CHRISTOPHER KOTSAKIS

☆

The role of statistical collocation as a tool for constructing generalized multiresolution analyses in the $L^2(\Re)$ Hilbert space

The role of statistical collocation as a tool for constructing generalized multiresolution analyses in the $L^2(\Re)$ Hilbert space

CHRISTOPHER KOTSAKIS

Department of Geomatics Engineering, University of Calgary, Calgary, Alberta, Canada

Summary. – An interesting connection between the statistical collocation method and the multi-resolution/wavelet framework of signal approximation is made in this study. The rapid developments in multi-resolution analysis theory over the past few years have provided very useful (theoretical and practical) tools for approximation and spectral studies of irregularly varying signals, opening thus new possibilities for «non-stationary» gravity field modelling. In this paper we demonstrate that the classic multiresolution formalism according to Mallat's pioneering work lies at the very core of some of the approximation principles traditionally used in physical geodesy problems. In particular, it will be shown that the use of a spatio-statistical (non-probabilistic) minimum mean square error criterion, for optimal linear estimation of deterministic signals, always gives rise to a generalized multiresolution analysis in the Hilbert space $L^2(\Re)$, under some mild constraints on the spatial covariance function and the power spectrum of the unknown field under consideration. Using the theory and the approximation algorithms associated with statistical collocation, a new constructive (frequency domain based) framework for building generalized multiresolution analyses in $L^2(\Re)$ is presented, without the need of 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.

IL RUOLO DELLA COLLOCAZIONE STATISTICA COME STRUMENTO PER SVILUPPARE ANALISI GENERA-LIZZATE MULTIRISOLUZIONE NELLO SPAZIO DI HILBERT $L^2(\Re)$.

Sommario. – In questo studio si sviluppa un'interessante collegamento tra il metodo della collocazione statistica e l'apparato multirisoluzione/*wavelet* dell'approssimazione del segnale. I rapidi sviluppi nella teoria dell'analisi multirisoluzione negli ultimi anni hanno reso disponibili utilissimi strumenti (sia teorici che pratici) per l'approssimazione e lo studio dello spettro di segnali irregolarmente variabili, aprendo così nuove possibilità nella modellizzazione del campo di gravità «non-stazionario». In questo lavoro si dimostra che il classico

formalismo multirisoluzione secondo il pionieristico contributo di Mallat, ha un ruolo centrale in alcuni dei principi di approssimazione utilizzati tradizionalmente nei problemi di geodesia fisica. In particolare, si dimostra che l'adozione di un criterio di minimo errore quadratico medio spaziale-statistico (non-probabilistico), per la stima lineare ottimale di segnali deterministici, consente sempre un'analisi generalizzata multirisoluzione nello spazio di Hilbert $L^2(\Re)$ sotto alcuni (non troppo stringenti) vincoli relativi alla funzione spaziale di covarianza e allo spettro di intensità del campo incognito in considerazione. Applicando la teoria e gli algoritmi di approssimazione associati alla collocazione statistica, viene qui presentato un nuovo apparato costruttivo (basato sul dominio di frequenza) per la realizzazione di analisi generalizzate multirisoluzione in $L^2(\Re)$, senza la necessità dell'usuale *ipotesi restrittiva diadica* che sussiste nella classica teoria delle *wavelet*. Vengono inoltre discussi gli aspetti di multirisoluzione e «non-stazionari» della procedura di approssimazione della collocazione statistica e, infine, vengono presentate alcune conclusioni e raccomandazioni per il lavoro futuro.

1. – INTRODUCTION

The concept of least-squares collocation (LSC) represents one of the major theoretical (and practical) foundations in modern physical geodesy. Closely related to Bjerhammar's initial idea on discrete underdetermined boundary value problems, collocation has evolved into a powerful optimal estimation method for either global or local gravity field modelling. Despite the various different interpretations and their associated mathematical concepts upon which LSC has been based (see, e.g., Tscherning, 1986; Sansò, 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 has long been established by Sansò (1980). Such an approach has eliminated, to some degree, most of the «pitfalls» in each individual original formulation (e.g. reproducing kernel or norm choice problem, non-stochasticity of the actual gravity field); see also Moritz (1980), Moritz and Sansò (1980). In this way, collocation is usually considered as a rigorous linear *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.

One of the main characteristics of Sansò's spatio-statistical formulation for the collocation problem is that it leads to the same solution algorithm as the purely deterministic/stochastic approaches. In this case, however, instead of using a reproducing kernel in some arbitrary Hilbert space or a covariance (CV) function of a stochastic signal, we only need a *spatial CV function* defined through a certain spatial averaging operator over the unknown deterministic signal. The «stationary» form of this spatial CV function has created the false belief among many geodesists that we still need to model the gravity field of the Earth as a stationary stochastic process, which is furthermore perceived as a strong limitation of the statistical collocation framework since the actual behavior of the gravity field is «non-stationary». However, such a claim is absolutely meaningless because no stochastic nature is assigned to the unknown field, and the property of stationarity is not defined at all for deterministic signals; see also the related discussion given in Sansò (1980).

In order to eliminate any stationarity concerns about the statistical collocation framework, and to additionally encourage the transition towards the use of wavelet-based approximation techniques in gravity field modelling, the aim of this paper is to show that the spatio-statistical *mean square error criterion* (for optimal linear estimation in deterministic fields), in conjunction with regularly gridded data, is directly related to certain sequences of nested multiresolution subspaces in $L^2(\Re)$, similar to the ones encountered in the classic (dyadic) multiresolution approximation theory which was originally developed by Mallat (1989a, b). Wavelet bases and multiresolution signal analysis are two strongly related and relatively new mathematical concepts, that have been developed at an explosive rate during the last years in both theoretical and practical sense. Their constantly increasing popularity within various scientific areas is due to their ideal ability to analyze *locally* many transient («non-stationary») physical phenomena, and to approximate them at different (spatial/time) scale levels according to a *zoom-in/zoom-out* approach. Using the theory and the actual approximation algorithms associated with statistical collocation, the paper will present a new constructive (frequency domain based) framework for building generalized multiresolution analyses (and thus wavelet-type bases) in the Hilbert space $L^2(\Re)$, without the need of the usual *dyadic restriction* that exists in classic wavelet theory.

2. – MULTIRESOLUTION APPROXIMATION 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, have 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 these 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 brief overview from these two vast mathematical subjects, restricting ourselves to what is only necessary in order to follow the discussion in the following sections. A fully comprehensive reference including some of the most recent developments, can be found 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 (1996), and Keller (1998), among many others.

2.1. - Multiresolution analysis

A *multiresolution analysis* (MRA) in the Hilbert space $L^2(\Re)$ is defined as an infinite sequence of closed linear Hilbert subspaces $V_j \subset L^2(\Re)$, having the following five properties (Jawerth and Sweldens, 1994):

a)
$$V_j \subset V_{j+1}, \quad \forall j \in Z \tag{1.a}$$

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

c)
$$f(x) \in V_j \Leftrightarrow f(x+n2^{-j}) \in V_j, \quad \forall n \in Z \tag{1.c}$$

d)
$$\overline{\bigcup_{j=-\infty}^{+\infty} V_j} = L^2(\Re) \quad \text{and} \quad \bigcap_{j=-\infty}^{+\infty} V_j = \{0\} \tag{1.d}$$

e) A *scaling function* $\varphi(x) \in V_0$, with a non-vanishing integral, exists such that the family $\varphi(x - n)_{n \in Z}$ is a Riesz basis of V_0.

