Wavelets and Collocation: An Interesting Similarity

Christopher Kotsakis
Department of Geomatics Engineering, University of Calgary
2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4

Abstract

The rapid developments in the fields of multiresolution approximation theory and wavelets over the past few years have created an enormous amount of important theoretical knowledge and useful practical tools to be used for various signal processing applications. One of the most attractive properties of multiresolution/wavelet theory is the ability to study the details of a signal locally in various scale levels, according to a *zoom-in/zoom-out* approach. In gravity field modelling, on the other hand, we are used to employing (both theoretically and practically) the concept of collocation in order to approximate and study the behavior of unknown signals based on discrete data. One of the standard formulations of collocation, as a spatio-statistical linear approximation problem in purely deterministic fields, requires the use of a covariance function which is defined via a certain spatial averaging operator over the signal's domain. The "stationary" form of this CV function is often thought to provide strong limitations in the approximation framework, because the actual behavior of the gravity field is "non-stationary". The aim of this paper is to show that there does not really exist any "signal stationarity restriction" problem in statistical collocation, since it can be proven to be equivalent to a wavelet-type expansion. The connection between the two concepts is discussed, and also some recommendations for further work are given.

1 Introduction

The method of least-squares collocation (LSC) represents one of the major 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 collocation has been based (see, e.g., Tscherning 1986; Sanso 1986), a rigorous unified approximation approach that merges both the deterministic (Krarup's formulation) and the stochastic (Wiener's linear prediction theory) viewpoints behind LSC has long been established by Sanso (1980). Such an approach has eliminated, to some degree, most of the "pitfalls" in each individual formulation (e.g., reproducing kernel choice problem, non-stochasticity of the actual gravity field); see also Moritz (1980), Moritz and Sanso (1980). In this way, collocation is usually considered as a rigorous linear *spatio-statistical* method for gravity field approximation, where the term "statistical" is used not to describe some underlying stochastic behavior for the gravity field, but rather to specify the 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 Sanso'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 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 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 in Sanso (1980).

In order to eliminate any stationarity concerns about the spatio-statistical collocation framework, and to additionally support the transition towards the use of wavelet/multiresolution approximation techniques in gravity field modelling, the aim of this paper is to show that the optimal signal approximation according to the statistical collocation approach can be expressed in the wavelet-like linear form (e.g., for 1D signals):

$$\hat{g}(x) = \sum_{n} g(nh)\,\varphi\left(\frac{x}{h} - n\right) \tag{1}$$

where $g(x)$ is the unknown field under consideration, h is the resolution level of the discrete data $g(nh)$, and $\varphi(x)$ is a kernel related in a specific way to the "stationary" spatial CV function of $g(x)$. Convolution-based approximation models of the form (1) are a standard tool nowadays in most signal processing applications (Unser and Daubechies 1997; Blu and Unser 1999). Their constantly increasing popularity is due to their close connection with wavelet signal expansions which provide the best available mathematical tool today for localized signal analysis. This important link has actually resulted in the development of the vast field of *multiresolution approximation theory*, originally formulated by Mallat (1989).

2 Statistical Collocation and Data Resolution

In this section, the linear approximation problem for an unknown field $g(x) \in L^2(\Re)$ will be solved in such a way that the immediate connection between the estimated signal $\hat{g}(x)$ and the discrete data resolution will explicitly appear in the solution algorithm. An interesting discussion on the important interplay between data resolution and optimal approximation in gravity field modelling can be found in Sanso (1987). We will assume that the available data represent noiseless point values $g(nh)$ taken on a uniform grid with known resolution level h. The unknown field is considered as 1D for simplicity. The treatment for higher dimensions, i.e. in $L^2(\Re^2)$ or $L^2(\Re^3)$, is just a straightforward extension of the following derivations.

Since we are seeking a linear approximation, the recovered signal will have the general form

$$\hat{g}(x) = \sum_n g(nh)\,\varphi_{n,h}(x) \tag{2}$$

where $\varphi_{n,h}(x)$ is a family of unknown base functions whose dependence (if any) on the data resolution is introduced through the subscript h. If we further impose the condition of *translation-invariance* for the linear approximation with respect to the spatial reference system (in the multi-dimensional case this becomes invariance under more general affine transformations), then the family $\varphi_{n,h}(x)$ should be generated by a single kernel $\varphi_h(x)$ such that $\varphi_{n,h}(x) = \varphi_h(x - nh)$. In this way, eq.(2) is reduced to the simplified form

