

**UCGE Reports
Number 20138**

Department of Geomatics Engineering

**Multiresolution Aspects of Linear Approximation
Methods in Hilbert Spaces Using Gridded Data**

by

Christophoros Kotsakis

June 2000



UNIVERSITY OF
CALGARY

(URL: <http://www.geomatics.ucalgary.ca/GradTheses.html>)

UNIVERSITY OF CALGARY

Multiresolution Aspects of Linear Approximation Methods in Hilbert Spaces
Using Gridded Data

by

Christophoros Kotsakis

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

June, 2000

© Christophoros Kotsakis 2000

ABSTRACT

This thesis presents a novel optimal methodology for dealing with linear estimation problems in spatial deterministic fields, using discrete and regularly gridded data. More specifically, a unified study of various important issues that affect the theoretical analysis and practical computations associated with signal approximation problems (namely, stability, convergence, error analysis and choice of estimation model restrictions) is performed with respect to the data resolution parameter. A combination of different mathematical tools is employed for our theoretical developments, with the underlying ideas originating from the areas of deterministic collocation in Hilbert spaces, frame signal expansions, spatio-statistical collocation and multiresolution signal analysis theory. The spatio-statistical collocation principle is used to develop a new generalized multiresolution signal analysis scheme, which offers increased flexibility (in terms of scale level restrictions) and it is more powerful (in terms of approximation performance) than the classic dyadic multiresolution analyses that are associated with standard wavelet theory. Additional investigations are conducted on interpolation error analysis with respect to the data resolution level and the used estimation kernel, as well as on aliasing error propagation in convolution integral formulas using discrete gridded input data. Most of the theoretical developments are made with practical applications in mind, which means that an extensive (and original) treatment of the optimal noise filtering problem is also included, considering the most general case with non-stationary additive noise in the gridded input data.

PREFACE

This is a corrected version of the author's Doctor of Philosophy thesis of the same title. This thesis was accepted by the Faculty of Graduate Studies in June, 2000. The faculty supervisor of this work was Dr. M.G. Sideris. The other members of the examining committee were Dr. J.A.R. Blais, Dr. L.P. Bos, Dr. G. Lachapelle, Prof. F. Sanso and Dr. K.P. Schwarz.

ACKNOWLEDGEMENTS

After having survived the grueling experience of writing a Ph.D. thesis, it is a great pleasure to finally have the opportunity to thank all those who contributed in many various ways to its completion.

I wish to thank my supervisor Dr. M.G. Sideris for his continuous support, guidance and constructive criticism throughout the course of my graduate studies. Dr. J.A.R. Blais is gratefully acknowledged for his many stimulating comments to all my theoretical questions. I am deeply grateful to Dr. L.P. Bos for his suggestions in the purely mathematical aspects of my thesis. He helped me overcome some major obstacles during the preparation of Chapter 4. Prof. F. Sanso from Politecnico di Milano provided some enlightening remarks at the beginning of my work, which became really important at the final stages of this dissertation. I deeply thank all my other colleagues in the Department of Geomatics Engineering for their various contributions and, especially, my very good friend Georgia Fotopoulos for her support and encouragement during the final completion of my thesis, as well as for her thorough proofreading of the initial draft of the manuscript.

Financial support for this work was obtained from research grants awarded to my supervisor Dr. M.G. Sideris and to my co-supervisor Dr. G. Lachapelle by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and also from research grants awarded to Dr. M.G. Sideris by the Geomatics for Informed Decisions (GEOIDE) Network of Centres of Excellence (NCE) in Canada. Additional financial support was

received in the form of various Special Awards, Graduate Research Scholarships and Graduate Teaching Assistantships, as well as by the Helmut Moritz Graduate Scholarship, awarded to the author by the Department of Geomatics Engineering, University of Calgary. All this generous support is gratefully acknowledged.

Special thanks are extended to all my academic teachers during my undergraduate studies at the Department of Rural and Surveying Engineering in the University of Thessaloniki, especially Profs. A. Dermanis, E. Livieratos, D. Rossikopoulos, I.N. Tziavos, D. Arabelos, and A. Fotiou, who taught me that solid basic foundations is the most important step for innovative scientific work.

Finally, I would like to thank my family and all my friends in Greece for their love and support during the last four years. Above all, I want to thank my mother Agatha for without her help and advice this achievement would not have been possible.

Στην Αγάθη

στον Κώστα

και στη μνήμη του Τάκη Μελισσινού

and to the 'happy few'

TABLE OF CONTENTS

APPROVAL PAGE.....	ii
ABSTRACT	iii
PREFACE	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	viii
LIST OF FIGURES.....	xi
LIST OF SYMBOLS.....	xiv
1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Thesis Objectives.....	6
1.3 Thesis Outline.....	8
2 OVERVIEW OF MULTIREOLUTION APPROXIMATION THEORY.....	11
2.1 Multiresolution Analysis – Basic Definitions.....	12
2.2 Biorthogonal and Orthonormal Bases in Multiresolution Analyses.....	18
2.3 Sampling Bases in Multiresolution Analyses.....	22
2.4 Multiresolution Approximation via Orthogonal Projectors.....	25
2.5 Multiresolution Interpolation via Oblique Projectors.....	29
2.6 Wavelet Bases.....	35
3 FROM DETERMINISTIC COLLOCATION TO MULTIREOLUTION APPROXIMATION.....	42
3.1 What is Collocation ?	43
3.1.1 Deterministic Collocation.....	43
3.1.2 Stochastic-Probabilistic Collocation.....	46
3.1.3 Spatio-Statistical Collocation: A Compromise.....	49
3.1.4 Some Important Existing Problems.....	53
3.2 Linear Approximation in Hilbert Spaces.....	55

3.2.1	Inversion Scheme for Case 1	60
3.2.2	Inversion Scheme for Case 2	61
3.2.3	Comments	63
3.2.4	Modelling Considerations	65
3.3	Numerical Stability and the Role of Frames	67
3.3.1	General Frame Theory	68
3.3.2	Gabor and Affine Frames	73
3.3.3	Frames and Linear Approximation in Hilbert Spaces	77
3.3.4	A Note on Ill-Posed Problems	82
3.4	The Hilbert Space Choice Problem	83
3.4.1	Data Type and Configuration	86
3.4.2	Stable and Convergent Deterministic Collocation (Interpolation Problem)	88
3.5	Trade-Off Between Data and Model Resolution	96
3.6	The Connection with the MRA Concept	103
3.6.1	Linear Estimation as a Multiresolution Approximation	106
3.6.2	Final Remarks – Summary	113
4	OPTIMAL MULTIREOLUTION APPROXIMATION	117
4.1	Basic Aspects of the Multiresolution Model	118
4.2	Spectral Aspects of the Multiresolution Model	125
4.3	Linear Approximation and Data Resolution	129
4.3.1	General Formulation	130
4.3.2	A Spatio-Statistical Optimal Principle	134
4.3.3	Comments	139
4.4	The Multiresolution Character of Statistical Collocation	143
4.5	Optimal Multiresolution Approximation Kernels Using Synthetic Signal Power Spectra	147
4.6	Generalized Multiresolution Analysis	153
4.6.1	MRA Properties of the Optimal Collocation Kernel	154
4.6.2	An Interesting Result	163
4.6.3	Stability of the Optimal Riesz Bases	166

4.6.4 Final Remarks.....	171
5 ALIASING ERROR AND NOISE FILTERING IN LINEAR MULTIRESOLUTION APPROXIMATION MODELS	177
5.1 Accuracy of Linear Multiresolution Approximation Models.....	178
5.1.1 Multi-Parameter Error Description – Error CV Functions.....	180
5.1.2 Decay Rate of the Mean L^2 Approximation Error	190
5.1.3 Numerical Examples with Synthetic Signals	196
5.1.4 Aliasing Error Propagation in Convolution-Type Integral Formulas	202
5.1.4.1 Special Case 1: Band-Limited Data Referencing Model	205
5.1.4.2 Special Case 2: ‘No’ Data Referencing Model	208
5.1.5 Comparison with Wiener Filtering.....	214
5.2 Noise Filtering	218
5.2.1 Continuous Versus Discrete Noise	219
5.2.2 General Formulation	221
5.2.3 Optimization of the Noise Filter	225
5.2.4 The Cascade Structure of the Optimal Noise Filter	228
5.2.5 Additional Remarks	234
5.3 A Physical Geodesy Example.....	240
6 CONCLUSIONS AND RECOMMENDATIONS.....	246
6.1 Summary – Conclusions.....	246
6.2 Recommendations for Further Research	250
REFERENCES.....	253
APPENDIX A.....	265
APPENDIX B.....	268
APPENDIX C.....	272

LIST OF FIGURES

No.	Page
2.1 Optimal multiresolution signal approximation via orthogonal projection onto a dyadic MRA subspace V_j	29
2.2 Multiresolution signal interpolation in a dyadic MRA subspace V_j	32
2.3 Hierarchical decomposition of a multiresolution analysis $\{V_j\}$ in terms of a sequence of ‘detail-wavelet’ orthogonal subspaces $\{W_j\}$	38
3.1 Different ‘positions’ for the same geometry of a 2D data point network	51
3.2 Mapping type of the observation equations for Case 1 and Case 2	59
3.3 The geometry of the linear operator U for Case 1 and Case 2	59
3.4 Unacceptable function for the solution space $V_{\Delta x}$ in the case of stable optimal linear approximation using discrete samples with resolution Δx	90
3.5 Minimum-norm collocation using a ‘low-resolution’ r.k (lower graph) and a ‘higher-resolution’ r.k (upper graph)	101
3.6 Linear estimation as an isomorphic mapping T	109
4.1 Adaptation of the approximation kernel to the data resolution level ($\varphi(x)$ is an interpolating B-spline of fifth order in this example)	126
4.2 Filtering configuration of linear translation-invariant signal approximation using discrete samples	132
4.3 Different signal sampling configurations at a given resolution level h	135
4.4 Scale-invariant signal approximation at a certain data resolution level h (the value of the scaling parameter is assumed $\alpha > 1$)	145
4.5 Various models for the signal power spectrum $C(\omega)$	149
4.6 Fourier transform $\Phi(\omega, h)$ of the optimal approximation kernel for various data resolution levels h . The left column corresponds to a Gaussian model for the signal power spectrum, whereas the right column corresponds to an exponential model	150

LIST OF FIGURES (continued)

No.	Page
4.7	Fourier transform $\Phi(\omega, h)$ of the optimal approximation kernel for various data resolution levels h . The left column corresponds to the signal power spectrum model given by eq.(4.24c), whereas the right column corresponds to the model given by eq.(4.24d)..... 151
4.8	Plots of the 2π -periodic function $\overline{M}_{2\pi}(\omega)$, given by eq.(4.41), for various scale levels. The signal power spectrum is assumed to follow a Gaussian model 168
4.9	Plots of the 2π -periodic function $\overline{M}_{2\pi}(\omega)$, given by eq.(4.41), for various scale levels. The signal power spectrum is assumed to follow an exponential model 168
4.10	Plots of the 2π -periodic function $\overline{M}_{2\pi}(\omega)$, given by eq.(4.41), for various scale levels. The signal power spectrum is assumed to follow the model given by eq.(4.24c)..... 169
4.11	Plots of the 2π -periodic function $\overline{M}_{2\pi}(\omega)$, given by eq.(4.41), for various scale levels. The signal power spectrum is assumed to follow the model given by eq.(4.24d)..... 169
5.1	Multi-parameter description of the signal approximation error in linear translation-invariant MR models using 1D gridded data..... 182
5.2	Mean error power spectrum for Gaussian signal model using various interpolating kernels (the data resolution level is $h=4$) 187
5.3	Mean error power spectrum for Gaussian signal model using various interpolating kernels (the data resolution level is $h=1.5$) 187
5.4	Mean error power spectrum for the ‘experimental’ signal model using various interpolating kernels (the data resolution level is $h=1.5$) 188