The definition given above is not minimal, in the sense that some of the conditions *a)* - *e)* can be derived from the remaining ones (Wojtaszczyk, 1997). However, it has been customary to use all these 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 $\varphi_n(x)_{n \in Z}$ in a Hilbert space H, then there always exists a unique *biorthonormal* system $\tilde{\varphi}_n(x)_{n \in Z}$ which also forms a Riesz basis for H. The biorthonormality property between the two systems can be expressed through the relation

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

where $<,>$ 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 Heil and Walnut (1994), Young (1980), and Wojtaszczyk (1997).

If a Riesz basis (under the usual L^2 inner product) is formed by the translates $\varphi(x - n)_{n \in Z}$ 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):

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

for some strictly positive, finite bounds A and B. If equation (3) is true, then the sets $\varphi(x/h - n)_{n \in Z}$ form Riesz bases for their corresponding closed linear spans, for any

non-zero value of the scaling parameter h (Unser and Daubechies, 1997). In this way, the translates of the scaling function $\varphi(2^j x - n)_{n \in Z}$ will also form a Riesz basis in each corresponding subspace V_j of an MRA. Condition (3), for the special case where the family $\varphi(x - n)_{n \in Z}$ is an *orthonormal* set, takes the simple form

$$0 < A \leq \sum_k \left| \Phi(\omega + 2\pi k) \right|^2 \leq B < +\infty. \tag{4}$$

In every nested subspace V_j of an MRA, an *infinite number* of complete orthonormal sets can be constructed from a given Riesz basis $\varphi(2^j x - n)_{n \in Z}$, according to the orthonormalization trick given in Young (1980, p. 48); see also Holschneider (1995, p. 187), Wojtaszczyk (1997, pp. 24-25). These orthonormal sets will, too, be comprised of integer translates of a basic function. For the same MRA in $L^2(\Re)$, therefore, we can have many different choices for the generating scaling function.

Let us now 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 reproducing kernel (r.k) 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 r.k of the «unit» resolution subspace V_0. For some technical mathematical conditions, see the paper by Walter (1992). It can also be shown that the collection of functions $k(x, n)_{n \in Z} = k(x - n, 0)_{n \in Z}$ provides an alternative Riesz basis for the MRA subspace V_0 (Walter, 1992). The biorthonormal basis correspon- ding to $k(x, n)_{n \in Z}$ has some very special properties, namely being a *sampling basis* for the same subspace V_0. The expansion of an arbitrary signal $f(x) \in V_0$, with respect to such a basis, takes the form of a *sampling theorem* associated with the specific subspace V_0, i.e.

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

The situation can easily be extended for any MRA subspace V_j. More details for the connection between sampling theorems and L^2 multiresolution theory can be found in Walter (1992), Zayed (1993), Xia and Zhang (1993), Aldroubi and Unser (1992, 1994). An 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. - MULTIRESOLUTION APPROXIMATION

The original definition of an MRA, according to Mallat (1989.b), differs slightly from the one given in eq. (1). In Mallat's conception, instead of intro-ducing 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. These projectors are used to determine the best linear L^2 approximation of an arbitrary signal $f(x) \in L^2(\Re)$ 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 (1989.b).

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

$$(P_j f)(x) = \sum_n a(n) \, \varphi(2^j x - n) \qquad (8.a)$$

$$a(n) = 2^j \int f(x) \, \varphi(2^j x - n) \, dx. \qquad (8.b)$$

The projection procedure is illustrated in fig. 1, where the first filter $\Phi(-2^{-j}\omega)$ from eq. (8.b) 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 frame-work.
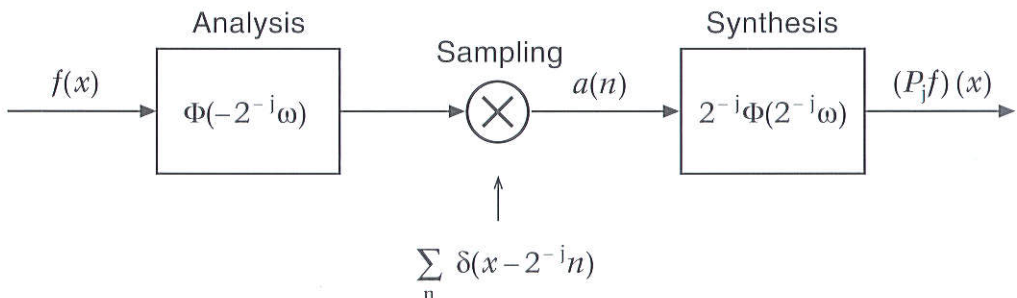

Fig. 1 – Orthogonal projection onto a subspace V_j of an MRA (use of orthonormal scaling function).

2.3. - WAVELETS

Associated with every MRA in $L^2(\mathfrak{R})$ is a corresponding *wavelet basis*, which provides the means to connect consistently and efficiently signal information from different 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 $\psi(2^j x - n)_{n \in Z}$ will provide a Riesz basis for every W_j. The Hilbert subspace $W_j \subset L^2(\mathfrak{R})$ contains basically the signal «details» needed to go from the dyadic resolution level 2^j to the next upper resolution 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 $L^2(\mathfrak{R})$; see, e.g., Jawerth and Sweldens (1994).

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

$$f_k(x) = \sum_n a(n) \, \varphi(2^k x - n) = \sum_{j=-\infty}^{k-1} \sum_n b(n, j) \, \psi(2^j x - n), \quad \forall f_k \in V_k. \tag{9}$$

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 the classic 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. – STATISTICAL COLLOCATION AND DATA RESOLUTION

In this section, the optimal linear approximation problem for an unknown *deterministic* field $g \in L^2(\mathfrak{R})$ 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 will be seen to depend only on a single basic kernel $\varphi \in L^2(\mathfrak{R})$, 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 is considered as 1D for simplicity. The multi-dimensional case (i.e. when the unknown field belongs in the $L^2(\mathfrak{R}^2)$ or the $L^2(\mathfrak{R}^3)$ 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 spatial reference system (in the multi-dimensional case this becomes invariance under more general affine 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 approximation formula (12) can now be illustrated in terms of the linear filtering procedure shown in fig. 2. Applying the Fourier transform to the above convolution equation, we get:

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

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 $\overline{G}_h(\omega)$ corresponds to the *periodic* Fourier transform of the generalized function

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

and it has the form (Oppenheim and Schafer, 1989)

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

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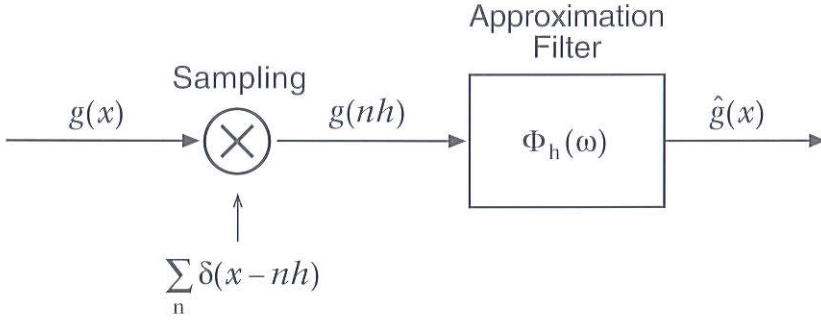

Fig. 2 – Filtering configuration of linear, translation-invariant signal approximation using discrete samples.