$$\hat{g}(x) = \sum_n g(nh)\,\varphi_h(x - nh) \tag{3}$$

The above approximation formula can now be illustrated in terms of the linear filtering system shown in Figure 1. Applying the Fourier transform to the "mixed" convolution equation (3), we get

$$\hat{G}(\omega) = \Phi_h(\omega)\ \overline{G}_h(\omega) \tag{4}$$

where $\hat{G}(\omega)$ and $\Phi_h(\omega)$ are the Fourier transforms of the approximated signal and the approximation kernel $\varphi_h(x)$, respectively. The term $\overline{G}_h(\omega)$ corresponds to the aliased periodic Fourier transform:

$$\overline{G}_h(\omega) = \frac{1}{h}\sum_k G(\omega + \frac{2k\pi}{h}) = \sum_n g(nh)\,e^{-i\omega nh} \tag{5}$$

where $G(\omega)$ is the Fourier transform of the original unknown signal (see, Oppenheim and Schafer 1989).

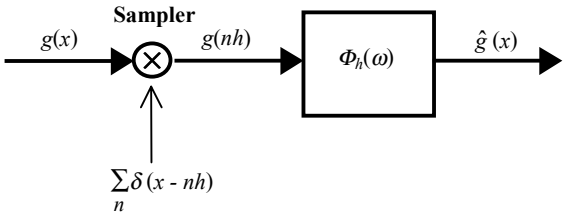

Fig. 1 Linear, translation-invariant signal approximation using discrete samples

Note that the above frequency domain formalism implies that we have sampled our unknown signal over its entire (finite or infinite) support. If the available data grid $g(nh)$ covers only some part of this (finite or infinite) support, then eq.(4) is certainly not valid and a window function should be additionally used. In order to avoid such complications, we will assume that the unknown field we try to approximate covers only the region inside the given grid boundaries. Although such an assumption is unacceptable in applications involving temporal signals (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 noted that, although $g(x)$ is assumed zero outside the given data grid boundaries, its approximation by eq.(3) may exhibit a non-zero pattern in this region.

2.1 A Spatio-Statistical Optimal Principle

The sequence $g(nh)$ 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 by an amount x_o, an infinite number of different data sequences can be obtained, which all represent different sampling schemes for the same signal at the same resolution. The situation is illustrated in Figure 2, 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 between the limits $-h/2 \le x_o \le h/2$.

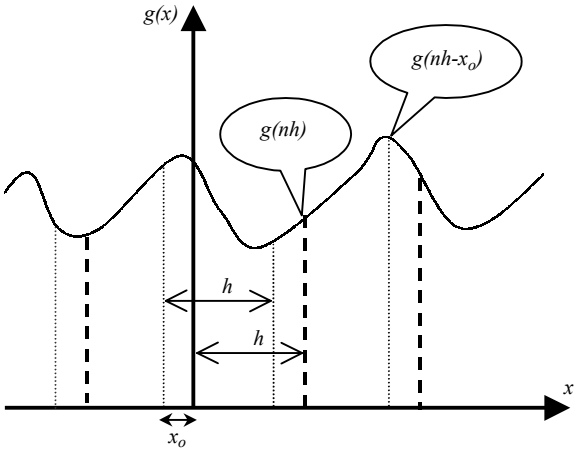

Fig. 2 Different signal sampling configurations at a given resolution level h

In accordance with the translation-invariance condition, we will now have the following general linear equation for the approximated signal from an arbitrary sampled sequence at resolution level h:

$$\hat{g}(x,x_o) = \sum_n g(nh - x_o)\,\varphi_h(x + x_o - nh) \tag{6}$$

Thus, the approximation error produced by eq.(6) becomes also a function of the sampling phase value x_o associated with the given data set, i.e.

$$e(x,x_o) = g(x) - \hat{g}(x,x_o) \tag{7}$$

Taking into account eq.(6) and applying the Fourier transform to the last equation (considered as a function of x only), we get

$$E(\omega,x_o) = G(\omega) - \frac{1}{h}\Phi_h(\omega)\sum_k G(\omega + \frac{2k\pi}{h})\, e^{-i\frac{2k\pi}{h}x_o} \tag{8}$$

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

$$P_e(\omega) = \frac{1}{h}\int_{-h/2}^{h/2} \left| E(\omega,x_o) \right|^2 dx_o = \min \tag{9}$$