LIST OF FIGURES (continued)

No.	Page
5.5	Mean error power spectrum for the ‘experimental’ signal model using various interpolating kernels (the data resolution level is $h=10$) 188
5.6	Mean error power spectrum for the $O(\omega^{-2})$ signal model using various interpolating kernels (the data resolution level is $h=1.5$) 189
5.7	Mean error power spectrum for the $O(\omega^{-2})$ signal model using various interpolating kernels (the data resolution level is $h=4$) 189
5.8	Filtering scheme of the linear MR estimation formula in eq.(5.1) for the case of band-limited signal interpolation..... 194
5.9	Tested MR approximation filters $\Phi(\omega)$ corresponding to different orthonormal scaling kernels $\varphi(x)$ 197
5.10	Decay rate of the mean error variance using various MR estimation kernels. The unknown signal follows a Gaussian power spectrum model..... 199
5.11	Decay rate of the mean error variance using various MR estimation kernels. The unknown signal follows an $O(\omega^{-2})$ -type power spectrum model..... 200
5.12	Decay rate of the mean error variance using various MR estimation kernels. The unknown signal follows an ‘experimental’ power spectrum model..... 200
5.13	Multiresolution filtering configuration of convolution-based integral formulas using discrete input data 203
5.14	Diagrammatic comparison between the classic Wiener estimation filter (bottom system) and optimal translation-invariant linear MR approximation using deterministic gridded data (top system)..... 216
5.15	Linear translation-invariant signal estimation using discrete noisy data 222
5.16	Two-step optimal translation-invariant filtering of discrete noisy data 229
5.17	Optimal geoid estimation from gridded noisy gravity anomalies using a certain multiresolution reference filter 241

LIST OF SYMBOLS

The most frequently used symbols and abbreviations throughout the thesis are listed below.

General symbols

\sum_n	summation over all integer numbers n
$\sum_{n,m}$	double summation over all integer numbers n and m
$\sum_{\vec{k}}$	summation over all (2×1) integer vectors \vec{k}
$\int dx$	integration over the whole real line
$\int_{\Omega} d\Omega$	integration over the domain Ω
\langle, \rangle	inner product
$\ \ $	norm (for signals)
$\ \ _{\text{sup}}$	supremum norm (for operators)
$ $	magnitude of a complex quantity, absolute value of a real quantity
\oplus	direct sum of two disjoint linear subspaces
$*$	complex conjugate (as superscript)

General symbols (continued)

\forall	for every
$\delta_{n,m}$	Kronecker delta
n, m, k, j	integer numbers
$\bar{\mathbf{n}}, \bar{\mathbf{m}}, \bar{\mathbf{k}}$	(2×1) vectors of integer numbers
i	imaginary unit

Sets and spaces

\Re	set of real numbers
\mathbb{Z}	set of integer numbers
\mathbb{Z}^+	set of positive integer numbers
$L^2(\Re)$	Hilbert space of square-integrable functions (over the real line)
$L^2(\Re^2)$	Hilbert space of square-integrable functions (over the real plane)
$l^2(\mathbb{Z})$	Hilbert space of square-summable infinite sequences
$l^2(\Gamma)$	Hilbert space of square-summable sequences over the index set Γ
H	arbitrary Hilbert space
V	linear subspace of H

Sets and Spaces (continued)

$V_{\Delta x}, V_h$	signal approximation subspaces associated with a certain data resolution level
$\{V_j\}$	multiresolution sequence of linear subspaces

Operators

U	general linear operator
U^{-1}	inverse of the linear operator U
\tilde{U}^{-1}	pseudo-inverse of the linear operator U
\mathfrak{S}	Fourier transform operator
\mathfrak{S}^{-1}	inverse Fourier transform operator
$g(x) \xleftrightarrow{\mathfrak{S}} G(\omega)$	Fourier transform pair
E	probabilistic expectation operator
F	frame operator
P_j, Q_j	projective operators associated with an approximation subspace V_j
*	discrete, continuous and semi-continuous convolution (as operator)
*	adjoint operator (as superscript)

Special functions and kernels

$\delta(x)$	Dirac delta function
$\delta_n, \delta[n]$	discrete delta sequence
$\chi_{a,b}$	indicator function over the interval $[a,b]$
$\text{sinc}(x)$	$\frac{\sin \pi x}{\pi x}$, sinc function
$K(P,Q), k(x,y)$	reproducing kernel
$\psi(x)$	wavelet kernel
$\varphi(x)$	scaling kernel
$\varphi_h(x)$	dilated version of the scaling kernel by h , $\varphi_h(x) = \varphi(x/h)$
$\Phi(\omega)$	Fourier transform of the scaling kernel
$\Phi_h(\omega)$	Fourier transform of the dilated scaling kernel
$\overline{W}_h(\omega)$	discrete noise filter at data resolution level h

Miscellaneous symbols and functions

$\Delta x, h$	data resolution level, grid sampling interval
x_o	sampling phase
$f(x), g(x)$	deterministic spatial signals

Miscellaneous symbols and functions (continued)

$c(x)$	signal spatial covariance function
$C(\omega)$	Fourier transform of the signal spatial covariance function
$G(\omega)$	Fourier transform of the signal $g(x)$
$ G(\omega) ^2$	power spectrum of $g(x)$ [same as $C(\omega)$]
$e(x)$	signal approximation error
$E(\omega)$	Fourier transform of $e(x)$
$e_h(x)$	resolution-dependent part of the signal approximation error
$e_v(x)$	noise-dependent part of the signal approximation error
$e(x, x_o, h)$	signal approximation error at a certain sampling phase and data resolution level
$ E(\omega, x_o, h) ^2$	power spectrum of the signal approximation error at a certain sampling phase and data resolution level
$c_e(\xi, x_o, h)$	error spatial covariance function at a certain sampling phase and data resolution level
$c_e^{aver}(\xi, h)$	mean error spatial covariance function at a certain data resolution level
$P_e(\omega, h)$	Fourier transform of $c_e^{aver}(\xi, h)$, mean error power spectrum
$\sigma^2(h)$	mean error variance as a function of the data resolution, $c_e^{aver}(0, h)$

Miscellaneous symbols and functions (continued)

$v(x)$	input data noise
σ_v	noise variance-covariance
$\bar{P}_v(\omega)$	discrete data noise PSD function

Abbreviations

BVP	boundary value problem
CV	covariance
FFT	fast Fourier transform
LSC	least squares collocation
MMSE	minimum mean square error
MR	multiresolution
MRA	multiresolution analysis
MSE	mean square error
PSD	power spectral density
r.k	reproducing kernel
RKHS	reproducing kernel Hilbert space
RMS	root-mean-square

Abbreviations (continued)

SNR	signal-to-noise ratio
vs.	versus
W-K	Wiener-Kolmogorov
1D	one-dimensional
2D	two-dimensional

Chapter 1

INTRODUCTION

1.1 Background

One of the most important tools in modern operational physical geodesy, as in many other areas of applied sciences and engineering, is the use of optimal approximation (or estimation) methods. Although physical systems can usually be described very precisely according to a certain theoretical model (e.g. Newtonian gravity field theory), their actual realization through discrete *observations* requires additional optimal estimation procedures. A wide variety of such procedures have been developed over the years, ranging from classic linear methods (Gaussian least-squares theory, Tikhonov regularization, Wiener-Kolmogorov theory, etc.), to more complicated non-linear estimation schemes (adaptive basis selection, maximum entropy estimation, Bayesian estimation, fuzzy methods, etc.). The choice of a specific estimation model is, to a certain extent, arbitrary, but usually we prefer linear methods due to their theoretical simplicity and straightforward implementation. Furthermore, the formalism of functional analysis offers

a convenient compact treatment of linear approximation methods within a geometrical Hilbert space framework, which, in turn, allows for a unified analysis of many diverse physical problems (Naylor and Sell, 1982; Kirsch, 1996; Debnath and Mikusinski, 1999).

In geodesy, there are numerous situations where the incorporation of an optimal estimation procedure is necessary. For example, Stokes's integral formula provides a very accurate theoretical basis for geoid determination (Heiskanen and Moritz, 1967), but its practical use with a discrete set of gravity measurements leads to an ill-posed problem with no unique solution. An external optimal estimation model/principle is then required in order to approximate the geoid signal in a unique manner. Stated in a general fashion, every geodetic application that utilizes discrete (and possibly noisy) spatial data, for the recovery of an unknown field or signal, is associated with a corresponding estimation problem that needs to be solved in some optimal interpolatory sense (Bjerhammar, 1987).

Regardless of the specific method used, the solution of such operational geodetic approximation problems should obey some basic properties. Perhaps the most important among these properties is the stability of the solution algorithm. An estimation model that is very sensitive to small perturbations in the discrete input data can produce large errors in the final results, which may deviate significantly from reality (Gerstl and Rummel, 1981). Robust methods are always preferable in order to ensure a well-conditioned signal approximation scheme, with small numerical distortions from computer round-off errors and external noise effects. An equally important aspect, within a spatial estimation

framework, is also the convergence of the solution algorithm to the true field as the amount of discrete data increases (Eeg and Krarup, 1973; p. 39). It would be useless to acquire high-resolution data sets from modern satellite, airborne or terrestrial sensors, if we do not have the ability to process them in a consistent (in the sense of estimation theory) manner. The problems of stability and convergence are actually strongly related to each other. Not only do we want to use a convergent estimation methodology that can fully recover an unknown spatial field using infinitely dense data, but we must also ensure that its algorithm remains reasonably stable as the data density increases. It is rather ironic that the most celebrated operational method in geodesy (i.e. collocation), along with its many different facets, becomes highly ill-conditioned for increasing data resolution (Sjoberg, 1978; Rummel et al., 1979).

Apart from the stability and convergence issues, optimal estimation methods should also be able to provide an insightful view to the behaviour of the underlying unknown fields, as well as to the quality of their approximation. In this respect, spectral and error analysis procedures have to be applied to the results of the signal estimation algorithms. Until recently, the only mathematical tools available to geodesists for studying the spectral characteristics of their signals were the classic Fourier decompositions/transformations, which measure the signal spectral content exclusively in the frequency domain, in terms of either spatial frequencies or spatial wavelengths (Kaula, 1959; Schwarz, 1984). This approach, however, can only offer a very narrow viewpoint to their underlying behaviour, since it overlooks important localized trends and signal irregularities. Localized spectral

information is far more insightful than the ‘global averaging’ implied in the Fourier methods of harmonic analysis, and it is already used in various types of signal processing applications (Cohen, 1995; Hlawatsch and Boudreaux-Bartels, 1992). More importantly, the results of any spectral analysis method depend directly on the estimation model within which we choose to approximate our unknown fields from their discrete data. For example, if we use the collocation concept for gravity field approximation in a Hilbert space of piecewise-constant harmonic functions, it would not be very illuminating to employ a Fourier-based spectral analysis for the estimated signals. Although this specific example is quite extreme, it nevertheless reveals the need to adapt the spectral analysis procedures to the signal estimation models (or vice versa).