Note that the previous frequency domain formulas imply that we have sampled the unknown signal $g(x) \in L^2(\mathfrak{R})$ 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 $g(x)$ we try to approximate 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 given data grid boundaries, its approximation $\hat{g}(x)$ by eq. (12) may exhibit a non-zero pattern at points outside the data grid. Of course, the theoretical case of in-finitely 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:

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

Indeed, under such a condition the periodic Fourier transform $\overline{G}_h(\omega)$ in eq. (14.b) 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.

$$\left| E(\omega) \right|^2 = E(\omega)\, E^*(\omega) = G(\omega)\, G^*(\omega) - \Phi_h^*(\omega)\, G(\omega)\, \overline{G}_h^*(\omega) -$$
$$- \Phi_h(\omega)\, \overline{G}_h(\omega)\, G^*(\omega) + \Phi_h(\omega)\, \Phi_h^*(\omega)\, \overline{G}_h(\omega)\, \overline{G}_h^*(\omega) \tag{17}$$

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 $\sum_n \delta(x - nh)$ by an amount x_0, 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. The situation is illustrated in fig. 3, 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 $g(nh - x_0)$, where the *sampling phase* parameter x_0 varies between $-h/2 \le x_0 \le h/2$.
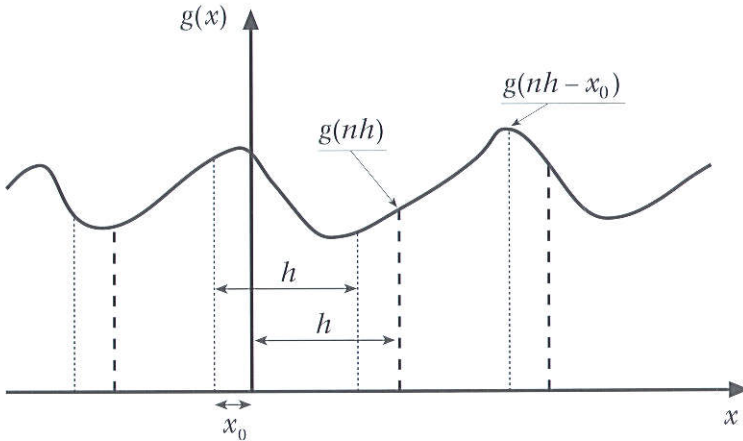

Fig. 3 – Different sampling configurations at a given resolution level h.

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

$$\hat{g}(x, x_0) = \sum_n g(nh - x_0)\, \varphi_h(x + x_0 - nh). \tag{18}$$

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

$$\hat{G}(\omega, x_0) = \frac{1}{h}\, \Phi_h(\omega) \sum_k G\!\left(\omega + \frac{2\pi k}{h}\right) e^{-i\frac{2\pi k}{h} x_0} \tag{19}$$

where it is again assumed that all possible sampled sequences $g(nh - x_0)$ of the unknown field are always measurable in the sense of eq. (15). Thus, for each different sampling phase value x_0 we will have a correspondingly *different* approximation error $e(x, x_0)$ i.e.

$$e(x, x_0) = g(x) - \hat{g}(x, x_0) \tag{20.a}$$

whose Fourier transform is

$$E(\omega, x_0) = G(\omega) - \frac{1}{h}\, \Phi_h(\omega) \sum_k G\!\left(\omega + \frac{2\pi k}{h}\right) e^{-i\frac{2\pi k}{h} x_0}. \tag{20.b}$$

The optimal criterion for choosing the best approximation kernel $\varphi_h(x)$ will be

$$P_e(\omega) = \frac{1}{h} \int_{-h/2}^{h/2} \left| E(\omega, x_0)\right|^2 dx_0 = \min. \tag{21}$$

Equation (21) represents a *minimum mean square error* (MMSE) principle, expressed in the frequency domain. The quantity $P_e(\omega)$ is nothing else 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 it has a rather *spatio-statistical* meaning. In other words, the optimization of the linear approximation algorithm does not employ the classic averaging expectation operator considering different «experiment repetitions», but it is based on the average error over all possible different sampling configurations for the *given* data resolution level h. In Appendix A it is proven that

$$\int_{-h/2}^{h/2} \left| E(\omega, x_0)\right|^2 dx_0 = hC(\omega) - \Phi_h^*(\omega)\, C(\omega) - \Phi_h(\omega)\, C(\omega) + \Phi_h(\omega)\, \Phi_h^*(\omega)\, \overline{C}_h(\omega) \tag{22}$$

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

$$c(x) = \int g(y)\, g(y + x)\, dy \quad \xrightarrow{\ \mathfrak{I}\ } \quad C(\omega) = G(\omega)\, G^*(\omega) = \left| G(\omega)\right|^2 \tag{23}$$

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 $\overline{C}_h(\omega)$ in eq. (22) denotes its following periodization (see Appendix A):

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

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

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

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). In this way, the corresponding optimal space domain kernel $\varphi_h(x)$ can be 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:

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

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 approximation formula for an unknown deterministic field $g(x)$ according to the MMSE principle (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 (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., Unser and Daubechies, 1997; Blu and Unser, 1999). In such cases, the selection of the basic approximation kernel $\varphi(x)$ is usually made a priori (e.g. sinc-based inter-polation, polynomial spline interpolation, etc.), and its performance is evaluated according to an assumed behavior for the unknown signal (e.g. bandlimitedness, 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 mean 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 approximation 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 assumed 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 approximation kernel $\varphi(x)$, defined by eq. (27), is a *cardinal* (sampling) function. This simply means that

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

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

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

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

A very 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. (14.a)]. In our case, the noise takes the form of the lost information due to the discretization of the original signal (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*, that has been used throughout this section, should be understood in a purely spatial deterministic sense [eq. (23)] and *not* in any stochastic context under some stationarity and ergodicity assumption. This is especially important in view of the so-called «stationarity restriction» problem which is believed to exist in the statistical collocation framework. Our present 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 Sansò (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. – THE MULTIRESOLUTION CHARACTER OF STATISTICAL COLLOCATION

The final result of section 3.2 is quite general and it did not involve any particular concepts from Mallat's multiresolution theory. It is a rather remarkable 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. However, there is a significant difference between the collocation model of eq. (28) and the classic wavelet-based approximation methodology, due to the fact that the optimal approximation 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 frequency domain form in eq. (27), is to characterize it as: *a) translation-invariant*; *b) scale-invariant*; and *c) 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. Let us briefly demonstrate the scale-invariance aspect of the collocation approximation 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 approximation formula (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.

