eq. (27)]:

$$\varphi(x,h) \overset{\Im}{\longleftrightarrow} \Phi(\omega,h) = \frac{C\left(\frac{\omega}{h}\right)}{\sum_k C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right)} \tag{31a}$$

where the data resolution h plays just the role of an additional constant parameter in the last equation. According to the fundamental scaling property of the Fourier transform, the scaled version $\varphi(x/h) = \varphi_h(x)$ of the optimal approximation kernel $\varphi(x)$ will thus have the following frequency domain form:

$$\varphi\left(\frac{x}{h},h\right) \overset{\Im}{\longleftrightarrow} h\Phi(h\omega,h) = h\frac{C(\omega)}{\sum_k C\left(\omega + \frac{2\pi k}{h}\right)} \tag{31b}$$

which is identical to the optimal Wiener-like approximation filter that was determined in Sect. 3.2, Eq. (25). For each different value of the data resolution parameter h, the optimal kernel $\varphi(x)$ in Eq. (31a) will assume a correspondingly different waveform, and hence the basic estimation model of Eq. (28) will not employ scaled versions of the *same* $\varphi(x)$ for every data sampling level h. Therefore, we see that the statistical collocation concept not only produces a scale-variant signal approximation, but in addition also 'forces' the behavior of its basic approximation kernel to be adapted to the current data resolution in a certain optimal fashion, as suggested by Eq. (31a). It is very important to note that, regardless of the actual value of h, the function $\varphi(x,h)$ *always* corresponds to a cardinal (sampling) kernel, as was explained in Sect. 3.3.

The varying behavior of the optimal approximation kernel $\varphi(x,h)$, for different data resolution levels h, is shown in Figs. 5 and 6. Two different models for the power spectrum $C(\omega)$ of the underlying unknown signal are used. In particular, Fig. 5 shows the Fourier transform $\Phi(\omega,h)$ from Eq. (31a) for the case where the signal power spectrum has a Gaussian form, i.e.

$$C(\omega) = B\mathrm{e}^{-\omega^2} \tag{32a}$$

whereas Fig. 6 illustrates the Fourier transform of the optimal approximation kernel for the case where the signal power spectrum follows a slower decaying pattern than the Gaussian, as follows:

$$C(\omega) = \frac{B}{1+\omega^2} \tag{32b}$$

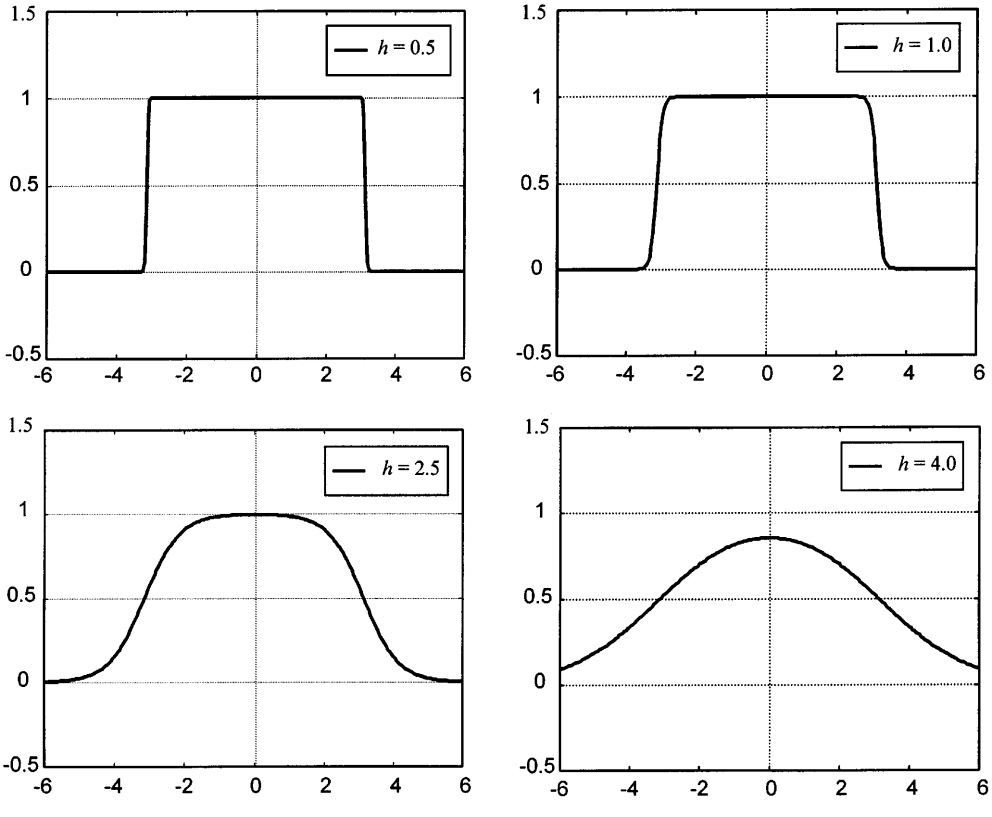

Fig. 5. Fourier transform $\Phi(\omega,h)$ of the optimal approximation kernel $\varphi(x,h)$ for various data resolutions levels h. The underlying unknown signal is assumed to follow a Gaussian power spectrum $C(\omega)$