In terms of accuracy analysis for the results of operational approximation algorithms, there exists a wide variety of qualitative and quantitative signal error measures, depending on the specific properties of the geodetic estimation technique used. The deterministic version of collocation, for example, can offer rigorous upper bound values for the signal error norm within a certain Hilbert space (Dermanis, 1976; Tscherning, 1986), which are often difficult to admit a practically useful interpretation, and even more difficult to compute. The stochastic/probabilistic facet of collocation (Dermanis, 1976; Sanso, 1986), on the other hand, is overall problematic for interpolation error analysis in geodetic signals using noiseless discrete data, since the underlying fields do not alter their behaviour over different observation (‘experiment’) repetitions, as the variance-covariance propagation law requires. An additional issue of special importance, for both

theoretical and practical studies, is the development of an algorithmic procedure that computes the decay rate of some functional of the signal estimation error with respect to the data resolution level. Such a resolution-dependent error modelling scheme, however, is not yet available within the signal approximation framework used in geodesy.

Since estimation methods are only artificial mathematical constructions, their application to physical problems is always associated with a number of additional modelling choices or assumptions. In geodetic problems, we often need to impose specific a-priori restrictions to our signals before we can employ a certain operational approximation technique. Typical examples of such situations include the norm (or reproducing kernel) choice problem in deterministic collocation (Dermanis, 1977), the signal stationarity and ergodicity assumption for the practical implementation of stochastic collocation (Moritz, 1980), and the additional noise stationarity assumption for the application of Wiener-type optimal filtering schemes (Sideris, 1995; Sanso and Sideris, 1995). The implications of these modelling issues are usually neglected in practical applications, although their importance is quite significant and it should always remind us of the limitations and/or the drawbacks of our optimal estimation tools.

To this end, we have identified a number of major factors that affect both the theoretical analysis and the practical computations associated with signal estimation problems, namely, stability, convergence, spectral and error analysis, and choice of model restrictions. A unified study of these important issues, with respect to the data resolution

parameter, is the central focus of this research work, whose specific objectives are discussed in the following section.

1.2 Thesis Objectives

The overall objective of this thesis is to present a new optimal methodology for dealing with linear approximation problems in spatial deterministic fields, using discrete and regularly gridded data. The key task is to develop a stable estimation framework for increasing data resolution, that can be reduced to a well defined signal description when the data become infinitely dense. In addition, no a-priori smoothing restrictions should be imposed to the true unknown fields, which are allowed to exhibit irregular local variations at any scale level.

A number of different mathematical tools are employed for the theoretical developments of this thesis. In particular, we combine various ideas originating from the deterministic collocation concept (Moritz, 1980), from the spatio-statistical interpretation of collocation according to Sanso (1978), and finally from the multiresolution analysis (MRA) concept according to Mallat (1989a,b). The latter corresponds to a relatively new tool of approximation theory that has been developed at an explosive rate over the last decade. Its increasing popularity stems from its immediate connection with the wavelet theory, which currently represents the most sophisticated framework for signal analysis and estimation (Sweldens, 1996; Mallat, 1998a,b).

A primary objective of the thesis is to combine, in a cascading manner, the three aforementioned theoretical concepts in order to eliminate their individual limitations. As a first step, the use of MRA methods will be introduced as a necessary regularization scheme for dealing with the stability, convergence and modelling issues of deterministic collocation, when gridded data of increasing resolution are used. Some important aspects of the general linear approximation problem in Hilbert spaces are also presented, using a tool of functional analysis known as frame expansions (Young, 1980). The next step involves the incorporation of the spatio-statistical collocation principle in order to determine an optimal MRA model for the stable estimation of an unknown field from its gridded data at a given resolution.

Another important objective, which is directly associated with the previous cascading methodology, is to extend the classic multiresolution analysis methods for more general cases than the ones implied in Mallat's approximation model. The original MRA concept requires that the data resolution level is always given in a dyadic form, a fact that restricts its applicability in many practical situations. Again, the spatio-statistical collocation principle will be used to develop a new generalized MRA scheme, which is more flexible in terms of scale level restrictions and more powerful in terms of approximation performance.

The problem of error analysis in multiresolution signal estimation forms the basis of one more objective of this research. The emphasis is given on the derivation of a simple

rigorous algorithmic procedure that can measure the decay rate of the mean square interpolation error as a function of the data resolution and the used estimation kernel. In addition, the problem of aliasing error propagation in convolution-type integral formulas with gridded input data is studied, which is of great importance in many geodetic applications.

The last topic that is addressed in the thesis is noise filtering. Although the previous objectives concentrate on important aspects of the linear approximation problem in spatial deterministic fields using errorless data, the most realistic situation arises when the discrete input observations are influenced by random noise. A new linear multiresolution method for optimal noise filtering is presented, containing certain similarities with the existing Wiener-type filtering schemes. The latter, however, rely entirely on the stationarity for the input data noise, an assumption that will not be imposed in our developments.

1.3 Thesis Outline

The analysis and the results of this research work are presented in the next five chapters. In the sequel, the main structure and the general contents of each chapter are outlined.

In Chapter 2, the necessary mathematical background on multiresolution analysis methods for signal approximation is given. The presentation closely follows Mallat's

original formulation, with the addition of some more recent important results. Although the developments in this dissertation do not explicitly incorporate the concept of wavelet signal expansions, their relationship with the multiresolution analysis framework is briefly explained in order to follow some of the discussions in the forthcoming chapters.

In Chapter 3, a descriptive theoretical overview of the various versions of the collocation method is provided and some important existing problems are identified. A detailed analysis of the linear signal approximation problem in Hilbert spaces is also presented, revealing interesting modelling aspects that have not been previously discussed in the geodetic literature. The concept of frame signal expansions is then explained, and it is used for studying the stability and convergence properties of deterministic collocation with gridded data at uniform resolution. The last part of the chapter introduces the use of the MRA principle, within a Hilbert space linear estimation framework, as a necessary regularization tool for obtaining unperturbed stability and convergence for increasing data density.

The problem of optimal multiresolution signal approximation is treated in Chapter 4, which contains many of the original contributions of this thesis. A brief discussion on the modelling and spectral-related advantages for the use of multiresolution estimation techniques in geodesy is first given. The spatio-statistical collocation principle is then employed for the development of a new linear estimation method, which results in a certain class of optimal resolution-dependent interpolating kernels. A few examples for

the behaviour and the stability performance of these optimal kernels, at different data resolution levels, are provided using some synthetic signal models. This chapter also contains the construction of a generalized MRA scheme for signal analysis, that overcomes the usual dyadic restriction for the scale level values. Finally, a detailed discussion on the similarities and differences between this new multiresolution structure and the traditional dyadic MRA schemes concludes the chapter.

Chapter 5 deals with the problems of aliasing error analysis and noise filtering within a multiresolution signal estimation model using gridded data at a certain resolution. A simple frequency-domain algorithm is constructed for the computation of the decay rate of the mean square estimation error as a function of the data resolution level and the used estimation kernel. Numerical examples are also given to test this algorithm with the help of simple synthetic signal models and various scaling estimation kernels. The issue of aliasing error propagation in convolution-type integral formulas is also studied, and some special cases of interest in geodetic practice are identified. The second part of the chapter deals with the problem of optimal noise filtering in multiresolution signal interpolation models. An original Wiener-type linear method is developed that has the ability to be easily implemented using FFT techniques, even in the presence of non-stationary observational noise. Finally, some conclusions and recommendations for further research work are given in Chapter 6.

Chapter 2

OVERVIEW OF MULTIREOLUTION APPROXIMATION THEORY

Generally stated, multiresolution methods deal with the analysis and synthesis of signals in terms of a scale-based representation (Cohen, 1993). This offers a more flexible (and often more natural) framework than the usual Fourier techniques of harmonic analysis, which consider only frequency-based signal representations. The development of such methods was initiated for various purposes in many different scientific fields, a factor that has resulted in numerous different approaches for their underlying theoretical background. For example, electrical engineers and signal analysts use the concepts of filter banks, multi-rate systems and joint time-frequency or time-scale distributions (Rioul and Vetterli, 1994; Hlawatsch and Boudreaux-Bartels, 1992; Cohen, 1995), pure mathematicians and physicists employ ideas from integral transform theory to obtain a continuous and/or a discretized scale-based decomposition of functions (Heil and Walnut, 1994; Daubechies et al., 1986; Daubechies, 1990), whereas approximation theorists study the subject from a functional analysis point of view (De Vore and Lucier, 1992; De Boor

et al., 1994; Unser and Daubechies, 1997). For some historical reviews and related applications, see the papers by Cohen (1989) and Daubechies (1996).

In this chapter, we will follow just one among these many ‘roots’, which is directly related with the specific problems tackled later in this thesis. This is also the approach that has managed to unify with a common language the disciplines of signal processing, wavelet analysis and approximation theory (Mallat, 1998a). It is important to emphasize that a comprehensive introduction to multiresolution methods in signal approximation requires an extensive amount of discussion, which is not possible to be given within the limits of a single chapter. Therefore, in the following sections we will only present some basic fundamental concepts that are going to be used in the rest of the thesis. For more mathematical details, the textbooks by Walter (1994), Holschneider (1995), Mertins (1996) and Mallat (1998b) should be consulted.

2.1 Multiresolution Analysis – Basic Definitions

The concept of multiresolution signal analysis is a relatively recent one, originally developed by Mallat (1989a,b). Wavelet signal expansions, on the other hand, appeared long before Mallat’s formulation, with the most common example being the asymptotic approximation of L^2 signals by the translates of piecewise-constant base functions (i.e. Haar wavelet bases). Since there exists a very strong connection between these two

mathematical subjects, they are usually viewed as the ‘two sides of the same coin’, although there do exist pathological cases of wavelet signal expansions that cannot be identified under Mallat’s multiresolution framework, which is described below.

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

$$(i) \quad V_j \subset V_{j+1} \quad , \quad \forall j \in \mathbb{Z} \quad (2.1)$$

$$(ii) \quad f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1} \quad (2.2)$$

$$(iii) \quad f(x) \in V_j \Leftrightarrow f(x+n2^{-j}) \in V_j \quad , \quad \forall n \in \mathbb{Z} \quad (2.3)$$

$$(iv) \quad \text{span} \bigcup_{j=-\infty}^{+\infty} V_j = L^2(\mathfrak{R}) \quad \text{and} \quad \bigcap_{j=-\infty}^{+\infty} V_j = \{0\} \quad (2.4)$$

$$(v) \quad \text{A scaling function } \varphi(x) \in V_0, \text{ with a non-vanishing integral, exists such that the family } \varphi(x-n)_{n \in \mathbb{Z}} \text{ is a Riesz basis for } V_0. \quad (2.5)$$

The above definition is not minimal, in the sense that some of the conditions (i)-(v) can be derived from the remaining ones (Wojtaszczyk, 1997). However, it has been customary to treat all these five MRA properties as independent statements. Condition (ii) implies that all the individual subspaces $\{V_j\}$ of a multiresolution analysis are dyadically

scaled versions of each other, which in addition form a nested sequence according to condition (i). The third condition assigns a translation-invariant property to the MRA subspaces with respect to their associated dyadic resolution interval $\Delta x_j = 2^{-j}$. Finally, condition (iv) imposes a kind of causal behaviour in the nested sequence $\{V_j\}$, which should be dense in $L^2(\mathfrak{R})$ (for high resolution) and it should shrink to the zero space (for low resolution).