Eq. (9) corresponds to 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" now has a purely spatio-statistical deterministic meaning, in contrast to Wiener's prediction theory where the mean error is defined in the sense of "experiment repetitions" via an expectation operator. The criterion (9) will minimize the mean error power spectrum over all possible sampling schemes for the given data resolution level h. It is similar to the use of the classic spatial averaging operator $\boldsymbol{M}$ in Moritz's (1980) book and in Sanso's (1980) paper, for the special case of 1D gridded data. It can be easily shown (see, Kotsakis 1999) that

$$P_e(\omega) = C(\omega) - \frac{\Phi_h^*(\omega)C(\omega)}{h} - \frac{\Phi_h(\omega)C(\omega)}{h} + \frac{\Phi_h(\omega)\Phi_h^*(\omega)\overline{C}_h(\omega)}{h} \tag{10}$$

where the asterisk * denotes complex conjugation, and $C(\omega)$ is just the Fourier transform of the usual spatial CV function $c(\xi)$ of the unknown deterministic signal $g(x)$, i.e.

$$c(\xi) = \int g(x)\, g(x+\xi)\, dx = \Im^{-1}\left\{ C(\omega) = G(\omega)G^*(\omega) \right\} \tag{11}$$

The symbol $\Im^{-1}$ denotes the inverse Fourier transform operator, and the term $\overline{C}_h(\omega)$ has the usual periodized form

$$\overline{C}_h(\omega) = \frac{1}{h}\sum_k C(\omega + \frac{2k\pi}{h}) \tag{12}$$

2.2 Resolution-Dependent Optimal Approximation

Using eqs.(9) and (10), we can easily solve the corresponding variational problem and obtain the optimal approximation kernel (for the analytical procedure, see, e.g., Sideris 1995). The frequency domain form of the optimal kernel will be finally given by the equation

$$\Phi_h(\omega) = \frac{C(\omega)}{\overline{C}_h(\omega)} = \frac{C(\omega)}{\frac{1}{h}\sum_k C(\omega + \frac{2k\pi}{h})} \tag{13}$$

The linear approximation procedure of spatio-statistical collocation, therefore, will be based on the following data filtering formula:

$$\hat{G}(\omega) = \frac{C(\omega)}{\sum_k C(\omega + \frac{2k\pi}{h})} \sum_k G(\omega + \frac{2k\pi}{h}) \tag{14}$$

which is verified by taking into account eqs.(4) and (5). It is seen from the last equation that, as the data resolution increases ($h \to 0$), the approximate will converge to the true unknown signal in the L^2 topology. In order for eq.(14) to correspond to a well defined filtering formula, the power spectrum $C(\omega)$ of the unknown signal has to satisfy some mild conditions that are explained in Kotsakis (1999).

Interestingly enough, the space domain form $\varphi_h(x)$ of the optimal approximation filter $\Phi_h(\omega)$ in eq.(13) can be expressed through the scaling relation

$$\varphi_h(x) = \varphi(\frac{x}{h}) \tag{15}$$

where the generating function $\varphi(x)$ is given by the following inverse Fourier transform

$$\varphi(x) = \Im^{-1}\left\{ \Phi(\omega) = \frac{C(\frac{\omega}{h})}{\sum_k C(\frac{\omega}{h} + \frac{2k\pi}{h})} \right\} \tag{16}$$

The justification of the above fact is trivial and it is based on the fundamental scaling property of the Fourier transform. If we finally substitute eq.(15) into the initial approximation model of eq.(3), we get the wavelet-like expression of eq.(1). It is worth mentioning that the basic optimal kernel $\varphi(x)$ will always be a *symmetric* function, since its Fourier transform $\Phi(\omega)$ given in eq.(16) is always a real-valued function.

2.3 Remarks

Convolution-based linear models of the form of eq.(1) are used in many signal processing applications in the context of classical interpolation, quasi-interpolation, and multiscale approximation via projections into multiresolution subspaces (see, e.g., Blu and Unser 1999). In such cases, the selection of the approximation model $\varphi(x)$ is usually made a priori and its performance is evaluated according to an assumed behavior for the unknown signal (e.g., bandlimitedness, smoothness, etc.), and/or other theoretical error bounds that depend on the form of the adopted model (i.e. Strang-Fix conditions); for more details, see Unser and Daubechies (1997). In the present paper, on the other hand, the selection of $\varphi(x)$ is adapted to the unknown signal itself through the use of an optimal MMSE principle, and its computation requires some knowledge of the signal's average behavior (i.e. spatial CV function). Note that the basic form of the optimal kernel $\varphi(x)$ depends directly on the data resolution level h, according to eq.(16). As a result, the linear model of eq.(1) in the collocation case will not employ scaled versions $\varphi(x/h)$ of the *same* kernel for each different value of h. This is in contrast to the classic wavelet approximation framework, where a *fixed* scaling kernel $\varphi(x)$ is used for any (dyadic) data resolution level. For more details and discussion, see Kotsakis (1999).