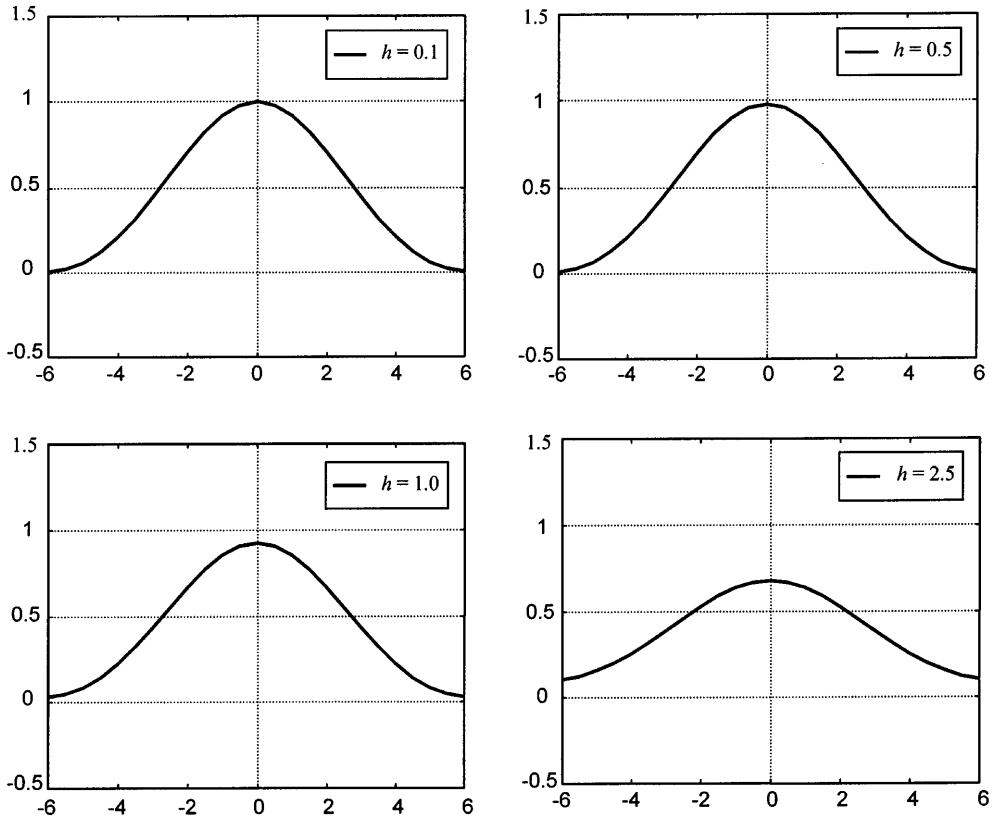

Fig. 6. Fourier transform $\Phi(\omega, h)$ of the optimal approximation kernel $\varphi(x, h)$ for various data resolutions levels h. The underlying unknown signal is assumed to follow the power spectrum model $C(\omega)$ given in Eq. (32b)

with the symbol B denoting just an arbitrary constant value for both cases.

These graphs help considerably in understanding the (somewhat peculiar) behavior of the optimal approximation kernel in the statistical collocation framework. Under proper mild conditions on the signal power spectrum $C(\omega)$, the function $\varphi(x,h)$ in Eq. (31a) will asymptotically converge to a well-defined $L^2(\mathbf{R})$ cardinal kernel as $h \to 0$. All the individual members (i.e. functions) of this convergent sequence will be $L^2(\mathbf{R})$ cardinal kernels as well. In the case of Fig. 5, for example, it is obvious that the optimal approximation kernel will gradually converge to the $\mathrm{sinc}(x)$ function. On the other hand, as the data sampling resolution decreases ($h \to \infty$), the optimal approximation kernel $\varphi(x,h)$ will gradually become the zero function in the $L^2(\mathbf{R})$ sense, as it is evident from the behavior of its Fourier transform in both Figs. 5 and 6. The rigorous mathematical proof of the above statements, as well as the derivation of the necessary mild conditions on the signal power spectrum $C(\omega)$, are beyond the scope of the present paper and they will not be presented here. Some relevant details can be found in the next section.

5 A new constructive framework for generalized multiresolution analyses

In this section we will explore in more detail the actual connection between the statistical collocation approximation model of Eqs. (27) and (28), and the classic MRA approximation framework which was briefly presented in Sect. 2. We will also attempt to clarify a few mathematical details that were left unjustified in the previous sections. In particular, it will be shown that under certain mild conditions on the spatial CV function and the power spectrum of the unknown signal $g(x)$, the corresponding optimal kernel $\varphi(x,h)$ of Eq. (31a) produces a generalized MRA-type approximation scheme in the Hilbert space $L^2(\mathbf{R})$.

5.1 Basic MRA properties of the optimal approximation kernel

First, we need to establish that the optimal kernel in statistical collocation, as it is given in Eq. (27) or Eq. (31a), is a well-defined function in the $L^2(\mathbf{R})$ Hilbert space for any real positive value of the data resolution h. Using Eq. (31a), the L^2 norm of the optimal kernel $\varphi(x,h)$ takes the following form:

$$\begin{aligned}\|\varphi(x,h)\|_{L^2}^2 &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |\Phi(\omega,h)|^2\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\left|C\left(\frac{\omega}{h}\right)\right|^2}{\left|\sum_k C\left(\frac{\omega}{h}+\frac{2\pi k}{h}\right)\right|^2}\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{0}^{2\pi} \frac{\sum_k \left|C\left(\frac{\omega+2\pi k}{h}\right)\right|^2}{\left|\sum_k C\left(\frac{\omega}{h}+\frac{2\pi k}{h}\right)\right|^2}\,\mathrm{d}\omega\end{aligned}$$

$$= \frac{1}{2\pi}\int_0^{2\pi} \frac{\sum_k \left(\frac{1}{h} C\left(\frac{\omega}{h}+\frac{2\pi k}{h}\right)\right)^2}{\left(\sum_k \frac{1}{h} C\left(\frac{\omega}{h}+\frac{2\pi k}{h}\right)\right)^2} \mathrm{d}\omega$$

$$= \frac{1}{2\pi}\int_0^{2\pi} \bar{M}_{2\pi}(\omega)\mathrm{d}\omega \tag{33}$$

where $\bar{M}_{2\pi}(\omega)$ is an auxiliary 2π-*periodic* function, given by the formula

$$\bar{M}_{2\pi}(\omega) = \frac{\sum_k \left(\frac{1}{h} C\left(\frac{\omega}{h}+\frac{2\pi k}{h}\right)\right)^2}{\left(\sum_k \frac{1}{h} C\left(\frac{\omega}{h}+\frac{2\pi k}{h}\right)\right)^2} = \frac{\sum_k a_k^2}{\left(\sum_k a_k\right)^2} \tag{34a}$$

and the discrete infinite sequence a_k has of course the general form

$$a_k = \frac{1}{h} C\left(\frac{\omega}{h}+\frac{2\pi k}{h}\right) \tag{34b}$$

Let it be reminded that the signal power spectrum $C(\omega)$ is always a real-valued, non-negative and even function, which belongs in the $L^1(\mathbf{R})$ space [i.e. since the unknown signal $g(x)$ is assumed to belong in the $L^2(\mathbf{R})$ Hilbert space]. The infinite series $\sum_k a_k$ corresponds to the 2π-*periodic* Fourier transform of a space domain sequence $b[n]$ constructed from the discrete signal covariance values as follows (see e.g. Oppenheim and Schafer 1989):

$$b[n] = c(nh) \tag{35}$$

Therefore, if the discrete sequence $b[n]$ is *absolutely summable*, the series $\sum_k a_k$ will always converge uniformly to a finite, *continuous*, 2π-*periodic* function of ω (Oppenheim and Schafer 1989, p. 47). In this way, we will impose the following basic condition on the signal CV function $c(x)$:

$$\textbf{Condition I:} \quad \sum_n |c(nh)| < \infty, \quad \forall\, h > 0 \tag{36}$$

Note that the above condition is always satisfied in the case where the underlying unknown field $g(x)$ has a finite support in the space domain. A simple example of a CV function with infinite support, for which the above condition is valid, is the Gaussian function. Under condition (36), the series $\sum_k a_k$ will converge uniformly for every value of ω and h, and since *all* its individual terms a_k are always non-negative, the series $\sum_k a_k^2$ will also converge to a finite 2π-*periodic* function of ω for every data resolution level h.