The concept of a **Riesz basis** is a simple generalization of the notion of orthonormal bases in Hilbert spaces, corresponding to a set of linearly independent functions that form a complete, oblique and stable system of reconstructing elements. A set of functions $\varphi_n(x)_{n \in \mathbb{Z}}$ that spans a Hilbert space H is said to be a Riesz basis, if and only if there exist constants $C \geq c > 0$ such that

$$c \left(\sum_n |a_n|^2 \right)^{1/2} \leq \left\| \sum_n a_n \varphi_n(x) \right\|_H \leq C \left(\sum_n |a_n|^2 \right)^{1/2} \quad (2.6)$$

for any square-summable sequence of scalars $\{a_n\}_{n \in \mathbb{Z}}$. If we have a Riesz basis $\varphi_n(x)_{n \in \mathbb{Z}}$ in an arbitrary Hilbert space, then there always exists a unique biorthogonal system $\tilde{\varphi}_n(x)_{n \in \mathbb{Z}}$ which also forms a Riesz basis for the same space (Wojtaszczyk, 1997). The biorthogonality property between the two systems is expressed by the formula

$$\langle \varphi_n(x), \tilde{\varphi}_m(x) \rangle = \delta_{n,m} \quad (2.7)$$

where \langle, \rangle denotes the inner product in the underlying Hilbert space, and $\delta_{n,m}$ corresponds to the Kronecker delta symbol. More details on the general theory of Riesz bases can be found in Young (1980) and Heil and Walnut (1994).

If a Riesz basis (in the usual L^2 norm) is formed by the integer translates $\varphi(x-n)_{n \in \mathbb{Z}}$ of a basic function $\varphi(x) \in L^2(\mathfrak{R})$, then the Fourier transform $\Phi(\omega)$ of the generating kernel should always satisfy the following relationship (Wojtaszczyk, 1997):

$$0 < A \leq \sum_k |\Phi(\omega + 2\pi k)|^2 \leq B < \infty \quad (2.8)$$

for some strictly positive finite bounds A and B . Equation (2.8) represents a very basic and important condition in the multiresolution analysis framework, which hereinafter will be referred to as the *Riesz condition*. It is actually equivalent to the space-domain definition of a Riesz basis given in eq.(2.6), when $\varphi_n(x)_{n \in \mathbb{Z}} = \varphi(x-n)_{n \in \mathbb{Z}}$ and $\| \cdot \|_{\mathbb{H}} = \| \cdot \|_{L^2}$. In the special case where the set $\varphi(x-n)_{n \in \mathbb{Z}}$ forms an orthonormal basis

for its closed linear span $V_o \in L^2(\mathfrak{R})$, the Riesz condition takes the simple form

$$\sum_k |\Phi(\omega + 2\pi k)|^2 = 1 \quad (2.9)$$

Throughout this thesis, Fourier transform functions $\Phi(\omega)$ that satisfy the above equation will be called orthonormal filters (or just orthogonal filters, for the cases where the constant on the right-hand side in eq.(2.9) differs from the value one).

If a kernel $\varphi(x) \in L^2(\mathfrak{R})$ satisfies the Riesz condition given in eq.(2.8), then it can be shown that the sets $\varphi(x/h - n)_{n \in \mathbb{Z}}$ form Riesz bases for their corresponding closed linear spans $V_h \subset L^2(\mathfrak{R})$, for any non-zero value of the scaling parameter h (Unser and Daubechies, 1997). In this way, the family of the dyadic translates $\varphi(2^j x - n)_{n \in \mathbb{Z}}$ will also form a Riesz basis for every nested subspace V_j of the MRA that is generated by the scaling function $\varphi(x)$.

From this brief introduction, it is clear that we can adopt two main ways for establishing the existence of a multiresolution analysis, i.e.

- We can take the subspaces $\{V_j\}$ as our basic, given objects. They have to satisfy conditions (i)-(iv), which usually are rather easy to check. Then, we need to find a scaling function that satisfies the last property (v). This is usually not so obvious.
- We can start with a scaling function $\varphi(x) \in L^2(\mathfrak{R})$ that satisfies the basic Riesz condition in eq.(2.8). We define the initial subspace V_0 as the closed linear span

of the set $\varphi(x-n)_{n \in \mathbb{Z}}$, and the other subspaces V_j are defined accordingly through the scaling condition (ii). Hence, the translation-invariance property (iii) is automatically satisfied, and we only need to check the validity of conditions (i) and (iv).

There is also a third way that is not as evident from the MRA definition. This method starts from the so-called *scaling* or *dilation equation* (Strang, 1989)

$$\varphi\left(\frac{x}{2}\right) = \sum_n a_n \varphi(x-n) \quad (2.10)$$

and then tries to build the scaling function from there by using an appropriate and unique choice of square-summable coefficients $\{a_n\}_{n \in \mathbb{Z}}$. The above equation is just an expression of the fact that $V_{-1} \subset V_0$, where the subspace V_0 is defined as the closed linear span of a translation-invariant Riesz basis $\varphi(x-n)_{n \in \mathbb{Z}}$ and the subspace V_{-1} is simply generated by the dyadic scaling condition (ii). Clearly, only very special sequences $\{a_n\}$ can satisfy a scaling equation as shown in eq.(2.10) and, furthermore, create a scaling function $\varphi(x)$ whose associated subspace sequence $\{V_j\}$ satisfies the completeness condition (iv). Nevertheless, this approach has been used extensively for the construction of various MRA models in $L^2(\mathbb{R})$ with certain optimal properties for their scaling function $\varphi(x)$, such as compact support, symmetry, cardinality,

orthonormality, number of vanishing moments, etc. For more details and examples, see Daubechies (1992), Mallat (1998b) and Hernandez and Weiss (1996).

2.2 Biorthogonal and Orthonormal Bases in Multiresolution Analyses

Every finite-energy signal that belongs into an arbitrary dyadic subspace $V_j \subset L^2(\mathfrak{R})$ of some multiresolution analysis model $\{V_j\}$ has the general linear expansion form

$$f(x) = \sum_n a_n \varphi(2^j x - n) \quad , \quad \forall f(x) \in V_j \quad (2.11)$$

where $\varphi(x) \in V_0$ is the scaling function associated with the underlying MRA, and $\{a_n\}_{n \in \mathbb{Z}}$ represents a square-summable sequence of coefficients that is uniquely determined through the conventional L^2 inner product, i.e.

$$\begin{aligned} a_n &= 2^j \langle f(x), \tilde{\varphi}(2^j x - n) \rangle \\ &= 2^j \int f(x) \tilde{\varphi}(2^j x - n) dx \end{aligned} \quad (2.12)$$

The analysis kernel $\tilde{\varphi}(x) \in V_0$ is called the *dual* (or biorthogonal) scaling function, and it generates the same multiresolution analysis as $\varphi(x)$. It is uniquely defined by the following equation (Unser and Daubechies, 1997):

$$\tilde{\varphi}(x) = \sum_n c_n^{-1} \varphi(x-n) \quad (2.13a)$$

or, in the frequency domain

$$\tilde{\Phi}(\omega) = \frac{\Phi(\omega)}{\sum_k |\Phi(\omega + 2\pi k)|^2} \quad (2.13b)$$

where c_n^{-1} is the so-called *convolution inverse* of the autocorrelation sequence c_n of the synthesis scaling function $\varphi(x)$, i.e.

$$\begin{aligned} c_n &= \varphi(-x) * \varphi(x) \Big|_{x=n} \\ &= \int \varphi(y) \varphi(y+n) dy \end{aligned} \quad (2.14a)$$

Note that the superscript in the coefficients c_n^{-1} does not mean the usual arithmetic inversion, but it just implies that

$$c_n * c_n^{-1} = \sum_m c_m c_{n-m}^{-1} = \delta_n \quad (2.14b)$$

where δ_n denotes the discrete delta sequence. The two sets of functions, $\varphi(2^j x - n)_{n \in \mathbb{Z}}$ and $\tilde{\varphi}(2^j x - n)_{n \in \mathbb{Z}}$, will constitute a pair of biorthogonal Riesz bases for the same MRA

subspace V_j at the given dyadic scale level. For more details, see Aldroubi and Unser (1993) and Mallat (1998b).

The concept of a Riesz basis in an infinite-dimensional Hilbert space V_j is identical with the existence of a linear topological isomorphism between V_j and the space $l^2(\mathbb{Z})$ of square-summable sequences. The latter contains the expansion coefficients $\{a_n\}_{n \in \mathbb{Z}}$ for all signals in V_j with respect to the Riesz basis under consideration; see eq.(2.6). Since linear topological isomorphisms are (by definition) continuous mappings, the linear MRA expansion in eq.(2.11) will always yield a stable signal representation (in the L^2 norm) for every function $f(x) \in V_j$. Furthermore, the stability of the individual Riesz bases $\varphi(2^j x - n)_{n \in \mathbb{Z}}$ will remain constant at every dyadic scale level, due to the self-similarity of the MRA subspaces $\{V_j\}$ imposed by eq.(2.2), and the scale-invariance property of the L^2 norm*. The ideal situation occurs when the basic Riesz basis $\varphi(x - n)_{n \in \mathbb{Z}}$ is additionally orthogonal, in which case the signal expansion in eq.(2.11) becomes a perfectly stable linear algorithm at any dyadic scale level (i.e. its condition number is always equal to one).

* By scale-invariance property of the L^2 norm, we mean that $\|f(x)\|_{L^2} = \sqrt{|a|} \|f(ax)\|_{L^2}$ for every $L^2(\mathbb{R})$ signal and for any non-zero scaling factor a .

In every MRA subspace V_j , an infinite number of orthonormal bases can be constructed from a given Riesz basis $\varphi(2^j x - n)_{n \in \mathbb{Z}}$ according to the orthonormalization trick given in Young (1980, p. 46); see also Holschneider (1995, p. 187) and Wojtaszczyk (1997, p. 24). These orthonormal bases will be comprised of normalized dyadic translates $\sqrt{2^j} \varphi_o(2^j x - n)_{n \in \mathbb{Z}}$ of a basic scaling function $\varphi_o(x) \in V_o$, which is usually called the *orthonormal scaling function*. For the same multiresolution analysis in $L^2(\mathfrak{R})$, we can thus have many different choices for its generating scaling function (Aldroubi and Unser, 1993). The most typical choice for the orthonormal scaling kernel $\varphi_o(x)$ is given by the frequency-domain normalization formula (Wojtaszczyk, 1997)

$$\Phi_o(\omega) = \frac{\Phi(\omega)}{\sqrt{\sum_k |\Phi(\omega + 2\pi k)|^2}} \quad (2.15)$$

where $\Phi(\omega)$ is the Fourier transform of some non-orthonormal scaling function $\varphi(x) \in V_o$. It can easily be verified that $\Phi_o(\omega)$, as defined in the last equation, satisfies the orthonormality Riesz condition that was given in eq.(2.9). For an alternative orthonormalization procedure within a multiresolution analysis framework, see Mallat (1989b).