The optimal kernel $\varphi(x)$ in the statistical collocation model of eq. (28) is appropriately *scaled* (shrinked 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, section 2).
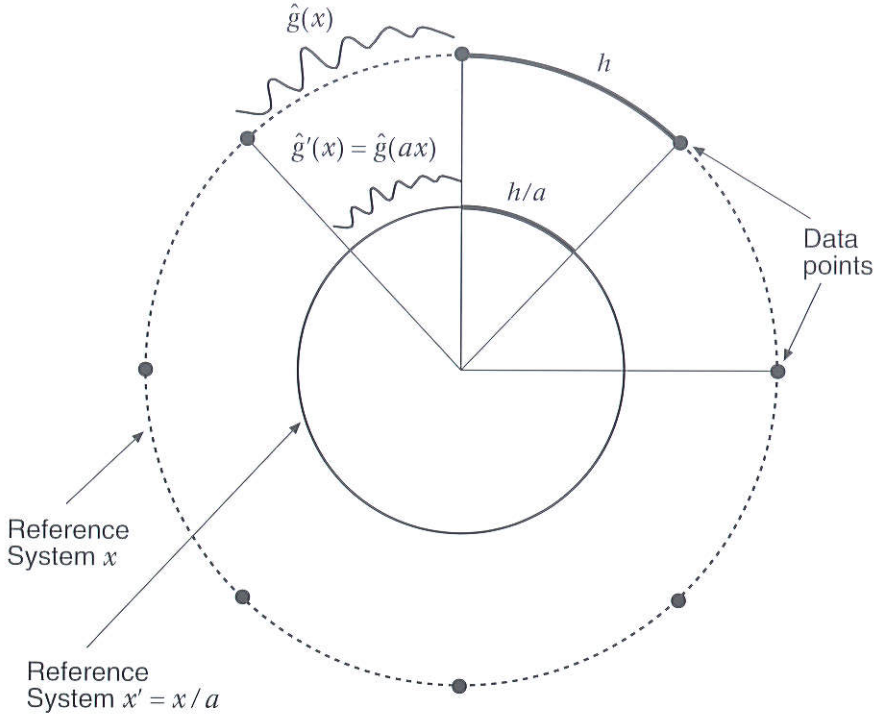
Fig. 4 – Scale-invariant signal approximation at a certain data resolution level h (the value of the scaling parameter a is assumed $a > 1$).

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)]:

$$\varphi(x, h) \xleftrightarrow{\ \mathfrak{I}\ } \Phi(\omega, h) = \frac{C\left(\dfrac{\omega}{h}\right)}{\displaystyle\sum_k C\left(\dfrac{\omega}{h} + \dfrac{2\pi k}{h}\right)} \qquad (31.\text{a})$$

where the data resolution h plays the role of a 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:

$$\varphi\left(\frac{x}{h}, h\right) \xleftrightarrow{\ \mathfrak{I}\ } h\Phi(h\omega, h) = h\,\frac{C(\omega)}{\displaystyle\sum_k C\left(\omega + \dfrac{2\pi k}{h}\right)} \qquad (31.\text{b})$$

which is identical with the Wiener-like approximation filter that was derived in section 3.2, eq. (25). For each different value of the data resolution parameter h, the basic optimal kernel in eq. (31.a) will assume a correspondingly different waveform, and hence the approximation 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 does it produce a scale-invariant signal approximation, but in addition it also «forces» the behavior of the basic approximation kernel to be adapted to the current data resolution in a certain optimal fashion, as suggested by eq. (31.a). 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 it was explained in section 3.3.

The varying behavior of the optimal approximation kernel $\varphi(x, h)$, for different data resolution levels h, is shown in figures 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. (31.a) for the case where the signal power spectrum has a Gaussian form, i.e.

$$C(\omega) = Be^{-\omega^2} \tag{32.a}$$

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:

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

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

These graphs help considerably to understand 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. (31.a) will asymptotically converge to a well defined $L^2(\mathfrak{R})$ cardinal kernel as $h \to 0$. All the individual members (i.e. functions) of this convergent sequence will be $L^2(\mathfrak{R})$ cardinal kernels as well. In the case of fig. 5, for example, it is obvious that the spatial expression of the optimal approximation kernel will gradually converge to the $sinc(x)$ 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 $L^2(\mathfrak{R})$ sense, as it is evident from the behavior of its Fourier transform in both figures 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.
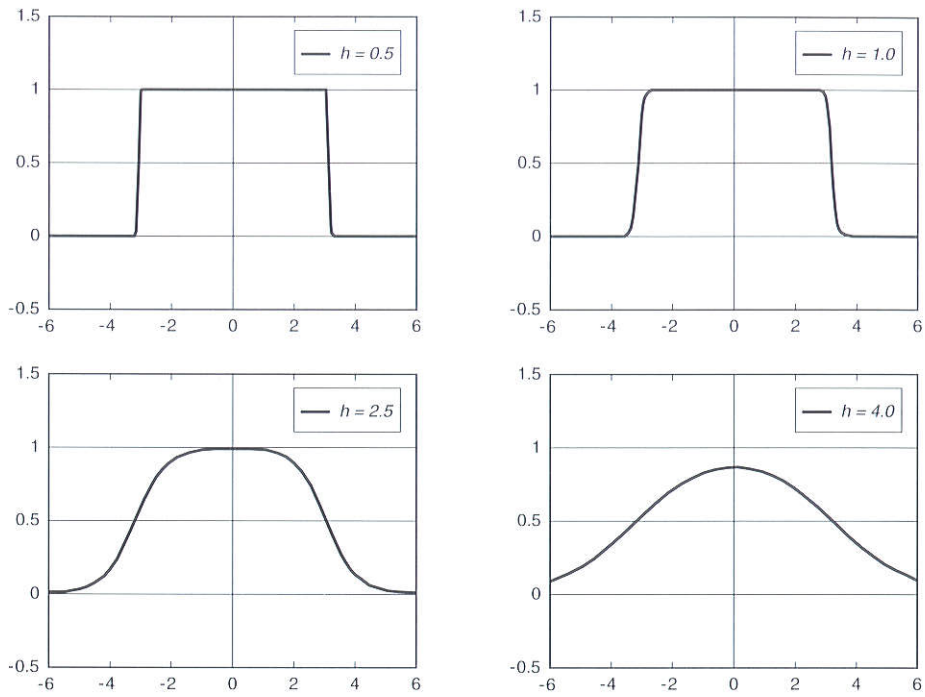
Fig. 5 – Fourier transform $\Phi(\omega, h)$ of the optimal approximation kernel $\varphi(x, h)$ for various data resolution levels h. The underlying unknown signal is assumed to follow a Gaussian power spectrum $C(\omega)$.
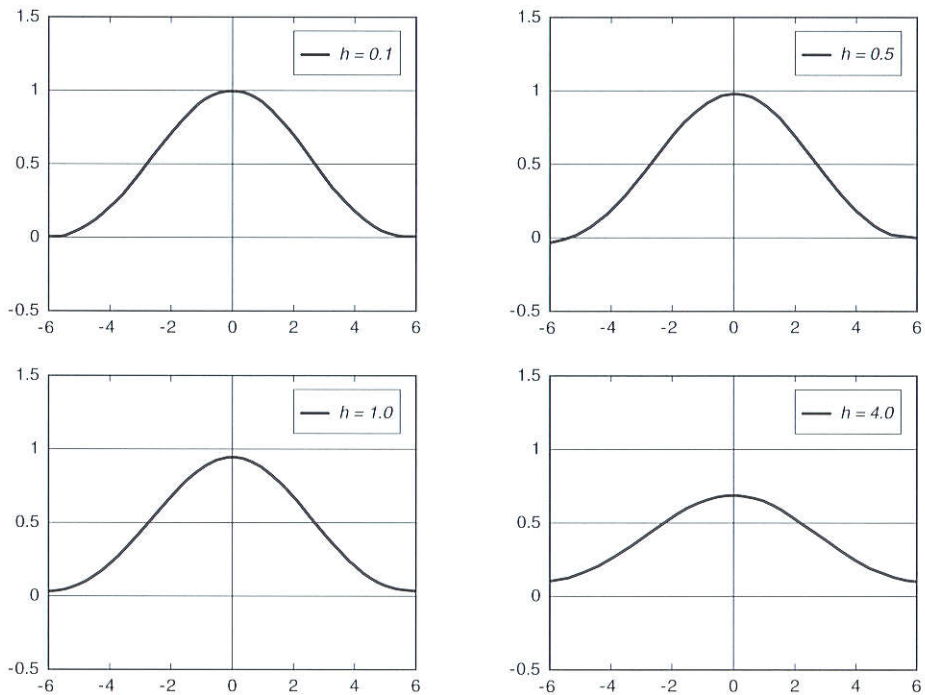