It is also essential to ensure the validity of the following relationship:

$$\sum_k a_k = \sum_k \frac{1}{h} C\left(\frac{\omega}{h}+\frac{2\pi k}{h}\right) \neq 0, \quad \forall\, \omega \in \mathbf{R}, \quad h > 0 \tag{37}$$

There are various different types of conditions, non-contradictory with the first condition given in Eq. (36), that can be imposed on the signal power spectrum $C(\omega)$ in order for Eq. (37) to be true. For the purpose of this paper, we shall simply assume one of the following:

Condition II: (38)

1. $C(\omega) = |G(\omega)|^2 > 0, \quad \forall\, \omega \in \mathbf{R}$ **or**
2. $C(\omega)$ is allowed to vanish only at a finite number of arbitrary isolated points, and/or in a finite number of closed frequency intervals. The signal power spectrum $C(\omega)$ is also allowed to vanish at an infinite number of isolated points without destroying the validity of Eq. (37), as long as these infinite points are not equidistant.

The justification of the previous restrictions on $C(\omega)$ depends basically on the physical properties of the unknown field that we want to approximate. The case where the signal power spectrum $C(\omega)$ vanishes in an infinite frequency interval [i.e. the unknown field $g(x)$ is a *band-limited* signal] requires special consideration, and it will be treated separately in a future publication.

If we further assume that the signal power spectrum $C(\omega)$ is a continuous function, i.e.

$$\textbf{Condition III:} \quad C(\omega) \text{ is continuous for every } \omega \in \mathbf{R} \tag{39}$$

then, under the three previous conditions, the auxiliary term $\bar{M}_{2\pi}(\omega)$ in Eq. (34a) will always converge to a well-defined, finite (bounded), *strictly-positive*, *continuous* and 2π-*periodic function*, and therefore its integral in Eq. (33) will always be a finite number. This makes the optimal approximation kernel $\varphi(x,h)$ a proper $L^2(\mathbf{R})$ function for any real positive value of the data resolution level h.

Finally, the condition that the optimal kernel $\varphi(x,h)$ in statistical collocation has a non-vanishing integral (just like the scaling function of an MRA should have a non-vanishing integral; see Sect. 2.1) requires that its Fourier transform $\Phi(\omega,h)$ does not vanish at the origin. Taking into account Eq. (31a), this is transformed to the following simple condition for the signal power spectrum:

$$\textbf{Condition IV:} \quad C(\omega)_{\omega=0} \neq 0 \tag{40}$$

We are now in position to consider an infinite sequence $\{V_j\}_{j\in\mathbf{Z}}$ of closed linear Hilbert subspaces in $L^2(\mathbf{R})$. Each element of this sequence is defined as the closed linear span of the set

$$\left\{ \varphi\left(\frac{x}{h_j} - n, h_j\right) | n \in \mathbf{Z} \right\}$$

where $\varphi(x,h_j)$ is the optimal approximation kernel given by Eq. (31a), and $h_j > 0$ denotes the data resolution level associated with each subspace V_j. We will further assume that

$$\textbf{Condition V:} \quad h_j > h_{j+1}, \quad \forall\, j \in \mathbf{Z} \tag{41}$$

which makes $\{V_j\}$ a subspace sequence of *increasing resolution* in the Hilbert space $L^2(\mathbf{R})$. Note that the

scaling parameter h_j is not restricted now to dyadic values (i.e. $h_j = 2^{-j}$), as happens in the classic MRA case (see Sect. 2.1). By definition, the above subspace sequence satisfies the third (translation-invariance) basic property of an MRA [see Eq. (1c), Sect. 2.1] for any possible form of the scaling parameter h_j, i.e.

$$f(x) \in V_j \Leftrightarrow f(x + nh_j) \in V_j, \quad \forall\, n \in \mathbf{Z} \tag{42}$$

In order for the specific sequence $\{V_j\}$ to satisfy the first 'nesting' property of an MRA [see Eq. (1a), Sect. 2.1], we have to impose some additional restriction on the way that the value of the scaling parameter h_j changes from one subspace V_j to the next V_{j+1}. In particular, we have to assume that for every $j \in \mathbf{Z}$

$$\textbf{Condition VI:} \quad \frac{h_j}{h_{j+1}} = a_j, \quad \text{where} \quad a_j \in \mathbf{Z}^+ - \{1\} \tag{43}$$

The above condition implies that any two successive scaling parameters associated with the subspace sequence $\{V_j\}$ should be related through an arbitrary positive *integer* number, different from unity. Note that the actual integer value a_j may *change* from one subspace pair (V_j, V_{j+1}) to another (V_{j+1}, V_{j+2}). However, Eq. (43) will ensure that the scaling parameters associated with an arbitrary pair of subspaces $(V_j, V_k)_{j<k}$ are always related through a positive integer number as follows:

$$\frac{h_j}{h_k} = a_j\, a_{j+1} \ldots a_{k-1}, \quad \forall\, j < k \in \mathbf{Z} \tag{44}$$

The special case where the scaling ratio in Eq. (43) assumes a *fixed* positive integer value a (i.e. independent from the index j) occurs if we restrict the data resolution level h_j to take the following exponential form:

$$h_j = a^{-j}, \quad \forall\, j \in \mathbf{Z} \tag{45}$$

where a is now a fixed positive integer number, different from unity. Dyadic subspace schemes (as in the classic MRA framework) will of course arise if we set the value of a in Eq. (45) to be equal to 2. Nevertheless, the general condition of Eq. (43) is all that we actually need in order for the specific subspace sequence $\{V_j\}$ to be nested. The proof is very easy and it can be found in Appendix B.