2.3 Sampling Bases in Multiresolution Analyses

Under mild decaying restrictions on the orthonormal scaling function $\varphi_o(x)$, every subspace $V_j \subset L^2(\mathfrak{R})$ of an MRA sequence is a *reproducing kernel Hilbert space* (RKHS). Its actual reproducing kernel $k_j(x, y)$ is given by the ‘quasi-stationary’ scaling expression (Walter, 1992)

$$k_j(x, y) = 2^j k_o(2^j x, 2^j y) \quad (2.16a)$$

and

$$k_o(x, y) = \sum_n \varphi_o(x - n) \varphi_o(y - n) \quad (2.16b)$$

where $k_o(x, y)$ is the reproducing kernel of the ‘unit scale’ or ‘reference’ MRA subspace V_o . As the resolution index j increases, the reproducing kernel $k_j(x, y)$ will gradually converge to the Dirac delta function $\delta(x - y)$, which corresponds to an informal reproducing kernel for all $L^2(\mathfrak{R})$ signals (Walter, 1993). Note that $k_o(x, y)$ can also be expressed with respect to a pair of biorthogonal scaling functions, as follows (Unser and Daubechies, 1997):

$$k_o(x, y) = \sum_n \varphi(x - n) \tilde{\varphi}(y - n) \quad (2.17)$$

In every dyadic multiresolution analysis in $L^2(\mathfrak{R})$ that possess a sequence of proper reproducing kernels, we can associate a certain **sampling theorem** with each of its nested subspaces $V_j \subset L^2(\mathfrak{R})$. This interesting result was originally derived by Walter (1992) and it was subsequently extended by many authors (Xia and Zhang, 1993; Janssen, 1993; Aldroubi and Unser, 1994; Djokovic and Vaidyanathan, 1997).

According to Walter (1992), the set $k_o(x, n)_{n \in \mathbb{Z}} = k_o(x - n, 0)_{n \in \mathbb{Z}}$ formed by the integer translates of the basic reproducing kernel $k_o(x, y)$ provides an alternative Riesz basis for the reference MRA subspace V_o . Its dual basis is generated by the integer translates of a specific kernel $\tilde{k}_o(x) \in V_o$ that possess the cardinal (sampling) property, i.e.

$$\tilde{k}_o(n) = \begin{cases} 1 & , n = 0 \\ 0 & , n = \pm 1, \pm 2, \pm 3, \dots \end{cases} \quad (2.18)$$

The expansion of an arbitrary signal $f(x) \in V_o$ with respect to such a cardinal Riesz basis takes the form of a generalized sampling theorem, as follows:

$$\begin{aligned} f(x) &= \sum_n a_n \tilde{k}_o(x - n) \\ &= \sum_n \langle f(x), k_o(x, n) \rangle_{L^2} \tilde{k}_o(x - n) = \end{aligned}$$

$$= \sum_n f(n) \tilde{k}_o(x-n) \quad (2.19)$$

The situation can easily be extended at any dyadic resolution level. In this way, for every function $f(x) \in V_j$, we have

$$f(x) = \sum_n f(n2^{-j}) \tilde{k}_o(2^j x - n) \quad (2.20)$$

The determination of the sampling scaling kernel $\tilde{k}_o(x)$ can be easily performed in the frequency domain, according to the formula (Walter, 1992)

$$\tilde{K}_o(\omega) = \frac{\Phi_o(\omega)}{\sum_k \Phi_o(\omega + 2\pi k)} \quad (2.21)$$

where $\Phi_o(\omega)$ is the Fourier transform of the orthonormal scaling function $\varphi_o(x)$ shown in eq.(2.16b).

The existence of uniform-sampling signal expansions in MRA subspaces provides a useful extension of the well known Shannon's sampling theorem for band-limited signals (Jerri, 1977). As a matter of fact, Shannon's interpolating formula fits perfectly into Mallat's multiresolution analysis setting, since its basic scaling kernel $\varphi_s(x)$ (the *sinc* function) generates a special MRA sequence $\{V_j\}$ of band-limited nested subspaces in

$L^2(\mathfrak{R})$, with their associated bandwidths given by the dyadic form $[-2^j\pi, 2^j\pi]$. For more details and examples on the connection between sampling theorems and MRA theory, see Zayed (1993), Walter (1994) and Aldroubi and Unser (1992, 1994). An additional excellent reference is Nashed and Walter (1991), where the notion of sampling theorems is studied in a more abstract Hilbert space framework.

2.4 Multiresolution Approximation via Orthogonal Projectors

A multiresolution analysis $\{V_j\}$ can be used to determine a certain linear approximation

$f_j(x) \in V_j$ of an arbitrary finite-energy signal $f(x) \in L^2(\mathfrak{R})$ at a dyadic resolution level

$\Delta x_j = 2^{-j}$. In fact, this was the main motivation behind the original formulation of the

MRA concept; see the discussion in Mallat (1989a). The first and the fourth basic

properties of an MRA, according to eqs.(2.1) and (2.4), imply that the accuracy of such an

approximation scheme can be made arbitrarily high by selecting a suitable resolution

interval Δx_j . Higher resolution approximations $f_j(x)$ ($j \rightarrow +\infty$) will yield smaller

errors $\|f(x) - f_j(x)\|_{L^2}$, whereas lower resolution approximations $f_j(x)$ ($j \rightarrow -\infty$)

will gradually yield no information at all about the field $f(x)$ under consideration.

The most straightforward way to perform such a multiresolution signal approximation is to employ a sequence of *orthogonal projectors* $\{P_j\}$ that can be linked with the nested subspace sequence $\{V_j\}$. The optimal signal approximation $\hat{f}_j(x)$, at a certain dyadic resolution level, can then be determined through the orthogonal projection of the original signal $f(x) \in L^2(\mathfrak{R})$ onto the corresponding MRA subspace V_j , i.e.

$$\hat{f}_j(x) = P_j f(x) \tag{2.22}$$

which, of course, satisfies the minimum L^2 error norm criterion

$$\|e_j(x)\|_{L^2} = \|\hat{f}_j(x) - f(x)\|_{L^2} = \min \tag{2.23}$$

among any other signals $f_j(x) \in V_j$ at the same scale level. Such a projective scheme is usually called the **least-squares solution** for the signal approximation problem (Unser and Daubechies, 1997).

Due to the nature of the L^2 inner product, the computation of the optimal approximation $\hat{f}_j(x)$ implies a very interesting algorithmic procedure, which is described by the following linear formula (Mallat, 1989a):

$$\begin{aligned}
\hat{f}_j(x) &= P_j f(x) = \sum_n a_n \varphi(2^j x - n) \\
&= 2^j \sum_n \langle f(y), \tilde{\varphi}(2^j y - n) \rangle_{L^2} \varphi(2^j x - n) \\
&= 2^j \sum_n \left[\int f(y) \tilde{\varphi}(2^j y - n) dy \right] \varphi(2^j x - n) \\
&= 2^j \sum_n \left[f(\xi) * \tilde{\varphi}(-2^j \xi) \right]_{\xi=n2^{-j}} \varphi(2^j x - n) \\
&= 2^j \left[f(\xi) * \tilde{\varphi}(-2^j \xi) \right]_{\xi=n2^{-j}} * \varphi(2^j x)
\end{aligned} \tag{2.24}$$

where the symbol $*$ is used to denote both continuous and semi-continuous convolution operations. The kernels $\varphi(x)$ and $\tilde{\varphi}(x)$ correspond to a pair of biorthogonal scaling functions for the specific MRA model in which the linear multiresolution approximation takes place. We can also express the optimal signal estimate $\hat{f}_j(x)$ in terms of a more compact projective operation, as follows:

$$\begin{aligned}
\hat{f}_j(x) &= P_j f(x) = \langle k_j(x, y), f(y) \rangle_{L^2} \\
&= \int k_j(x, y) f(y) dy \\
&= 2^j \int k_o(2^j x, 2^j y) f(y) dy
\end{aligned} \tag{2.25}$$

where $k_j(x, y)$ is the reproducing kernel of the approximation subspace $V_j \subset L^2(\mathfrak{R})$ at the desired resolution level; see eq.(2.16).

The algorithm for computing the optimal signal approximation $\hat{f}_j(x) = P_j f(x)$ consists of three basic steps, connected in a linear cascading manner. Initially, the original field $f(x)$ is filtered through a certain *analysis* kernel, an operation which stems from the first (continuous) convolution shown in eq.(2.24). The analysis kernel $\tilde{\varphi}(-2^J x)$ corresponds to a flipped version of the dual scaling function that is tuned to the desired scale level of approximation by a proper dyadic dilation. Such a prefiltering operation has a kind of ‘anti-aliasing’ role within the linear approximation procedure. The output signal is then sampled at the corresponding dyadic rate $\Delta x_j^{-1} = 2^J$ and the result is the sequence of coefficients $\{a_n\}$ appearing in eq.(2.24). Finally, a *synthesis* kernel $\varphi(2^J x)$ is applied to this coefficient sequence through the second (semi-continuous) convolution operator shown in eq.(2.24), in order to obtain the optimal signal estimate. Of course, when an orthonormal scaling function $\varphi_o(x)$ is employed in the approximation procedure then both the analysis and the synthesis kernels coincide with the same function, i.e. $\varphi(x) = \tilde{\varphi}(x) = \varphi_o(x)$. The above three-step projective scheme is illustrated with the help of the linear system shown in Figure 2.1.

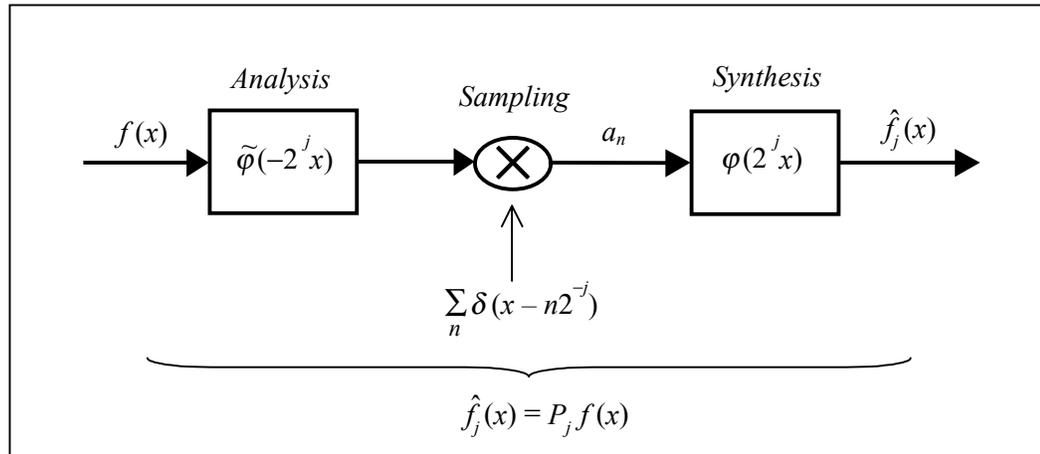


Figure 2.1 Optimal multiresolution signal approximation via orthogonal projection onto a dyadic MRA subspace V_j

An exhaustive error analysis of the previous approximation algorithm can be found in Blu and Unser (1999b). A more general study for the performance of the least-squares signal estimate $\hat{f}_j(x) = P_j f(x)$, in terms of pointwise, asymptotic and L^2 error analysis, is also given in Unser and Daubechies (1997) and Blu and Unser (1999a), where the various approximation subspaces $V_j \subset L^2(\mathfrak{R})$ at each scale level are not necessarily restricted to form a nested MRA sequence.