Fig. 6 – Fourier transform $\Phi(\omega, h)$ of the optimal approximation kernel $\varphi(x, h)$ for various data resolution levels h. The underlying unknown signal is assumed to follow the power spectrum model $C(\omega)$ given in eq. (32.b).

5. – GENERALIZED MULTIRESOLUTION ANALYSIS IN $L^2(\mathfrak{R})$

In this section we will explore in more detail the actual connection between the statistical collocation model of eqs. (27) and (28), and the classic multi-resolution/wavelet approximation framework which was briefly presented in section 2. We shall 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. (31.a) produces a generalized MRA-type approximation scheme in the Hilbert space $L^2(\mathfrak{R})$.

5.1. - MRA PROPERTIES OF THE OPTIMAL APPROXIMATION KERNEL

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

$$
\begin{aligned}
\left\| \varphi(x, h) \right\|_{L^2}^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \Phi(\omega, h) \right|^2 d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\left| C\left(\dfrac{\omega}{h} \right) \right|^2}{\left| \sum\limits_k C\left(\dfrac{\omega}{h} + \dfrac{2\pi k}{h} \right) \right|^2} d\omega \\
&= \frac{1}{2\pi} \int_{0}^{2\pi} \frac{\sum\limits_k \left| C\left(\dfrac{\omega + 2\pi k}{h} \right) \right|^2}{\left| \sum\limits_k C\left(\dfrac{\omega}{h} + \dfrac{2\pi k}{h} \right) \right|^2} d\omega = \frac{1}{2\pi} \int_{0}^{2\pi} \frac{\sum\limits_k \left(\dfrac{1}{h} C\left(\dfrac{\omega}{h} + \dfrac{2\pi k}{h} \right) \right)^2}{\left(\sum\limits_k \dfrac{1}{h} C\left(\dfrac{\omega}{h} + \dfrac{2\pi k}{h} \right) \right)^2} d\omega \quad (33) \\
&= \frac{1}{2\pi} \int_{0}^{2\pi} \overline{M}_{2\pi}(\omega) \, d\omega
\end{aligned}
$$

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

$$
\overline{M}_{2\pi}(\omega) = \frac{\sum\limits_k \left(\dfrac{1}{h} C\left(\dfrac{\omega}{h} + \dfrac{2\pi k}{h} \right) \right)^2}{\left(\sum\limits_k \dfrac{1}{h} C\left(\dfrac{\omega}{h} + \dfrac{2\pi k}{h} \right) \right)^2} = \frac{\sum\limits_k a_k^2}{\left(\sum\limits_k a_k \right)^2} \quad (34.a)
$$

and the discrete sequence a_k has of course the general form

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

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 $L^1(\Re)$ Hilbert space (i.e. since the unknown signal $g(x)$ is assumed to belong in the $L^2(\Re)$ Hilbert space). Obviously, the infinite series $\sum_k a_k$ corresponds to the 2π-*periodic* Fourier transform of a space domain sequence $b[n]$ constructed from the 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 $c(x)$:

Condition 1:

$$\sum_n \left| c(nh) \right| < \infty, \quad \forall h > 0. \tag{36}$$

Note that the above condition is always satisfied in the case where the underlying unknown field $g(x)$ has a compact 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:

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

There are various 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 2: (38)

a) $C(\omega) = |G(\omega)|^2 > 0,\ \ \forall \omega \in \Re$ or

b) $C(\omega)$ 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 (and adoption) 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 *bandlimited* 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.

Condition 3:
$$C(\omega) \text{ is continuous for every } \omega \in \Re \qquad (39)$$

then, under the three previous conditions, (36), (38) and (39), the auxiliary term $\bar{M}_{2\pi}(\omega)$ in eq. (34.a) 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 $\varphi(x, h)$, defined by eq. (31.a), a proper $L^2(\Re)$ function for any real positive value of the data resolution level h.

Finally, the condition that the optimal approximation 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, section 2.1) requires that its Fourier transform $\Phi(\omega, h)$ does not vanish at the origin. Taking into account eq. (31.a), this is transformed to the following simple condition for the signal power spectrum:

Condition 4:
$$\mathbf{C}(\omega)_{\omega=0} \neq 0. \qquad (40)$$

We are now in position to consider an infinite sequence $\{V_j\}_{j \in Z}$ of closed linear subspaces in $L^2(\Re)$. Each element of this sequence is defined as the closed linear span of the set $\{\varphi(x/h_j - n, h_j) \,|\, n \in Z\}$, where $\varphi(x, h_j)$ is the optimal approximation kernel given by eq. (31.a), and $h_j > 0$ denotes the data resolution level associated with each subspace V_j. We will further assume that

Condition 5:
$$h_j > h_{j+1}, \quad \forall j \in Z \qquad (41)$$

which makes $\{V_j\}$ a subspace sequence of *increasing resolution* in the Hilbert space $L^2(\Re)$. Note that the scaling parameter h_j is not restricted now to dyadic values (i.e. $h_j = 2^{-j}$), as it happens in the classic MRA case. By definition, the above subspace sequence satisfies the third (translation-invariance) basic property of an MRA (see, eq. (1.c), section 2.1) for any possible form of the scaling parameter h_j, i.e.

$$f(x) \in V_j \Leftrightarrow f(x + nh_j) \in V_j, \quad \forall n \in Z. \qquad (42)$$

In order for the specific sequence $\{V_j\}$ to satisfy the first («nesting») property of an MRA (see, eq. (1.a), section 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 $j \in Z$

Condition 6:

$$\frac{h_j}{h_{j+1}} = a_j, \quad \text{where} \quad a_j \in Z^+ - \{1\}. \tag{43}$$

The above condition implies that **any** two successive scaling parameters associated with the subspace sequence $\{V_j\}$ should be related through a 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:

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

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

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

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 **sub**space sequence $\{V_j\}$ to be nested. The proof is very easy and it can be found **in App**endix B.