Furthermore, the subspace sequence $\{V_j\}$ constructed from the optimal approximation kernel $\varphi(x, h_j)$ will also satisfy the fourth ('completeness') basic property of an MRA [see Eq. (1d), Sect. 2.1]. In order to see that, we have to recall the fact that the optimal kernel $\varphi(x, h_j)$ is always a cardinal/sampling function, regardless of the actual value of the data resolution level h_j. In this way, every signal that belongs in an arbitrary subspace $V_j \subset L^2(\mathbf{R})$ of the multiresolution sequence $\{V_j\}$ will have the general form

$$f(x) = \sum_n f(nh_j)\varphi\left(\frac{x}{h_j} - n, h_j\right), \quad \forall\, f(x) \in V_j \tag{46a}$$

By taking into account Eq. (31a) and applying the Fourier transform to the last equation, we obtain the general frequency domain form of every signal $f(x)$ belonging in an arbitrary subspace of the multiresolution sequence $\{V_j\}$, i.e.

$$F(\omega) = \frac{C(\omega)}{\sum_k C\left(\omega + \frac{2\pi k}{h_j}\right)} \sum_n F\left(\omega + \frac{2\pi n}{h_j}\right), \quad \forall\, f(x) \in V_j \tag{46b}$$

As the resolution index j increases, the data resolution level h_j associated with the corresponding subspace V_j becomes smaller and smaller, according to the general condition imposed by Eq. (41). Obviously, when h_j becomes infinitely small $(j \to \infty \Leftrightarrow h_j \to 0)$, then Eq. (46b) will be reduced to a simple identity, i.e.

$$F(\omega) = F(\omega) \tag{47a}$$

which is naturally satisfied by every signal in the Hilbert space $L^2(\mathbf{R})$. In other words

$$\lim_{j \to \infty} V_j = L^2(\mathbf{R}) \tag{47b}$$

On the other hand, when the resolution index j decreases, then the magnitude of the corresponding scaling parameter h_j will be increasing. In the limit, where h_j becomes arbitrarily large $(j \to -\infty \Leftrightarrow h_j \to \infty)$, the right-hand side of the frequency domain equation, Eq. (46b), will be reduced to the form $0 \cdot \infty$, which is equal to zero (Halmos 1991). This simply means that the multiresolution subspace sequence $\{V_j\}$ will be finally 'shrunk' to the zero space in the $L^2(\mathbf{R})$ sense, i.e.

$$\lim_{j \to -\infty} V_j = \{0\} \tag{48}$$

Lastly, we have to check if the family of translates

$$\left\{\varphi\left(\frac{x}{h_j} - n, h_j\right) | n \in \mathbf{Z}\right\}$$

of the optimal approximation kernel forms a *Riesz basis* for every element of the multiresolution subspace sequence $\{V_j\}$ that is spanned by this family. This final MRA property [see Eq. (1e), Sect. 2.1] is especially important, since it will ensure *stable* signal estimation schemes from their discrete samples within every multiresolution subspace V_j. A necessary and sufficient condition for this last property is (Unser and Daubechies 1997)

$$0 < A \le \sum_k |\Phi(\omega + 2\pi k, h_j)|^2 \le B < +\infty, \quad \forall\, \omega \in \mathbf{R},\ h_j > 0 \tag{49}$$

where A and B are some strictly positive bounds, and $\Phi(\omega, h_j)$ is the Fourier transform of the optimal approximation kernel $\varphi(x, h_j)$ at data resolution level h_j. If we take into account Eq. (31a), the above inequality can be easily expressed as a function of the signal power spectrum $C(\omega)$ in the following way:

$$0 < A \leq \frac{\sum_{\mathrm{k}} \left(\frac{1}{h_j} C\left(\frac{\omega}{h_j} + \frac{2\pi k}{h_j} \right) \right)^2}{\left(\sum_{\mathrm{k}} \frac{1}{h_j} C\left(\frac{\omega}{h_j} + \frac{2\pi k}{h_j} \right) \right)^2} = \bar{M}_{2\pi}(\omega) \leq B < +\infty,$$
$$\forall\ \omega \in \mathbf{R},\ h_j > 0 \tag{50}$$

where $\bar{M}_{2\pi}(\omega)$ is the same 2π-*periodic* auxiliary function that was defined and used previously in Eq. (34a). In the beginning of this section we had already established that (under conditions I, II and III) the term $\bar{M}_{2\pi}(\omega)$ will always converge to a well-defined, finite (bounded), *strictly-positive*, *continuous* and 2π-*periodic* function of ω, for every value of the data resolution level h_j. In this way, the existence of both bounds A and B in the double inequality of Eq. (50) is always guaranteed. Hence, the set of integer translates of the optimal approximation kernel

$$\left\{ \varphi\left(\frac{x}{h_j} - n, h_j \right) | n \in \mathbf{Z} \right\}$$

will provide a Riesz basis for every corresponding multiresolution subspace V_j associated with the scaling parameter h_j. Note that the actual numerical values of the two bounds, A and B, will change as h_j changes, which basically means that the *level of stability* of the individual Riesz bases formed by the optimal approximation kernel will not be the same for each V_j.

5.2 Remarks

We have established the fundamental result that: the solution of the linear approximation problem for an unknown deterministic field from its discrete and regularly gridded samples, under the condition of translation-invariance and the spatio-statistical MMSE optimal principle of Eq. (21), gives rise to a generalized MRA-type structure $\{V_j\}$ in the Hilbert space $L^2(\mathbf{R})$. The main difference between this multiresolution subspace structure and the classic MRAs according to Mallat (1989a, b) is that its basic scaling kernel does not have a fixed form, but varies for every different scale level h_j associated with the corresponding subspace V_j. In this case, the power spectrum of the unknown signal under consideration provides the 'generator' of the scaling kernel $\varphi(x, h_j)$ at each resolution level h_j, according to the frequency domain form given in Eq. (31a). Certain conditions have also to be satisfied by the spatial CV function and the power spectrum of the unknown signal, which were discussed in detail in the previous section.

The only traditional MRA property that will not necessarily be satisfied by the subspace sequence $\{V_j\}$, which is generated through the optimal kernel $\varphi(x, h_j)$ of statistical collocation, is the 'self-similar' dyadic scaling condition between the individual subspaces [see Eq. (1b), Sect. 2.1], i.e.

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1} \tag{51}$$

In a way, the above property has now been replaced by the freedom to use a much more flexible rule according to which the scaling parameter (data resolution level) h_j decreases from one nested subspace V_j to the next V_{j+1}, based on the general formula of Eq. (43). Note that the optimal kernel $\varphi(x, h_j)$ essentially generates not just a *single* nested sequence $\{V_j\}$ of multiresolution subspaces in $L^2(\mathbf{R})$, but an infinite number of such subspace sequences. Each of these sequences will depend on a specific formula that we choose to generate the various scale levels h_j [based on the two general conditions of Eqs. (41) and (43)], as well as on the value of a *reference scale level* h_o. A list of such different alternatives is given in Table 1. The classic case, where the nested subspace sequence $\{V_j\}$ is associated with a dyadic scale parameter h_j, is shown in the last two columns of Table 1 for some selected reference values. Even for such dyadic scaling schemes, however, the property in Eq. (51) will not necessarily be satisfied by the generalized MRA sequence $\{V_j\}$ associated with the optimal approximation kernel, unless we impose some further conditions on the signal power spectrum $C(\omega)$.