2.5 Multiresolution Interpolation via Oblique Projectors

Another simpler approach to obtain a multiresolution approximation for a finite-energy signal $f(x) \in L^2(\mathfrak{R})$, within some MRA model $\{V_j\}$, is to use only its discrete values at a uniform sampling resolution $\Delta x_j = 2^{-j}$. By applying the sampling theorem associated

with the corresponding MRA subspace at the given data scale level, we can compute an interpolating signal approximation as follows:

$$\hat{f}_j^{\text{int}}(x) = \sum_n f(n2^{-j}) s(2^j x - n) \quad (2.26)$$

where $s(2^j x - n)_{n \in \mathbb{Z}}$ denotes the **Riesz sampling basis** in the used MRA model (see section 2.3). In general, we have that $\hat{f}_j^{\text{int}}(x) \neq f(x)$ since the original signal does not necessarily belong in the specific approximation subspace V_j . Moreover, the signal interpolant $\hat{f}_j^{\text{int}}(x)$ will generally be different from the least-squares approximation $\hat{f}_j(x)$ obtained through the orthogonal projective algorithm of eq.(2.24).

Strictly speaking, the multiresolution interpolation scheme in eq.(2.26) corresponds to a signal estimation procedure, whereas the orthogonal projective formula in eq.(2.24) is related to a signal approximation methodology. Such a distinction follows the rigorous mathematical terminology, which reserves the term ‘approximation’ to describe the procedure of replacing a known quantity with a simpler (smoother) one, whereas ‘estimation’ is usually used to denote the approximation of an unknown quantity from incomplete or imperfect data. In this thesis, however, these two terms will be used interchangeably.

As in the case of the least-squares multiresolution approximation, the interpolating signal estimate in eq.(2.26) can also be expressed as a projective operation

$$\hat{f}_j^{\text{int}}(x) = Q_j f(x) \quad (2.27)$$

where Q_j denotes a certain projector from the Hilbert space $L^2(\mathfrak{R})$ onto an MRA subspace V_j . The role of the above operator is to sample the original signal $f(x)$ at a dyadic resolution level $\Delta x_j = 2^{-j}$, and then reconstruct a continuous waveform

$\hat{f}_j^{\text{int}}(x) \in V_j$ using these samples and the interpolating scaling function $s(x) \in V_0$. In order to establish the projective property of the interpolator Q_j , it is sufficient to show

that $Q_j^2 = Q_j$. Obviously, we have that

$$g(x) = Q_j g(x) \quad , \quad \forall \quad g(x) \in V_j \quad (2.28)$$

since the set $s(2^j x - n)_{n \in \mathbb{Z}}$ is an interpolating basis for every MRA subspace $V_j \subset L^2(\mathfrak{R})$. As a result, we can write the following:

$$\begin{aligned} \hat{f}_j^{\text{int}}(x) &= Q_j \hat{f}_j^{\text{int}}(x) \\ &= Q_j Q_j f(x) = \end{aligned}$$

$$= Q_j^2 f(x) \quad (2.29)$$

If we now compare eqs.(2.27) and (2.29), we can conclude that the operator Q_j possess the projective property. Note that the multiresolution interpolation corresponds to an *oblique* signal projection onto the approximation subspace V_j . Thus, it is less accurate than the least-squares approximation solution, which is associated with an orthogonal signal projection onto V_j . For theoretical and practical comparisons between the two methods, as well as for cases where the basic scaling kernel $s(x) \in V_0$ in eq.(2.26) is not strictly interpolating, see Unser and Daubechies (1997) and Blu and Unser (1999a,b). The previous convolution-based interpolation procedure, at a certain dyadic resolution level, is illustrated in Figure 2.2

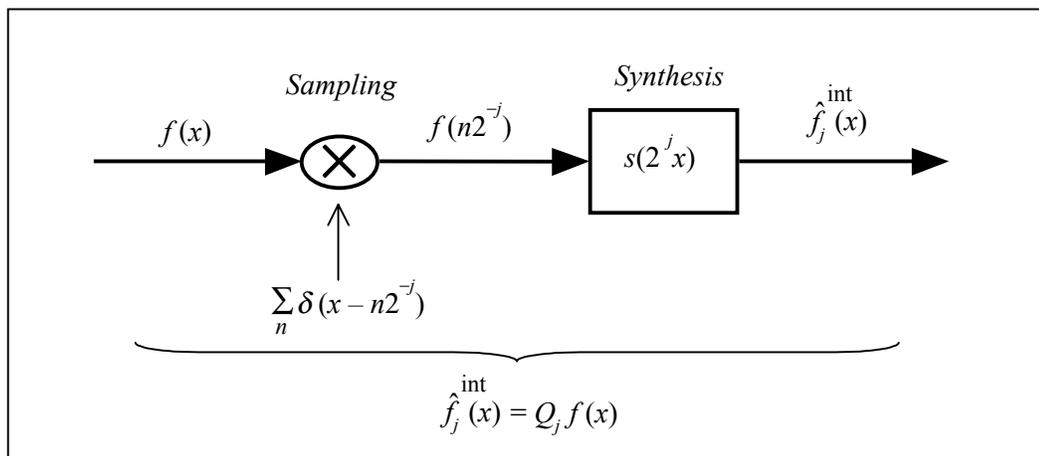


Figure 2.2 Multiresolution signal interpolation in a dyadic MRA subspace V_j

It should be mentioned that the input unknown signal $f(x)$ in the projective interpolation algorithm of eq.(2.26) or (2.27) cannot be *any* finite-energy function, because the sampling operation in the $L^2(\mathfrak{R})$ Hilbert space may result in non square-summable sequences $f(n\Delta x_j)$. However, this is of little concern in physical/practical problems where it can safely be assumed that the underlying fields always produce square-summable data sequences, at any finite sampling level. On the other hand, *any* square-integrable signal can be approximated arbitrarily well (in the L^2 norm) by an MRA sampling expansion of increasing resolution, since MRA subspace sequences $\{V_j\}$ are dense in the whole Hilbert space $L^2(\mathfrak{R})$. For more details, see Blu and Unser (1999a,b) and Unser (2000).

A very useful property of Riesz sampling bases, within the multiresolution analysis framework, is that they provide the same stability level for the signal interpolation algorithm of eq.(2.26) at every dyadic value of the data resolution. If $s(2^j x - n)_{n \in \mathbb{Z}}$ denotes such a Riesz sampling basis in an MRA sequence $\{V_j\}$, then we have

$$g(x) = \sum_n g(n2^{-j}) s(2^j x - n) \quad , \quad \forall g(x) \in V_j \quad (2.30)$$

From the general definition of a Riesz basis (see section 2.1), we also have the analogous isomorphic condition

$$c \left(\sum_n |g(n2^{-j})|^2 \right)^{1/2} \leq \left\| \sum_n g(n2^{-j}) s(2^j x - n) \right\|_{L^2} \leq C \left(\sum_n |g(n2^{-j})|^2 \right)^{1/2} \quad \forall g(x) \in V_j \quad (2.31)$$

for some strictly positive constants $c \leq C$. Using the self-similarity property of the MRA subspaces $\{V_j\}$ according to eq.(2.2) and the scale-invariance property of the L^2 norm (see the footnote in page 20), it can be shown that the condition number of the linear algorithm in eq.(2.30) remains constant at every sampling resolution $\Delta x_j = 2^{-j}$. Its actual value is determined by the ratio $\sqrt{B/A}$ of the maximum lower and the minimum upper bounds that appear in the fundamental frequency-domain Riesz condition, i.e.

$$0 < A \leq \sum_k |S(\omega + 2\pi k)|^2 \leq B < \infty \quad (2.32)$$

where $S(\omega)$ denotes the Fourier transform of the basic sampling kernel $s(x) \in V_0$. Note that the condition number of the linear expansion formula in eq.(2.30) is defined as the product of two *operator (supremum) norms* associated with the forward and the inverse isomorphic mappings between V_j and $l^2(\mathbb{Z})$, which are implied in the Riesz condition of eq.(2.31); see Strang (1988) and Phillips and Taylor (1996).

2.6 Wavelet Bases

A function $\psi(x) \in L^2(\mathfrak{R})$ is called a **wavelet** (or mother wavelet) if the set of its normalized dyadic translates $\sqrt{2^j} \psi(2^j x - n)_{j,n \in \mathbb{Z}}$ generates an orthonormal basis for the Hilbert space $L^2(\mathfrak{R})$. If we use the more compact notation $\psi_{j,n}(x)$ to denote an arbitrary wavelet basis, we can write the following expansion formula:

$$\begin{aligned} f(x) &= \sum_{j,n} \langle f(x), \psi_{j,n}(x) \rangle_{L^2} \psi_{j,n}(x) \\ &= \sum_{j,n} a_{j,n} \psi(2^j x - n) \quad , \quad \forall f(x) \in L^2(\mathfrak{R}) \end{aligned} \tag{2.33}$$

where the 2D sequence of wavelet coefficients $\{a_{j,n}\}$ is determined by the linear filtering procedure

$$\begin{aligned} a_{j,n} &= 2^j \int f(y) \psi(2^j y - n) dy \\ &= 2^j \left[f(x) * \psi(-2^j x) \right]_{x=n2^{-j}} \quad , \quad j, n \in \mathbb{Z} \end{aligned} \tag{2.34}$$

The values of the above coefficients correspond to gridded samples of successively filtered versions of the original signal $f(x)$ by the wavelet kernel $\psi(x)$. The resolution interval of this sampling operation is entirely adapted to the spread of the wavelet filter,

which varies in a dyadically exponential fashion. In the signal processing literature, such an analysis scheme is usually called *octave-band* or *Q-constant analysis* (Mertins, 1996).

Among other methodologies, the concept of multiresolution analysis provides a very general and straightforward approach for building wavelet bases in $L^2(\mathfrak{R})$. This fundamental result was first established by Mallat (1989b) and it marked the beginning of an enormous amount of theoretical and practical developments in the fields of signal processing and approximation theory, which continue at an explosive rate up to date. As a matter of fact, wavelet bases that do not arise from an MRA structure are essentially exceptional cases (Weiss and Hernandez, 1996; p. 47).