Furthermore, the subspace sequence $\{V_j\}$ **con**structed from the optimal approximation kernel $\varphi(x, h_j)$ will also satisfy the **fourth** («completeness») basic property of an MRA (see, eq. (1.d), section 2.1). In order **to** prove that, we have to recall the fact that the 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 $V_j \subset L^2(\Re)$ of the **multi**resolution sequence $\{V_j\}$ will have the general form

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

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

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

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. (46.b) will be reduced to a simple identity, i.e.

$$F_j(\omega) = F_j(\omega) \qquad (47.a)$$

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

$$\lim_{j \to \infty} V_j = L^2(\mathfrak{R}). \qquad (47.b)$$

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 (46.b) 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 «shrinked» to the zero space in the $L^2(\mathfrak{R})$ sense, i.e.

$$\lim_{j \to -\infty} V_j = \{0\}. \qquad (48)$$

Lastly, we have to check if the family of translates $\{\varphi(x/h_j - n, h_j) \mid n \in Z\}$ of the optimal approximation kernel forms a *Riesz basis* for every element V_j that is spanned by this family. This final MRA property (see, eq. (1.e), section 2.1) is especially important, since it will ensure *stable* signal reconstruction schemes from their discrete samples within every multiresolution subspace V_j. A necessary and sufficient condition for this last property is (see section 2.1, and Unser and Daubechies, 1997)

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

where A and B are some strictly positive constants, 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. (31.a), the above inequality can be easily expressed as a function of the signal power spectrum $C(\omega)$ in the following way:

$$0 < A \leq \frac{\sum\limits_{k} \left(\dfrac{1}{h_j} C\left(\dfrac{\omega}{h_j} + \dfrac{2\pi k}{h_j} \right) \right)^2}{\left(\sum\limits_{k} \dfrac{1}{h_j} C\left(\dfrac{\omega}{h_j} + \dfrac{2\pi k}{h_j} \right) \right)^2} = \overline{M}_{2\pi}(\omega) \leq B < +\infty, \quad \forall \omega \in \mathfrak{R}, \quad h_j > 0 \quad (50)$$

where $\overline{M}_{2\pi}(\omega)$ is the same *2π-periodic* auxiliary function that was defined and used previously in eq. (34.a). At the beginning of this section we had already established that (under *conditions 1, 2 and 3*) the term $\overline{M}_{2\pi}(\omega)$ 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 the lower bound A and the upper bound B in the double inequality (50) is always guaranteed. Hence, the set of integer translates of the optimal approximation kernel $\{\varphi(x/h_j - n, h_j) \,|\, n \in Z\}$ 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 (21), gives rise to a generalized MRA-type structure $\{V_j\}$ in the Hilbert space $L^2(\mathfrak{R})$.* The main difference between this multiresolution subspace structure $\{V_j\}$ and the classic MRAs according to Mallat (1989a, b) is that its basic scaling kernel does not have a fixed form, but it 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. (31.a). 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 classic MRA property that will not necessarily be satisfied by the subspace sequence $\{V_j\}$, which is generated through the optimal approximation

kernel $\varphi(x, h_j)$ of statistical collocation is the «self-similar» dyadic scaling condition between the individual subspaces (see eq.(1b), section 2.1), i.e.

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

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 (43). Note that the optimal kernel $\varphi(x, h_j)$ essentially generates not just a *single* nested sequence $\{V_j\}$ of dense multiresolution subspaces in $L^2(\Re)$, 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) in section 5.1), as well as on the specific value of a *reference scale level* h_0. A list of such different alternatives is given in the following table. 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 scale values. Even for such dyadic scaling schemes, however, the self-similar property of eq. (51) will not necessarily be satisfied by the generalized MRA sequence associated with the optimal approximation kernel, unless we impose some further conditions on the signal power spectrum $C(\omega)$.

Table 1 – Sample of scale level values h_j associated with different generalized MRA sequences $\{V_j\}$

	Scale level generator (see eq. (43))					
	$\frac{h_j}{h_{j+1}} = 2j^2 + 3$		$\frac{h_j}{h_{j+1}} = 2^{\|j\|+1}$		$\frac{h_j}{h_{j+1}} = 2$	
	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_{-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

It is worth mentioning that all the derivations in section 5.1 are valid even if the frequency domain function $C(\omega)$ in the basic eq. (31.a) 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 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, the signal approximation obtained by the statistical collocation algorithm will still converge to the true field in the $L^2(\Re)$ sense, as the data resolution increases ($h_j \to 0$). However, the optimal MMSE principle of eq.(21) will not be rigorously satisfied in such cases. Nevertheless, the frequency domain structure of eq. (31.a) provides a useful general recipe for building generalized MRAs, based on appropriately selected functions $C(\omega)$ that satisfy the mild conditions given in the previous section.

The preceding 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 $V_j \subset L^2(\Re)$ 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 *sampling theorem* associated with the specific subspace V_j, since the set of translates of the 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 classic 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., Schmidt, 1981; Moritz, 1976).

It is also interesting to mention the essential difference between the original 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 section 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 like starting from the top of a pyramid (i.e. MRA) and by successive orthogonal projections onto more and more detailed resolution subspaces 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 scaled 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 classic 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 sub-spaces 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 $\psi(x, h_j)$ much similar to the scaling kernel given in eq. (31.a). If such a step becomes successful, we could essentially generate a «non-stationary» system of base functions in $L^2(\mathfrak{R})$ that will be directly associated with the actual statistical collocation formula (28); i.e. the 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 presented in future publications.

6. – CONCLUSIONS AND FUTURE WORK

The central idea of the present study was to demonstrate that the concept of multiresolution/wavelet theory lies at the very core of some of the approximation principles involved in physical geodesy problems. The method of spatio-statistical collocation, as expressed by the optimal estimation 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 Sansò's formulation for the collocation problem (Sansò, 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 auton-omous modelling tool, with remarkable connections to multiresolution approxi-mation theory. The importance of this link is that it provides basically the means to develop a useful generalization of Mallat's classic MRA approximation theory, where the data resolution level is not restricted only to 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.

Other additional theoretical extensions of the issues discussed herein are also needed in order to cover the spectrum of possible applications in gravity field modelling. 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 approxi-mation models. In the noiseless case, efficient algorithms for studying the behavior of the signal approximation error caused by the basic formula (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 unknown 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 in order to be realizable for actual gravity field applications.