It is worth mentioning that all the derivations in Sect. 5.1 are valid even if the frequency domain function $C(\omega)$ in the basic equation (31a) does not correspond to the *true* signal power spectrum. This means that we are allowed to use a certain *model* for the signal power spectrum in the construction of the approximation kernel $\varphi(x, h_j)$, without destroying its cardinal/sampling and MRA properties (as long as this model is compatible with the basic conditions given previously, or any other conditions that may be equivalently derived for the same purpose). More importantly, in this case the signal approximation obtained by the statistical collocation algorithm will still converge in a stable way to the true field in the $L^2(\mathbf{R})$ topology, as the data resolution increases ($h_j \to 0$). The optimal MMSE principle of Eq. (21), however, will not be rigorously satisfied in such cases. Nevertheless, the frequency domain structure of Eq. (31a) provides a useful general recipe for building generalized MRAs, based on appropriately selected Fourier transform functions $C(\omega)$ that satisfy the mild conditions given in the previous section.

Table 1. Sample of scale level values h_j associated with different generalized MRA sequences $\{V_j\}$ which are generated from the same scaling kernel $\varphi(x, h_j)$

	Scale level generator [see Eq. (43)]					
	$\frac{h_j}{h_{j+1}} = 2j^2 + 3$		$\frac{h_j}{h_{j+1}} = 2^{\lvert j \rvert + 1}$		$\frac{h_j}{h_{j+1}} = 2$	
	Reference scale value					
	$h_0 = 1$	$h_0 = 0.3$	$h_0 = 1$	$h_0 = 0.3$	$h_0 = 1$	$h_0 = 0.3$
h_3	1/165	1/550	1/64	3/640	1/8	0.0375
h_2	1/15	0.02	1/8	0.0375	1/4	0.075
h_1	1/3	0.1	0.5	0.15	1/2	0.15
h_0	1	0.3	1	0.3	1	0.3
h_{-1}	5	1.5	4	1.2	2	0.6
h_{-2}	55	16.5	32	9.6	4	1.2
h_{-3}	1155	346.5	512	153.6	8	2.4

The previous developments open a new interesting viewpoint for the result of the statistical collocation algorithm in Eq. (28). Under certain conditions, the approximated field $\hat{g}(x)$ will always belong in some multiresolution Hilbert subspace $V_j \subset L^2(\mathbf{R})$ of a generalized MRA sequence, the scale level h_j of which is dictated from the sampling resolution of the available discrete data. The actual collocation approximation algorithm can be considered as the application of a stable *sampling theorem* associated with the specific subspace V_j, since the set of translates

$$\left\{\varphi\left(\frac{x}{h_j} - n, h_j\right) | n \in \mathbf{Z}\right\}$$

of the optimal approximation kernel will always constitute a sampling Riesz basis for V_j. This result is in very close connection with similar mathematical studies, where it was shown that for (almost) every dyadic MRA there exists a unique sampling Riesz basis in each of its nested subspaces (see e.g. Walter 1992; Xia and Zhang 1993). The idea of using sampling expansions for representing gravity field signals is certainly not new, and it has already been discussed by many authors in the context of optimal linear approximation (see e.g. Moritz 1976; Schmidt 1981).

It is also interesting to mention the essential difference between the approximation concept of classic dyadic MRAs according to Mallat (1989a, b), and the present collocation-based multiresolution approximation scenario. Mallat's initial idea was based on the orthogonal projection of the unknown signal $g(x)$ onto a dyadic MRA subspace V_j (see Sect. 2.2). Under this approach, the approximation $\hat{g}(x)$ and the original signal $g(x)$ will not necessarily agree at the data points $x_n = n2^{-j}$, which is not a desirable property within a noiseless data setting. Mallat's procedure could be thought of as starting from the top of a pyramid (i.e. MRA), and by successive orthogonal projections onto more and more detailed resolution subspaces V_j we finally return to the top. In the statistical collocation approach, on the other hand, we start from the 'bottom' of a generalized MRA structure and by obtaining denser and denser sampled values of the unknown field (and correspondingly applying the sampling theorem associated with the 'pyramid') we finally reach the top. It can actually be shown that this 'bottom-to-top' multiresolution approximation, through the use of a scaling cardinal kernel $\varphi(x, h_j)$, corresponds to a certain *oblique* projection scheme within the subspace sequence $\{V_j\}$; see Blu and Unser (1999).

The previous extensions of the MRA concept suggest that we may be able to achieve a similar extension of the classic wavelet bases associated with Mallat's dyadic MRAs. For example, the orthogonal complements W_j of the various nested subspaces in the generalized MRA structure $\{V_j\}$, which is constructed by the optimal approximation kernel $\varphi(x, h_j)$, will most likely admit a Riesz basis generated from the translates of a 'wavelet' kernel $\psi(x, h_j)$. If such a step becomes successful, we could essentially generate a 'non-stationary' system of base functions in $L^2(\mathbf{R})$ that will be directly associated with the actual statistical collocation formula in Eq. (28); i.e. the optimal approximation of the unknown signal will give rise to a certain type of wavelet-like basis. The potential of such a connection is quite remarkable, in both theoretical and practical terms, and it will be explored in future publications.

6 Conclusions and future work

The aim of this paper was to show that the concept of multiresolution approximation theory lies at the very core of some of the general estimation principles involved in physical geodesy problems. This should not come as a surprise, since MRA methods were developed not as 'brand new' theories, but rather as a synthesis of various ideas which originated over the years from different disciplines including mathematics, physics and signal processing. The common link that physical geodesy shares with these areas is the archetypical problem of estimating an unknown field from its discrete values. It has been shown that the method of statistical collocation, as expressed by the optimal criterion in Eq. (21) and the classic translation-invariance condition, leads to signal approximation models similar to the ones encountered in Mallat's MRA theory. It is the opinion of the author that Sanso's spatio-statistical formulation for the collocation problem (Sanso 1980) should not be viewed only as a 'supplement' to Wiener's stochastic prediction theory for geodetic approximation problems. It actually constitutes a very powerful and autonomous modeling tool, with remarkable connections to multiresolution approximation theory and wavelets. One of the advantages of this link is that it provides basically the means to develop a useful generalization of Mallat's classic MRA scheme, where the data resolution level is not restricted to only dyadic values. Of course, much more theoretical work is needed to establish the existence of wavelet-like bases within the generalized MRA structure that was developed in this paper.