The transition from a multiresolution analysis model $\{V_j\}$ to an associated wavelet basis $\psi_{j,n}(x)$ is based on the following decomposition scheme for every MRA subspace:

$$V_j = V_{j-1} \oplus W_{j-1} \tag{2.35}$$

where the symbol \oplus denotes the direct sum between two disjoint linear subspaces. The subspace W_{j-1} is the orthogonal complement of V_{j-1} in V_j , and it basically contains the additional precision of the least-squares multiresolution signal approximation when the dyadic resolution increases from 2^{j-1} to 2^j . If we now introduce a reference scale level

2^{j_0} and apply the orthogonal decomposition of eq.(2.35) successively, we can get the general equation

$$V_j = V_{j_0} \bigoplus_{k=j_0}^{j-1} W_k \quad (2.36)$$

where all the individual complements W_k are orthogonal to each other due to the nesting property of the MRA subspaces $\{V_j\}$, i.e.

$$W_{j_0} \perp W_{j_0+1} \perp \dots \perp W_{j-1} \quad (2.37)$$

If we also take into account the fourth (completeness) basic property of a multiresolution analysis according to eq.(2.4), we can finally obtain a total orthogonal partition of the entire $L^2(\mathfrak{R})$ Hilbert space as follows:

$$L^2(\mathfrak{R}) = \bigoplus_{j=-\infty}^{+\infty} W_j \quad (2.38)$$

The elements of the orthogonal sequence $\{W_j\}$ are usually called the ‘detail subspaces’ in order to indicate the fact that they contain the signal information needed to go from a coarse resolution representation to a higher one, within a certain MRA model $\{V_j\}$. The above hierarchical decomposition framework is illustrated in Figure 2.3.

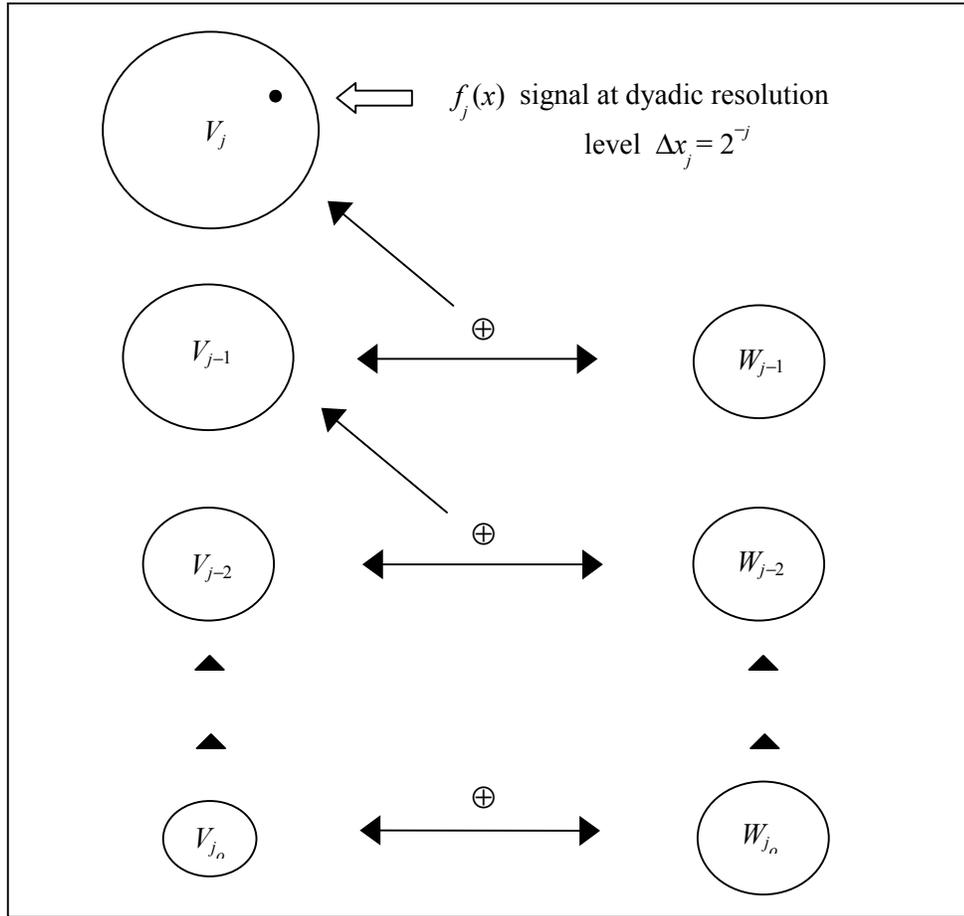


Figure 2.3 Hierarchical decomposition of a multiresolution analysis $\{V_j\}$ in terms of a sequence of ‘detail-wavelet’ orthogonal subspaces $\{W_j\}$

A fundamental theorem, proven by Mallat (1989a,b), states that for every dyadic MRA $\{V_j\}$ there exists a wavelet function $\psi(x)$, such that the family $\sqrt{2^j} \psi(2^j x - n)_{n \in \mathbb{Z}}$ provides an orthonormal basis for every detail (wavelet) subspace W_j associated with this MRA. Furthermore, the collection of all these orthonormal wavelet bases from every

dyadic scale level 2^j will generate a single wavelet basis $\sqrt{2^j} \psi(2^j x - n)_{j,n \in \mathbb{Z}}$ for the whole Hilbert space $L^2(\mathfrak{R})$, according to its orthogonal decomposition in eq.(2.38).

In this way, the study of a signal $f_j(x)$ that belongs in a certain multiresolution subspace $V_j \subset L^2(\mathfrak{R})$ is considerably enriched through its wavelet expansion at coarser resolution levels, according to the form

$$f_j(x) = \sum_n a_n \varphi(2^j x - n) = \sum_{k=-\infty}^{j-1} \sum_n b_{k,n} \psi(2^k x - n) \quad (2.39)$$

where $\varphi(x)$ is the scaling function (not necessarily orthonormal) of the MRA under consideration. A minimum resolution value 2^{j_0} is usually selected as a coarse signal reference for the wavelet decomposition, i.e.

$$\begin{aligned} f_j(x) &= f_{j_0}(x) + \sum_{k=j_0}^{j-1} \sum_n b_{k,n} \psi(2^k x - n) \\ &= \sum_n c_n \varphi(2^{j_0} x - n) + \sum_{k=j_0}^{j-1} \sum_n b_{k,n} \psi(2^k x - n) \end{aligned} \quad (2.40)$$

where $f_{j_0}(x)$ denotes the orthogonal projection of the original signal $f_j(x)$ onto the lower MRA subspace $V_{j_0} \subset V_j$. A variety of algorithms exist in the wavelet literature for the determination of the reference scaling coefficients $\{c_n\}$ and the wavelet coefficients $\{b_{k,n}\}$ from the values of the initial scaling coefficients $\{a_n\}$ in eq.(2.39). For more details on such computational issues, as well as for extensions in higher dimensions and/or compact signal domains, see Mallat (1998b), Wojtaszczyk (1997) and Daubechies (1992).

One of the important advantages, among many others, of wavelet bases is their ability to ‘capture’ the local signal details at various scale levels. Wavelet kernels $\psi(x)$ generally have a compact form with fast decay, which makes them an ideal tool for spectral analysis procedures in transient signals with high irregular variations. This is in contrast to the classic Fourier methods of harmonic analysis, where the underlying base functions (sinusoids) have an infinite support and uniform behaviour across their domain, and thus they cannot isolate any local features in the signal.

Although wavelet signal analysis is a very interesting and important topic that has started to receive increasing interest for many geodetic applications, it will not be discussed further here. In the following chapters, we will deal exclusively with the multiresolution analysis concept which basically corresponds to one of the many different facets of

wavelets. Due to this connection, however, their implicit association with all the forthcoming theoretical developments should always be kept in mind.

Chapter 3

FROM DETERMINISTIC COLLOCATION TO MULTIRESOLUTION APPROXIMATION

The purpose of this chapter is twofold. First, the general problem of linear signal approximation in arbitrary Hilbert spaces is reviewed. The formulation and the solution of this problem are presented in a very detailed way, emphasizing some important aspects that are often overlooked in many textbooks and monographs found in the geodetic literature. A relatively recent mathematical tool of functional analysis, called a *frame*, is also used to analyze the stability of the solution algorithm for the linear approximation problem. The second main scope of the chapter is to study the relationship between the Hilbert space in which we choose to perform the linear approximation, and the stability/convergence properties of its solution algorithm for increasing data resolution. In particular, it will be shown that a constantly stable and convergent (in the sense of infinitely dense data) linear signal estimation scheme in a Hilbert space H requires the incorporation of a multiresolution (MR) structure of subspaces in H , similar to the one introduced in the previous chapter.

3.1 What is Collocation ?

Before we study the linear approximation problem in Hilbert spaces and its associated stability and convergence issues, a descriptive (yet brief) overview of the general *collocation concept* in physical geodesy will be given in this section. Collocation is widely known as an optimal linear estimation method for gravity field modelling using discrete (and possibly heterogeneous) data. Behind this vague definition lie two fundamentally distinct viewpoints, with correspondingly different mathematical and physical models/assumptions associated with them. Both approaches have certain advantages and drawbacks, and they have been the subject of extensive debate in the geodetic scientific community over the past three decades.

3.1.1 Deterministic Collocation

The first approach, which will be called deterministic collocation, uses a functional analysis setting with the unknown field under consideration f (e.g. the anomalous potential of the Earth) being modelled as an individual element in an infinite-dimensional reproducing kernel Hilbert space (RKHS) H . The available discrete observations from the field are also modelled as continuous linear functionals $L_j f$, belonging in the dual Hilbert space H' of H . The optimally approximated field \hat{f} is then considered as the smoothest function, in the topology of H , that satisfies the given functionals. The situation corresponds to a typical inverse problem with no unique solution initially (ill-

posed), which is then regularized according to a simple Tikhonov-type projective scheme (Schwarz, 1979; Moritz, 1980). Such a linear approximation method is not of course an exclusive privilege of physical geodesy problems, but it has been borrowed from other areas of mathematics where it is often found under the name *minimum-norm interpolation in RKHS* (Davis, 1975). The original idea for using such deterministic Hilbert space methods in gravity field modelling is due to Torben Krarup, who in his famous publication (Krarup, 1969) developed a pioneering framework for solving discrete boundary value problems (BVPs) in Hilbert spaces of harmonic functions outside a certain spherical approximation of the Earth. However, the reformulation of the gravity field determination problem as an underdetermined/discrete BVP, instead of a continuous BVP in the sense of classic potential theory (i.e. Stokes, Molodensky), whose solution should employ only the finite number of the available discrete data in a certain optimal fashion, was made earlier than Krarup's developments by Bjerhammar (1964). The interpolatory character of Bjerhammar's discrete BVP, along with his proposed idea for the analytical downward reduction of the discrete observations on a certain reference sphere, have contributed in several ways to a clarification of the conceptual foundations in modern physical geodesy; see also Bjerhammar (1973, 1975, 1987). More details on the deterministic aspects of collocation can be found in Dermanis (1976), Krarup (1978), Leglemann (1979), Meissl (1976), Moritz (1980) and Tscherning (1978b, 1986).

Deterministic collocation 'suffers' from two important problems, one of which is the so-called *norm choice (or reproducing kernel choice) problem*; see Dermanis (1977). In

order to use the method and to actually compute an estimate for the unknown field from its discrete data, we need to know the reproducing kernel in the Hilbert space H , which in turn requires the prescription of a specific norm (topology) or inner product. An a-priori choice for the norm, inner product, or reproducing kernel is (to a certain extent) arbitrary, and not only does it affect the physical interpretation of the signal approximation, but it also controls other important aspects like the stability of the solution algorithm, as well as the admissible spatial configurations of the data needed to obtain such stable solutions (Eeg and Krarup, 1973; Rummel et al., 1979). These issues will be explored further later in this chapter.