In our derivations we never assumed that the optimally approximated signal should reproduce the available data, i.e. $\hat{g}(nh) = g(nh)$. However, this will always be satisfied since the optimal kernel, defined by eq.(16), is a *cardinal/sampling* function. This simply means that

$$\varphi(n) = \begin{cases} 1 & , \ n = 0 \\ 0 & , \ n = \pm 1, \pm 2, \pm 3, \ldots \end{cases} \tag{17}$$

The above property can be easily verified in the frequency domain using the following relationship:

$$\sum_n \Phi(\omega + 2n\pi) = 1 \tag{18}$$

which is of course satisfied by our optimal filter $\Phi(\omega)$ given in eq.(16), for *any* data resolution value h. The valid-

ity of the cardinal property (17) can then be ensured through the well known *Poisson summation formula*.

The mean error power spectrum, corresponding to the use of the optimal filter $\Phi_h(\omega)$ given in eq.(13), will be

$$P_e(\omega) = C(\omega) \left(1 - \frac{C(\omega)}{\sum_k C(\omega + \frac{2k\pi}{h})} \right) \qquad (19)$$

where $P_e(\omega)$ is the same quantity defined in the MMSE criterion of eq.(9). There is a remarkable similarity between the above error formula and the formula giving the power spectral density (PSD) of the prediction error in noisy stationary random signals according to Wiener's optimal prediction theory (see, e.g., Sideris 1995). The same type of similarity also exists between the optimal approximation filter $\Phi_h(\omega)$ in eq.(13) and the actual Wiener filter.

However, the two formulations correspond to entirely diverse physical situations and they are based on completely different mathematical concepts and assumptions. In the statistical collocation approach, instead of having continuous, noisy and stationary random *input* signals, we deal with purely discrete and deterministic *input* data. Also, in this case the noise takes the form of lost information due to the discretization of the original unknown signal $g(x)$ (see Figure 1). Furthermore, there is no signal stationarity assumption involved in the current formulation. It is actually the translation-invariance condition, imposed for the statistical collocation case, that makes the two approaches algorithmically comparable in terms of *signal-to-noise* ratio (SNR) linear filters which are applied to the input data of each case.

3 Statistical Collocation, Multiresolution Approximation and Wavelets

The previous developments can be considered quite general, and they did not involve any special concepts from Mallat's multiresolution approximation theory. It is quite remarkable the fact that the spatio-statistical collocation framework actually leads to a *scale-invariant* 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 optimal collocation model of eq.(1) and the classic wavelet-based approximation methodology, due to the fact that the associated kernel $\varphi(x)$ in the collocation case changes for every different data resolution level h, according to the frequency domain form given in eq.(16).

The most appropriate way to describe the behavior of the signal approximation model of eq.(1), with the associated kernel $\varphi(x)$ given by the optimal frequency domain form in eq.(16), is to characterize it as: (i) *translation-invariant*, (ii) *scale-invariant*, and (iii) *data resolution-dependent*. Regardless of the origin and the scale of the reference system that is 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 shape. On the other hand, as the *data point density* (h) changes, the linear approximation algorithm of eq.(1) will employ a constantly changing model-kernel $\varphi(x)$, which will be adapted to the average spatial characteristics of the unknown signal and the resolution level h of the currently used data points in a certain optimal fashion, as suggested by eq.(16). For a more detailed treatment, see Kotsakis (1999).