Among the benefits of the present multiresolution formulation for the collocation problem is the easiness with which the *non-stationarity* issue is overcome. Having its roots in the very much debated stochastic/non-stochastic interpretation of the gravity field, this problem has been 'amplified' over the years by the domination of the classic Fourier-based spectral techniques in gravity field modeling. Furthermore, the original spatio-statistical treatment of collocation according to Sanso (1980), where a 'stationary' spatial CV function is needed, has created the false belief among many geodesists that we still need to model the gravity field as a stationary ergodic process. Although stationarity is a stochastic term that cannot be theoretically justified in the present deterministic setting, I personally perceive this problem (in the context of optimal linear estimation in deterministic fields) as the ability to reconstruct and to study *locally* the unknown field in a

rigorous and consistent manner with the estimation principles. With such an understanding of the problem, MRA and wavelet theory can provide valuable tools without deviating from the universally acceptable collocation spirit (i.e. MMSE principle).

Many additional theoretical extensions of the issues discussed herein are needed in order to cover the spectrum of all possible applications in gravity field modeling. First, and most important, is the inclusion of the observational (non-stationary in general) noise in the multiresolution approximation framework, and the development of optimal noise filtering methods in multiresolution estimation models. In the noiseless case, efficient algorithms for studying the behavior of the signal error caused by the basic formula in Eq. (28), as a function of the data resolution level h and the used kernel $\varphi(x)$, will be quite useful especially for simulation studies with synthetic fields. The case where the available data grids include not only sampled values of the unknown field, but other linear functionals as well, should be also carefully treated. Finally, the extension of all the above in two and three dimensions (including compact spherical domains) should be made.

Practical and computational issues have not been discussed in this paper. Of special importance for our purposes would be the development of efficient (frequency domain) algorithms for the computation of the optimal scaling kernel in Eq. (31a) from the signal power spectrum, at various data resolution levels h_j. Also, the possibility to model empirically not the CV function (or the power spectrum) of the unknown field, but rather the optimal approximation kernel itself, should be explored. In the same line of thought, a reverse approach which would compute the 'induced' signal power spectrum from the analytical expression of already available scaling kernels [i.e. inversion of Eq. (31a)] might help to identify which types of kernels seem more realistic for approximating the actual behavior of the gravity field at various resolution levels. All these practical algorithmic issues are extremely important and rather complicated (especially for higher dimensions). It is interesting, finally, to note the computational efficiency that is achieved by using Eq. (28) over the classic matrix formula of collocation; see also Svensson (1983). Once the approximation kernel has been selected, Eq. (28) corresponds to a single multiplication between two n-dimensional vectors, whereas the matrix equation that utilizes the CV function of the unknown field would require $n+1$ multiplications of n-dimensional vectors plus an $n \times n$ symmetric matrix inversion, where n is the number of the available data points.

Acknowledgements. The author would like to thank Prof. M.G. Sideris, who provided valuable comments during the preparation of this paper. Profs. W. Keller, R. Lehmann, and an anonymous reviewer are also acknowledged for their very useful criticism of an early version of the paper. Financial support for this research has been provided by NSERC grants of Profs. M.G. Sideris and G. Lachapelle.

Appendix A

In this appendix we will prove the following equation:

$$\int_{-h/2}^{h/2} |E(\omega, x_o)|^2 \, \mathrm{d}x_o = hC(\omega) - \Phi_h^*(\omega)C(\omega) - \Phi_h(\omega)C(\omega) + \Phi_h(\omega)\Phi_h^*(\omega)\bar{C}_h(\omega) \tag{A1}$$

Taking into account Eq. (20b), the error power spectrum at an arbitrary value of the sampling phase parameter x_o has the form

$$\begin{aligned} |E(\omega, x_o)|^2 = {} & G(\omega)G^*(\omega) - \Phi_h^*(\omega)G(\omega)S^*(\omega) \\ & - \Phi_h(\omega)G^*(\omega)S(\omega) \\ & + \Phi_h(\omega)\Phi_h^*(\omega)S(\omega)S^*(\omega) \end{aligned} \tag{A2}$$

where the auxiliary function $S(\omega)$ is given by

$$S(\omega) = \frac{1}{h}\sum_k G\left(\omega + \frac{2\pi k}{h}\right) \mathrm{e}^{-i\frac{2\pi k}{h}x_o} \tag{A3}$$

Integrating Eq. (A2) over x_o, we obtain analytically for every term

$$\int_{-h/2}^{h/2} G(\omega)G^*(\omega)\mathrm{d}x_o = G(\omega)G^*(\omega)h = hC(\omega) \tag{A4}$$

$$\begin{aligned} & \int_{-h/2}^{h/2} \Phi_h^*(\omega)G(\omega)S^*(\omega)\mathrm{d}x_o \\ & = \Phi_h^*(\omega)G(\omega) \int_{-h/2}^{h/2} \frac{1}{h}\sum_k G^*\left(\omega + \frac{2\pi k}{h}\right) \mathrm{e}^{i\frac{2\pi k}{h}x_o} \, \mathrm{d}x_o \\ & = \frac{1}{h}\Phi_h^*(\omega)G(\omega) \sum_k G^*\left(\omega + \frac{2\pi k}{h}\right) \int_{-h/2}^{h/2} \mathrm{e}^{i\frac{2\pi k}{h}x_o} \, \mathrm{d}x_o \\ & = \frac{1}{h}\Phi_h^*(\omega)G(\omega) \sum_k G^*\left(\omega + \frac{2\pi k}{h}\right) \int_{-\pi}^{\pi} \frac{h}{2\pi}\mathrm{e}^{ik\xi} \, \mathrm{d}\xi \\ & = \Phi_h^*(\omega)G(\omega) \sum_k G^*\left(\omega + \frac{2\pi k}{h}\right) \frac{\sin k\pi}{k\pi} \\ & = \Phi_h^*(\omega)G(\omega)G^*(\omega) = \Phi_h^*(\omega)C(\omega) \end{aligned} \tag{A5}$$

Following similar derivations as in Eq. (A5), we obtain

$$\int_{-h/2}^{h/2} \Phi_h(\omega)G^*(\omega)S(\omega)\mathrm{d}x_o = \Phi_h(\omega)C(\omega) \tag{A6}$$