One of the advantages of the present multiresolution formulation for statistical collocation is the easiness with which the «non-stationarity problem» can be 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. Although stationarity is a stochastic term that cannot be theoretically justified in the present purely deterministic approximation 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 approximation principles. With such an understanding of the problem, MRA and wavelet theory can definitely provide valuable tools without deviating from the universally acceptable collocation spirit (i.e. MMSE principle).

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. (31.a) 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. (31.a) might help to identify which types of kernels seem more realistic for approximating the actual behavior of the gravity field. 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) basically corresponds to a single multiplication between two n-dimensional vectors, whereas the matrix equation that utilizes the spatial CV function of the unknown field would require $n + 1$ multiplications of n-dimensional vectors plus an $n \times n$ matrix inversion, where n is the number of the available data points.

ACKNOWLEDGEMENTS

REFERENCES

A. Aldroubi, M. Unser (1992), *Families of wavelet transforms in connection with Shannon's sampling theory and the Gabor transform*. In: Chui C.K. (ed.). Wavelets: A tutorial in Theory and Applications. Academic Press, New York, pp. 509-528.

A. Aldroubi, M. Unser (1994), *Sampling Procedures in Function Spaces and Asymptotic Equivalence with Shannon's Sampling Theory*. Numer. Funct. Anal. Optim., 15 (1-2), pp. 1-21.

L. Ballani (1996), *Solving the Inverse Gravimetric Problem: On the benefit of Wavelets*. IAG Symposia Proceedings, vol. 114. Springer-Verlag, Berlin, pp. 151-161.

J.S. Bendat A.G. Piersol (1986), *Random Data: Analysis and Measurement Procedures*. 2nd edition. John Wiley, New York.

T. Blu, M. Unser (1999), *Quantitative Fourier Analysis of Approximation Techniques: Part I (Interpolators and Projectors) and Part II (Wavelets)*. IEEE Transactions on Signal Processing, vol. 47, no. 10, pp. 2783-2806.

W. Freeden, F. Schneider (1998), *An Integrated Wavelet Concept of Physical Geodesy*. Journal of Geodesy, 72: 259-281.

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

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

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

B. Jawerth, W. Sweldens (1994), *An Overview of Wavelet Based Multiresolution Analyses*. Siam Review, vol. 36, no. 3, pp. 377-412.

W. Keller (1998), *Geoid computation by collocation in scaling spaces*. IAG Symposia Proceedings, vol. 119. Springer-Verlag, Berlin, pp. 176-182.

P. KUMAR, E. FOUFOULA-GEORGIOU (1997), *Wavelet Analysis for Geophysical Applications*. Reviews of Geophysics, vol. 34, no. 5, pp. 385-412.

Z. LI (1996), *Multiresolution approximation of the gravity field*. Journal of Geodesy, 70, pp. 731-739.

S.G. MALLAT (1989.a), *A Theory for Multiresolution Signal Decomposition: the Wavelet Representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674-693.

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

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

H. MORITZ (1976), *Least-Squares Collocation as a Gravitational Inverse Problem*. Report No. 249, Department of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio.

H. MORITZ (1980), *Advanced Physical Geodesy*. Herbert Wichmann Verlag, Karlsruhe.

H. MORITZ, F. SANSÒ (1980), *A dialogue on collocation*. Bollettino di Geodesia e Scienze Affini, vol. 39, no. 1, pp. 49-51.

M.Z. NASHED, G.G. WALTER (1991), *General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces*. Mathematics of Control, Signals and Systems, 4, pp. 363-390.

A.V. OPPENHEIM, R.W. SCHAFER (1989), *Discrete-Time Signal Processing*. Prentice-Hall, Englewood Cliffs.

F. SANSÒ (1980), *The Minimum Mean Square Estimation Error Principle in Physical Geodesy (Stochastic and Non-Stochastic Interpretation)*. Bollettino di Geodesia e Scienze Affini, vol. 39, no. 2, pp. 112-129.

F. SANSÒ (1986), *Statistical Methods in Physical Geodesy*. In: Sunkel H. (ed.) Mathematical and Numerical Techniques in Physical Geodesy. Springer-Verlag, Berlin, pp. 49-156.

F. SANSÒ (1987), *Talk on the Theoretical Foundations of Physical Geodesy*. Paper presented at the 19th IUGG General Assembly, Vancouver, August 9-22, 1987. Published in «Contributions to Geodetic Theory and Methodology», Report no. 60006, Department of Surveying Engineering, University of Calgary, Calgary, pp. 5-28, Alberta.

H.F. SCHMIDT (1981), *Sampling Function and Finite Element Method Representation of the Gravity Field*. Reviews of Geophysics, vol. 19, no. 3, pp. 421-436.

M.G. SIDERIS (1995), *On the use of heterogeneous noisy data in spectral gravity field modeling methods*. Journal of Geodesy, 70, pp. 470-479.

S.L. SVENSSON (1983), *A New Geodesy Based upon Inversion-free Bjerhammar Predictors*. Proceedings of the 8th Symposium on Mathematical Geodesy, Como, September 7-9, 1981, pp. 449-468.

C.C. TSCHERNING (1986), *Functional Methods for Gravity Field Approximation.* In: H. Sunkel (ed.) Mathematical and Numerical Techniques in Physical Geodesy. Springer-Verlag, Berlin, pp. 3-48.

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

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

P. WOJTASZCZYK (1997), *A Mathematical Introduction to Wavelets.* London Mathematical Society Student Texts 37. Cambridge University Press, Cambridge.

X-G. XIA, Z. ZHANG (1993), *On Sampling Theorem, Wavelets, and Wavelet Transforms.* IEEE Transactions on Signal Processing, vol. 41, no. 12, pp. 3524-3535.

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

A.I. ZAYED (1993), *Advances in Shannon's Sampling Theory.* CRC Press, Boca Raton, FL.

APPENDIX A

In this appendix we will prove the following equation

$$\int_{-h/2}^{h/2} \left| E(\omega, x_0) \right|^2 dx_0 = hC(\omega) - \Phi_h^*(\omega)\, C(\omega) -$$
$$- \Phi_h(\omega)\, C(\omega) + \Phi_h(\omega)\, \Phi_h^*(\omega)\, \overline{C}_h(\omega).$$

(A.1)

Taking into account eq. (20.b), the error power spectrum at an arbitrary value of the sampling phase parameter x_0 has the form:

$$\left| E(\omega, x_0) \right|^2 = G(\omega)\, G^*(\omega) - \Phi_h^*(\omega)\, G(\omega)\, S^*(\omega) -$$
$$- \Phi_h(\omega)\, G^*(\omega)\, S(\omega) + \Phi_h(\omega)\, \Phi_h^*(\omega)\, S(\omega)\, S^*(\omega)$$

(A.2)

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

$$S(\omega) = \frac{1}{h} \sum_k G\left(\omega + \frac{2\pi k}{h}\right) e^{-i\frac{2\pi k}{h} x_0}.$$

(A.3)