Furthermore, the optimal kernel $\varphi(x)$ in the statistical collocation framework creates a "generalised" type of *multiresolution analysis* (MRA) within the Hilbert space $L^2(\Re)$. Let us denote by $\{V_j\}_{j \in Z}$ an infinite sequence of subspaces in $L^2(\Re)$, each element of which is associated with a specific data resolution level ($h_j > 0$) and it is defined as the closed linear span of the family $\{\varphi(h_j^{-1}x - n) \,|\, n \in Z\}$, where the basic scaling kernel $\varphi(x)$ is given by eq.(16) for $h = h_j$. In this case, it can be shown (Kotsakis 1999) that the multiresolution subspace sequence $\{V_j\}$ will satisfy all the basic properties that define an MRA structure in $L^2(\Re)$ (see, e.g., Mallat 1989), except from the following one:

$$f(x) \in V_j \;\Leftrightarrow\; f(2x) \in V_{j+1} \qquad (20)$$

Nevertheless, the optimal kernel $\varphi(x)$ in statistical collocation can be viewed as a *scaling function* whose integer translates $\varphi(h_j^{-1}x - n)$, at each resolution level h_j, create a linearly independent and stable system of base functions (i.e. *Riesz basis*) that can span a *nested* sequence of subspaces $\{V_j\}$, which will asymptotically converge to the $L^2(\Re)$ Hilbert space (as $h_j \to 0$), or to the "zero space" (as $h_j \to \infty$). The way that the value of the scaling

parameter h_j changes from one subspace (V_j) to the next (V_{j+1}) cannot of course be arbitrary, but it should generally satisfy the following condition:

$$\frac{h_j}{h_{j+1}} = a_j \qquad \forall \ \ j \in Z \tag{21}$$

where a_j is some positive *integer*, different from unity. The actual value of a_j is not needed to remain constant for each subspace pair, and this provides great flexibility in contrast to the classic Mallat's MRA framework where only *dyadic* schemes (i.e. $a_j = 2$) are considered. In a way, the "self-similar" scaling property of eq.(20) between the various subspaces V_j of a classic MRA is now replaced by the freedom to use a much more flexible rule according to which the scaling parameter (data resolution level) h_j changes from one MRA subspace to the next. A couple of mild conditions that the signal power spectrum $C(\omega)$ has to satisfy, in order for the optimal kernel $\varphi(x)$ in eq.(16) to generate such a multiresolution subspace structure in $L^2(\Re)$, are discussed in Kotsakis (1999).

The previous extension of the classic MRA concept suggests that we may be able to achieve a similar extension of the classic wavelet bases associated with Mallat's dyadic MRAs. If such a step becomes successful, we would have essentially generated a "non-stationary" system of base functions in $L^2(\Re)$ that will be explicitly associated with the actual statistical collocation approximation formula (1); i.e. each 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.

4 Conclusions and Future Work

The concept of spatio-statistical collocation, as expressed by the optimal criterion (9) and a translation-invariance condition, leads to signal approximation models commonly encountered in MRA/wavelet theory. It is the opinion of the author that Sanso's 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 modelling tool, with remarkable connections to multiresolution approximation theory. As far as the "stationarity" issue is concerned, I personally perceive this problem (in the context of optimal estimation in noiseless deterministic fields) as the ability to study locally the approximated signal in a rigorous and consistent manner with the optimal principles. With such an understanding of the problem, MRA/wavelet theory can provide valuable tools without deviating from the widely acceptable collocation spirit (i.e. MMSE principle).

Many theoretical/practical extensions of the issues discussed herein are needed to cover all possible gravity field applications. Some of these topics are: multi-dimensional generalizations (including compact spherical domains), study of the approximation error as a function of the data resolution level h and the used kernel $\varphi(x)$, development of optimal noise filtering methods in multiresolution approximation models, and empirical/numerical determination of the optimal kernel $\varphi(x)$ in eq.(16).

References

Blu T, Unser M (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.

Kotsakis C (1999) The Multiresolution Character of Collocation. *Journal of Geodesy*. (in press)

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

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

Moritz H, Sanso F (1980) A Dialogue on Collocation. *Boll. di Geod. e Sci. Affi.*, vol.39, no.1, pp. 49-51.

Oppenheim AV, Schafer RW (1989) *Discrete-Time Signal Processing*. Prentice-Hall, Inc.

Sanso F (1980) The Minimum Mean Square Estimation Error Principle in Physical Geodesy (Stochastic and Non-Stochastic Interpretation). *Boll. di Geod. e Sci. Affi.*, vol.39, no.2, pp. 111-129.

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

Sanso F (1987) Talk on the Theoretical Foundations of Physical Geodesy. *Contributions to Geodetic Theory and Methodology*, Report No.60006, pp. 5-28, Dept. of Surveying Engineering, University of Calgary, Calgary, Alberta.

Sideris MG (1995) On the use of heterogeneous noisy data in spectral gravity field modelling methods. *Journal of Geodesy*, 70, pp. 470-479.

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

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