Finally, the integration of the last term in Eq. (A2) yields

$$\int_{-h/2}^{h/2} \Phi_h(\omega)\Phi_h^*(\omega)S(\omega)S^*(\omega)\,\mathrm{d}x_o$$

$$= \Phi_h(\omega)\Phi_h^*(\omega) \int_{-h/2}^{h/2} \frac{1}{h^2}\sum_n\sum_m G\left(\omega+\frac{2\pi n}{h}\right) \times G^*\left(\omega+\frac{2\pi m}{h}\right)\mathrm{e}^{i\frac{2\pi(m-n)}{h}x_o}\,\mathrm{d}x_o$$

$$= \frac{1}{h^2}\Phi_h(\omega)\Phi_h^*(\omega)\sum_n\sum_m G\left(\omega+\frac{2\pi n}{h}\right) \times G^*\left(\omega+\frac{2\pi m}{h}\right)\int_{-h/2}^{h/2}\mathrm{e}^{i\frac{2\pi(m-n)}{h}x_o}\,\mathrm{d}x_o$$

$$= \frac{1}{h^2}\Phi_h(\omega)\Phi_h^*(\omega)\sum_n\sum_m G\left(\omega+\frac{2\pi n}{h}\right) \times G^*\left(\omega+\frac{2\pi m}{h}\right)\int_{-\pi}^{\pi}\frac{h}{2\pi}\mathrm{e}^{i(m-n)\xi}\,\mathrm{d}\xi$$

$$= \frac{1}{h}\Phi_h(\omega)\Phi_h^*(\omega)\sum_n\sum_m G\left(\omega+\frac{2\pi n}{h}\right) \times G^*\left(\omega+\frac{2\pi m}{h}\right)\frac{\sin\pi(m-n)}{\pi(m-n)}$$

$$= \frac{1}{h}\Phi_h(\omega)\Phi_h^*(\omega)\sum_k G\left(\omega+\frac{2\pi k}{h}\right)G^*\left(\omega+\frac{2\pi k}{h}\right)$$

$$= \frac{1}{h}\Phi_h(\omega)\Phi_h^*(\omega)\sum_k C\left(\omega+\frac{2\pi k}{h}\right)$$

$$= \Phi_h(\omega)\Phi_h^*(\omega)\bar{C}_h(\omega) \tag{A7}$$

Combining together the results from Eqs. (A4) through (A7), we obtain the initially claimed statement of Eq. (A1).

Appendix B

In this appendix we will prove that the multiresolution subspace sequence $\{V_j\}$, which is constructed through the optimal approximation kernel $\varphi(x, h_j)$, has the basic nesting MRA property, i.e.

$$V_j \subset V_{j+1}, \quad \forall\, j \in \mathbf{Z} \tag{B1}$$

Each element $V_j \subset L^2(\mathbf{R})$ of this subspace sequence is defined as the closed linear span of the set

$$\left\{\varphi\left(\frac{x}{h_j}-n,\, h_j\right)\Big|n \in \mathbf{Z}\right\}$$

where the kernel $\varphi(x, h_j)$ is defined by Eq. (31a), and the scaling parameter h_j associated with each subspace V_j is assumed to satisfy the two general conditions given in Eqs. (41) and (43). Furthermore, the power spectrum and the CV function [through which the optimal approximation kernal $\varphi(x, h_j)$ is defined] of the unknown signal are assumed to satisfy all the conditions given in Sect. 5.1.

Every signal $f_j(x) \in V_j$ will have the general form

$$f_j(x) = \sum_n b_n \varphi\left(\frac{x}{h_j}-n,\, h_j\right), \quad \forall\, f_j(x) \in V_j \tag{B2}$$

where $\{b_n\}$ is a certain square-summable sequence of coefficients. Taking into account Eq. (31b), the last equation can be equivalently expressed in the frequency domain as follows:

$$F_j(\omega) = h_j\frac{C(\omega)}{\sum_k C\left(\omega+\frac{2\pi k}{h_j}\right)}\sum_n b_n \mathrm{e}^{-i\omega n h_j} = h_j\frac{C(\omega)}{\sum_k C\left(\omega+\frac{2\pi k}{h_j}\right)}\bar{B}_{2\pi/h_j}(\omega), \quad \forall\, f_j(x) \in V_j \tag{B3}$$

where $\bar{B}_{2\pi/h_j}(\omega)$ denotes a certain $(2\pi/h_j)$-periodic function with finite $L^2(0, 2\pi/h_j)$ norm. In the same way, every signal $f_{j+1}(x)$ that belongs in the subspace V_{j+1} will have the following frequency domain form:

$$F_{j+1}(\omega) = h_{j+1}\frac{C(\omega)}{\sum_k C\left(\omega+\frac{2\pi k}{h_{j+1}}\right)}\bar{B}_{2\pi/h_{j+1}}(\omega), \quad \forall\, f_{j+1}(x) \in V_{j+1} \tag{B4}$$

where h_{j+1} is the scaling parameter associated with V_{j+1}, and $\bar{B}_{2\pi/h_{j+1}}(\omega)$ denotes a certain $(2\pi/h_{j+1})$-periodic function with finite $L^2(0, 2\pi/h_{j+1})$ norm. It is now quite easy to transform Eq. (B3) in the form of Eq. (B4). Indeed, starting from Eq. (B3) we will have

$$F_j(\omega)$$

$$= h_j\frac{C(\omega)}{\sum_k C\left(\omega+\frac{2\pi k}{h_j}\right)}\frac{h_{j+1}\sum_k C\left(\omega+\frac{2\pi k}{h_{j+1}}\right)}{h_{j+1}\sum_k C\left(\omega+\frac{2\pi k}{h_{j+1}}\right)}\bar{B}_{2\pi/h_j}(\omega)$$

$$= h_{j+1}\frac{C(\omega)}{\sum_k C\left(\omega+\frac{2\pi k}{h_{j+1}}\right)}\frac{h_j\sum_k C\left(\omega+\frac{2\pi k}{h_{j+1}}\right)}{h_{j+1}\sum_k C\left(\omega+\frac{2\pi k}{h_j}\right)}\bar{B}_{2\pi/h_j}(\omega)$$

$$= h_{j+1}\frac{C(\omega)}{\sum_k C\left(\omega+\frac{2\pi k}{h_{j+1}}\right)}\bar{\Lambda}_{2\pi/h_{j+1}}(\omega)\bar{B}_{2\pi/h_j}(\omega)$$

$$= h_{j+1}\frac{C(\omega)}{\sum_k C\left(\omega+\frac{2\pi k}{h_{j+1}}\right)}\bar{N}_{2\pi/h_{j+1}}(\omega), \ \forall\, f_j(x) \in V_j \tag{B5}$$

where the auxiliary function $\bar{\Lambda}_{2\pi/h_{j+1}}(\omega)$ is defined by the formula

$$\bar{\Lambda}_{2\pi/h_{j+1}}(\omega) = \frac{h_j\sum_k C\left(\omega+\frac{2\pi k}{h_{j+1}}\right)}{h_{j+1}\sum_k C\left(\omega+\frac{2\pi k}{h_j}\right)} \tag{B6}$$

Obviously, the above function will be $(2\pi/h_{j+1})$-periodic, since the two scaling parameters (h_j and h_{j+1}) are

assumed to be related through a positive *integer* number, according to condition VI in Eq. (43). For the same reason, the product of the two periodic functions $\bar{\Lambda}_{2\pi/h_{j+1}}(\omega)$ and $\bar{B}_{2\pi/h_j}(\omega)$, which is denoted $\bar{N}_{2\pi/h_{j+1}}(\omega)$ in Eq. (B5), will also be a $(2\pi/h_{j+1})$-periodic function.