Integrating equation (A.2) over x_0, we get analytically for every term

$$\int_{-h/2}^{h/2} G(\omega)\, G^*(\omega)\, dx_0 = G(\omega)\, G^*(\omega)\, h = hC(\omega)$$

(A.4)

$$\int_{-h/2}^{h/2} \Phi_h^*(\omega)\, G(\omega)\, S^*(\omega)\, dx_0 = \Phi_h^*(\omega)\, G(\omega) \int_{-h/2}^{h/2} \frac{1}{h} \sum_k G^*\left(\omega + \frac{2\pi k}{h}\right) e^{i\frac{2\pi k}{h} x_0} dx_0$$
$$= \frac{1}{h} \Phi_h^*(\omega)\, G(\omega) \sum_k G^*\left(\omega + \frac{2\pi k}{h}\right) \int_{-h/2}^{h/2} e^{i\frac{2\pi k}{h} x_0} dx_0$$
$$= \frac{1}{h} \Phi_h^*(\omega)\, G(\omega) \sum_k G^*\left(\omega + \frac{2\pi k}{h}\right) \int_{-\pi}^{\pi} \frac{h}{2\pi} e^{ik\xi} d\xi$$
$$= \Phi_h^*(\omega)\, G(\omega) \sum_k G^*\left(\omega + \frac{2\pi k}{h}\right) \frac{\sin k\pi}{k\pi}$$
$$= \Phi_h^*(\omega)\, G(\omega)\, G^*(\omega) = \Phi_h^*(\omega)\, C(\omega).$$

(A.5)

Following similar derivations as in eq. (A.5), we obtain

$$\int_{-h/2}^{h/2} \Phi_h(\omega)\, G^*(\omega)\, S(\omega)\, dx_0 = \Phi_h(\omega)\, C(\omega).$$

(A.6)

Finally, the integration of the last term in eq. (A.2) yields

$$
\int_{-h/2}^{h/2} \Phi_h(\omega)\, \Phi_h^*(\omega)\, S(\omega)\, S^*(\omega)\, dx_0 =
$$

$$
= \Phi_h(\omega)\, \Phi_h^*(\omega) \int_{-h/2}^{h/2} \frac{1}{h^2} \sum_n \sum_m G\!\left(\omega + \frac{2\pi n}{h}\right) G^*\!\left(\omega + \frac{2\pi m}{h}\right) e^{\, i\,\frac{2\pi(m-n)}{h}\, x_0}\, dx_0
$$

$$
= \frac{1}{h^2}\, \Phi_h(\omega)\, \Phi_h^*(\omega) \sum_n \sum_m G\!\left(\omega + \frac{2\pi n}{h}\right) G^*\!\left(\omega + \frac{2\pi m}{h}\right) \int_{-h/2}^{h/2} e^{\, i\,\frac{2\pi(m-n)}{h}\, x_0}\, dx_0
$$

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

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

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

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

$$
= \Phi_h(\omega)\, \Phi_h^*(\omega)\, \overline{C}_h(\omega).
$$

(A.7)

Combining together the results from equations (A.4) through (A.7), we get the initially claimed statement (A.1).

APPENDIX B

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

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

Each element $V_j \subset L^2(\Re)$ of this subspace sequence is defined as the closed linear span of the set $\{\varphi(x/h_j - n, h_j) \mid n \in Z\}$, where the kernel $\varphi(x, h_j)$ is defined by eq. (31.a), 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 of the unknown signal are assumed to satisfy all the mild conditions given in section 5.1 of the paper. Every signal $f_j(x) \in V_j$ will have the general form

$$f_j(x) = \sum_n b_n \, \varphi\left(\frac{x}{h_j} - n, h_j\right), \quad \forall f_j(x) \in V_j \tag{B.2}$$

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

$$\begin{aligned}
F_j(\omega) &= h_j \, \frac{C(\omega)}{\displaystyle\sum_k C\left(\omega + \frac{2\pi k}{h_j}\right)} \sum_n b_n e^{-i\omega n h_j} \\[2em]
&= h_j \, \frac{C(\omega)}{\displaystyle\sum_k C\left(\omega + \frac{2\pi k}{h_j}\right)} \, \overline{B}_{2\pi/h_j}(\omega), \quad \forall f_j(x) \in V_j
\end{aligned} \tag{B.3}$$

where $\overline{B}_{2\pi/h_j}(\omega)$ 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:

$$F_{j+1}(\omega) = h_{j+1} \, \frac{C(\omega)}{\displaystyle\sum_k C\left(\omega + \frac{2\pi k}{h_{j+1}}\right)} \, \overline{B}_{2\pi/h_{j+1}}(\omega), \quad \forall f_{j+1}(x) \in V_{j+1} \tag{B.4}$$

where h_{j+1} is the scaling parameter associated with V_{j+1}, and $\overline{B}_{2\pi/h_{j+1}}(\omega)$ denotes a certain $(2\pi/h_{j+1})$-periodic function with finite $L^2(0, 2\pi/h_{j+1})$ norm. It is quite easy now to transform eq. (B.3) in the form of eq. (B.4). Indeed, starting from eq. (B.3) we will

have

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

$$= h_{j+1} \frac{C(\omega)}{\sum_k C\left(\omega + \frac{2\pi k}{h_{j+1}}\right)} \frac{h_j \sum_k C\left(\omega + \frac{2\pi k}{h_{j+1}}\right)}{h_{j+1} \sum_k C\left(\omega + \frac{2\pi k}{h_j}\right)} \overline{B}_{2\pi/h_j}(\omega) \qquad (B.5)$$

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

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

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

$$\overline{\Lambda}_{2\pi/h_{j+1}}(\omega) = \frac{h_j \sum_k C\left(\omega + \frac{2\pi k}{h_{j+1}}\right)}{h_{j+1} \sum_k C\left(\omega + \frac{2\pi k}{h_j}\right)}. \qquad (B.6)$$

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 6 in eq. (43). For the same reason, the product of the two periodic functions $\overline{\Lambda}_{2\pi/h_{j+1}}(\omega)$ and $\overline{B}_{2\pi/h_j}(\omega)$, which is denoted by $\overline{N}_{2\pi/h_{j+1}}(\omega)$ in eq. (B.5), will also be a $(2\pi/h_{j+1})$-periodic function.

Furthermore, under *conditions* 1 and 2 [see, eqs. (36) and (38)], both the numerator and denominator in eq. (B.6) will converge uniformly to finite, strictly-positive, continuous periodic functions, for any pair of values for the scaling parameters h_j and h_{j+1}. Hence, the periodic function $\overline{\Lambda}_{2\pi/h_{j+1}}(\omega)$ will certainly have a finite $L^2(0, 2\pi/h_{j+1})$ norm. As a result, the auxiliary periodic function $\overline{N}_{2\pi/h_{j+1}}(\omega)$ in eq. (B.5) will also have a finite $L^2(0, 2\pi/h_{j+1})$ norm. In this way, the final frequency domain form of eq. (B.5) corresponds exactly to the expression of a function belonging in the higher resolution subspace V_{j+1}, according to the general formula (B.4).