Furthermore, under conditions I and II both the numerator and denominator in Eq. (B6) will converge uniformly to finite, strictly-positive, continuous periodic functions, for any pair of values of the scaling parameters h_j and h_{j+1}. Hence, the periodic function $\bar{\Lambda}_{2\pi/h_{j+1}}(\omega)$ will certainly have a finite $L^2(0, 2\pi/h_{j+1})$ norm. As a result, the auxiliary periodic function $\bar{N}_{2\pi/h_{j+1}}(\omega)$ in Eq. (B5) will also have a finite $L^2(0, 2\pi/h_{j+1})$ norm. In this way, the final frequency domain form of Eq. (B5) corresponds exactly to the expression of a function belonging in the higher resolution subspace V_{j+1}, according to the general formula (B4).

References

Aldroubi A, Unser M (1992) Families of wavelet transforms in connection with Shannon's sampling theory and the Gabor transform. In: Chui CK (ed) Wavelets: a tutorial in theory and applications. Academic Press, New York, pp 509–528

Aldroubi A, Unser M (1994) Sampling procedures in function spaces and asymptotic equivalence with Shannon's sampling theory. Numer Funct Anal Optim 15(1, 2): 1–21

Ballani L (1996) Solving the inverse gravimetric problem: on the benefit of wavelets. IAG Symp Proc, vol 114. Springer Berlin Heidelberg New York, pp 151–161

Bendat JS, Piersol AG (1986) Random data: analysis and measurement procedures, 2nd edn. John Wiley, New York

Blu T, Unser M (1999) Quantitative Fourier analysis of approximation techniques: part I (interpolators and projectors) and part II (wavelets). IEEE Trans Signal Process 47(10): 2783–2806

Freeden W, Schneider F (1998) An integrated wavelet concept of physical geodesy. J Geod 72: 259–281

Halmos PR (1991) Measure theory. Graduate texts in mathematics series, vol 18. Springer, Berlin Heidelberg New York

Heil CE, Walnut DF (1994) Continuous and discrete wavelet transforms. Siam Rev 31(4): 628–666

Holschneider M (1995) Wavelets, an analysis tool. Oxford University Press, New York

Jawerth B, Sweldens W (1994) An overview of wavelet based multiresolution analyses. Siam Rev 36(3): 377–412

Keller W (1998) Geoid computation by collocation in scaling spaces. IAG Symp Proc, vol 119. Springer, Berlin Heidelberg New York, pp 176–182

Kumar P, Foufoula-Georgiou E (1997) Wavelet analysis for geophysical applications. Rev Geophys 34(5): 385–412

Li Z (1996a) Multiresolution approximation of the gravity field. J Geod 70: 731–739

Li Z (1996b) Multiresolution approximation in gravity field modelling. UCGE Rep 20103, Department of Geomatics Engineering, University of Calgary

Mallat SG (1989a) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11(7): 674–693

Mallat SG (1989b) Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbf{R})$. Trans Am Math Soc 315(1): 69–87

Mallat SG (1998) A wavelet tour of signal processing. Academic Press, New York

Moritz H (1976) Least-squares collocation as a gravitational inverse problem. Rep 249, Department of Geodetic Science and Surveying, The Ohio State University, Columbus

Moritz H (1980) Advanced physical geodesy. Herbert Wichmann, Karlsruhe

Moritz H, Sanso F (1980) A dialogue on collocation. Boll Geod Sci Affi 39(1): 49–51

Nashed MZ, Walter GG (1991) General sampling theorems for functions in reproducing kernel Hilbert spaces. Math Contr Signals Syst 4: 363–390

Oppenheim AV, Schafer RW (1989) Discrete-time signal processing. Prentice-Hall, Englewood Cliffs

Sanso F (1980) The minimum mean square estimation error principle in physical geodesy (stochastic and non-stochastic interpretation). Boll Geod Sci Affi 39(2): 112–129

Sanso F (1986) Statistical methods in physical geodesy. In: Sunkel H (ed) Mathematical and numerical techniques in physical geodesy. Springer, Berlin Heidelberg New York, pp 49–156

Sanso F (1987) Talk on the theoretical foundations of physical geodesy. Paper presented at the XIX IUGG General Assembly, Vancouver, 9–22 August

Schmidt HF (1981) Sampling function and finite element method representation of the gravity field. Rev Geophys 19(3): 421–436

Schwarz KP (1984) Data types and their spectral properties. Proc Int Summer School Local Gravity Field Approximation, Beijing, 21 August–4 September, pp 1–67

Sideris MG (1995) On the use of heterogeneous noisy data in spectral gravity field modelling methods. J Geod 70: 470–479

Svensson SL (1983) A new geodesy based upon inversion-free Bjerhammar predictors. Proc 8th Symp Mathematical Geodesy, Como, 7–9 September 1981, pp 449–468

Tscherning CC (1986) Functional methods for gravity field approximation. In: Sunkel H (ed) Mathematical and numerical techniques in physical geodesy. Springer, Berlin Heidelberg New York, pp 3–48

Unser M, Daubechies I (1997) On the approximation power of convolution-based least squares versus interpolation. IEEE Trans Signal Process 47(7): 1697–1711

Walter GG (1992) A sampling theorem for wavelet subspaces. IEEE Trans Info Theory 38(2): 881–884

Wojtaszczyk P (1997) A mathematical introduction to wavelets. London Mathematical Society Student Texts 37. Cambridge University Press, Cambridge

Xia X-G, Zhang Z (1993) On sampling theorem, wavelets, and wavelet transforms. IEEE Trans Signal Process 41(12): 3524–3535

Young RM (1980) An introduction to non-harmonic Fourier series. Academic Press, New York

Zayed AI (1993) Advances in Shannon's sampling theory. CRC Press, Boca Raton