The second important problem in deterministic collocation stems from the lack of efficient measures to evaluate the accuracy of the minimum-norm interpolation algorithm. Although there do exist rigorous upper bound values for the error norm $\|f - \hat{f}\|_H$, that can characterize the overall performance of the linear estimation procedure in a given Hilbert space H , their use is of rather limited practical importance and their actual computation requires the complete knowledge of the unknown field itself (Dermanis, 1976; Tscherning, 1986). Furthermore, the choice of a specific signal norm, reproducing kernel, or inner product directly influences the nature of the accuracy measure $\|f - \hat{f}\|_H$, which may not necessarily admit a practically useful interpretation (e.g. RMS-type approximation error). We will return to the accuracy evaluation issues in the following chapters.

3.1.2 Stochastic-Probabilistic Collocation

The second fundamental viewpoint in linear approximation problems of physical geodesy constitutes the stochastic version of collocation. According to this approach, the true field f is modelled as a zero-mean stochastic process (after a possible subtraction of a simple trend model) and the available discrete observations are considered as zero-mean random variables that are linearly related to the unknown random field. The optimal approximation (or more precisely, prediction) \hat{f} is now defined as the one satisfying the minimum mean square error (MMSE) principle, i.e. $E\{(f - \hat{f})^2\} = \min$, where E denotes the expectation operator in a probabilistic sense. The final solution is obtained by additionally requiring that \hat{f} is an unbiased estimator of f , which should be linearly related to the available discrete (stochastic) data.

As in the deterministic collocation case, this probabilistic estimation method has also been borrowed from other areas of mathematics and applied sciences (signal analysis, communication engineering) where it was originally developed. The underlying framework is formally known as the *Wiener-Kolmogorov* (W-K) linear prediction theory, pointing to the pioneering work of the two scientists back in the 1930s and 1940s. An excellent review paper on many different aspects of W-K theory, with more than 350 relevant references, is Kailath (1974). The original introduction of the W-K methodology in physical geodesy estimation problems should be attributed to the work of Moritz (1962) on optimal linear interpolation of gravity data. For a detailed treatment of the

stochastic principles in collocation theory and related applications in gravity field modelling, see Dermanis (1976), Dermanis and Sanso (1997), Bjerhammar (1982), Moritz (1970, 1980), Sanso (1986) and Rapp (1978).

For the computation of the optimal prediction \hat{f} in the W-K theory it is required to know the covariance (CV) function $C(P, Q) = E\{f(P)f(Q)\}$ of the unknown signal, which describes the average behaviour of f at specific points (or pairs of points) in its domain. Often, in practical applications, the unknown field is additionally modelled as a stationary and ergodic stochastic process. The main benefit of such an assumption is that the estimation of the (generally) unknown signal CV function becomes possible using the available realization of the unknown field (i.e. the discrete observations). Furthermore, in the stationary case it is highly beneficial to transform the approximation problem to the frequency domain using the Fourier transform, since the computational effort of the solution algorithm is significantly reduced due to the de-correlation ('whitening') property of the Fourier transform over stationary random signals; for more mathematical details, see Parzen (1967), Papoulis (1993), Priestley (1981), and for geodetic applications of frequency-domain collocation, see Eren (1980), Schwarz et al. (1990), Sideris (1995), Sanso and Sideris (1997) and Nash and Jordan (1978).

The main drawback of the probabilistic viewpoint in collocation is that it is not physically acceptable, since the external gravity field of the Earth is not a stochastic phenomenon as

the W-K theory requires. Excluding time-dependent variations (which are very weak compared to the mean spatial power of the ‘steady’ gravity field and which nowadays can be modelled and computed with very high accuracy) and random noise effects, repetitive gravity field measurements should always give the same result. For an interesting, as well as amusing, discussion on this aspect, see Moritz and Sanso (1980). Furthermore, treating the deterministic gravity field as a stochastic phenomenon creates important problems in the interpretation of the accuracy measures that we often use to evaluate the quality of the W-K linear prediction algorithm. The formalism of variance-covariance propagation allows us to easily obtain the variances and covariances of the prediction error, $e(P) = f(P) - \hat{f}(P)$, at the points where the prediction is applied. Such accuracy information has a purely probabilistic nature (i.e. average error over many experiment repetitions), that is meaningless in a physically deterministic/causal system. Note that all the discussions so far, as well as for the rest of this chapter, are referred to a noiseless discrete data setting.

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

3.1.3 Spatio-Statistical Collocation: A Compromise

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

$$C(P,Q) = M_{\tau} \{ f(P+\tau) f(Q+\tau) \} \quad (3.1)$$

where the operator M is applied to the translation parameter τ only. A common choice for M , which actually allows the efficient incorporation of Fourier methods into the statistical collocation framework, is the integral

$$C(P,Q) = \int f(P+\tau) f(Q+\tau) d\tau \quad (3.2)$$

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

$$C(P,Q) = C(P-Q) = C(\xi) = \int f(x) f(x+\xi) dx \quad (3.3)$$

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

As in the stochastic collocation case, the optimal signal approximation \hat{f} is again defined as the one satisfying a MMSE principle, which is now expressed in a purely deterministic spatio-statistical manner. In the 1D case, the optimal estimation criterion of statistical collocation takes the following form:

$$\frac{1}{|\Omega|} \int_{\Omega} M_{\tau} \{ e(P+\tau, x_o) e(Q+\tau, x_o) \} dx_o = C_e(P, Q) = \min \quad (3.4a)$$

where the estimation error has the general expression

$$e(P, x_o) = f(P) - \hat{f}(P, x_o) \quad (3.4b)$$

The above optimal principle takes into account the fact that, if we translate by x_o the available spatial configuration of the data points (but not the unknown field itself), we will generally obtain some new observation values that will produce a *different* approximation for the unknown field. In this way, the estimation error at an arbitrary

point P becomes also a function of the ‘position’ x_o of the data point configuration, with respect to the reference system used to describe the unknown field. The spatio-statistical collocation solution will then minimize the *mean error CV function* $C_e(P, Q)$ over all possible positions Ω of the given data point geometry. Note that for multi-dimensional problems, the position of the data point network is not determined solely through a single translation parameter x_o , but it includes additional translation and rotational parameters as well. Such a situation is illustrated in Figure 3.1, where we can see two different locations for the same geometry of a planar data point network. The network shown with dotted lines is just a translated (over the spatial coordinates x and y) and rotated (at 90° angle) version of the data point network shown with solid lines.

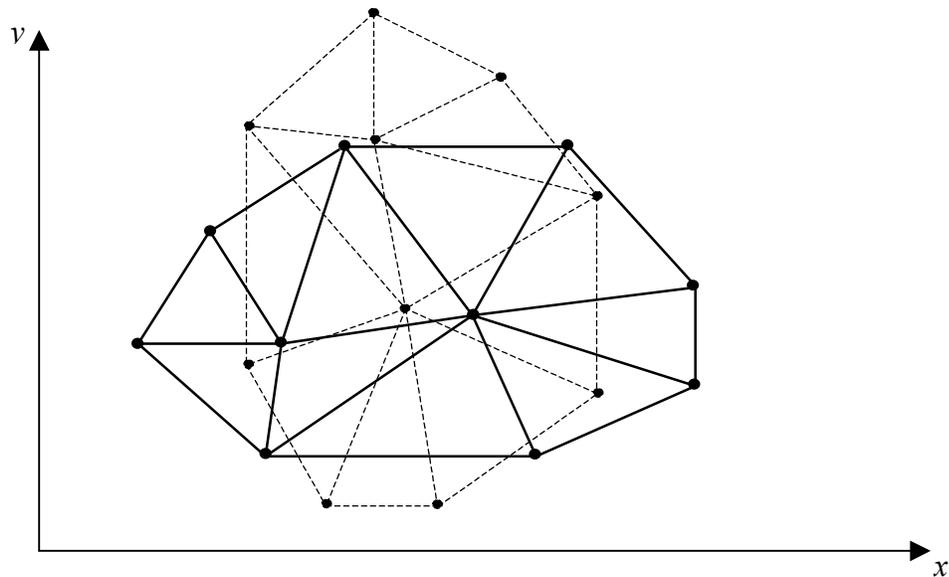


Figure 3.1 Different ‘positions’ for the same geometry of a 2D data point network

Two additional conditions are taken into account in the statistical collocation framework, which are: (i) linearity of the solution $\hat{f}(P)$ with respect to the discrete input data, and (ii) translation-invariance of the solution with respect to corresponding changes in the origin of the reference system. The last property will basically ensure that, if $\hat{f}(P)$ is the optimal linear approximation for the unknown field $f(P)$, then $\hat{f}(P + \alpha)$ should be the corresponding optimal linear approximation for the field $f(P + \alpha)$. In the multi-dimensional case, we should assume approximation invariance under more general rigid transformations of the reference system, which may possibly include rotational parameters. The use of this spatio-statistical version of collocation was proposed in the classic book of Heiskanen and Moritz (1967), but it was actually Sanso (1978) who first formulated a rigorous and complete mathematical setting for the method. A relevant and extensive discussion can also be found in Moritz (1978a, 1980). The statistical collocation framework will be studied in detail in Chapter 4, for the special case of regularly gridded data.

The general formula for the estimated field under the statistical collocation concept is similar to the ones obtained under either the deterministic or the stochastic approach, where now instead of a reproducing kernel or a probabilistic CV function, we use a spatial CV function that is defined in a purely deterministic sense as per eq.(3.1). In this way, the use of the spatio-statistical MMSE principle in eq.(3.4) automatically gives rise to a minimum-norm solution in a Hilbert space with reproducing kernel equal to the

signal spatial CV function. For an interesting topological paradox that exists during this identification procedure, see Tscherning (1977).

3.1.4 Some Important Existing Problems

Regardless of which of the three previous approaches we adopt, the linear estimation of an unknown field from discrete data always requires the inversion of an $N \times N$ symmetric matrix, where N is the number of the available observations. Usually, such a numerical task creates important problems in terms of the required computational effort and the stability of the solution algorithm. Although the computing time/storage requirements can be significantly reduced through special algorithmic and modelling techniques (Bottoni and Barzaghi, 1993; Sanso and Schuh, 1987) and/or fast Fourier transform (FFT) methods (Sideris, 1995; Schwarz et al., 1990; Eren, 1980; Thomas and Heller, 1976), the stability problem is not generally overcome by them. For example, if a high-resolution data point configuration is used, relative to the spread (correlation length) of the selected reproducing kernel (or CV function), then the resulting matrix will be highly ill-conditioned regardless of the domain in which we perform its inversion. In this way, the establishment of convergence for the collocation algorithm, as the data density increases, also becomes very difficult from a numerical/computational point of view.

Apart from the stability/convergence problem, there exists one more issue of special importance in collocation. The introduction of the spatio-statistical estimation principle in

eq.(3.4) by Sanso has been perceived by many geodesists only as an attempt for assigning a non-stochastic interpretation to Moritz's (or Wiener's) optimal prediction formulas, overcoming in this way the non-existence of a stochastic gravity field. Along with this perception, however, it has also remained the false belief that we need to model the gravity field as a stationary signal because of the 'stationary' form of Sanso's spatial CV function. This is believed to provide serious limitations in the whole approximation procedure of statistical collocation, since the actual behaviour of gravity field signals is 'non-stationary'. Besides the fact that the above claim is meaningless, since the notion of stationarity has no place in deterministic signals (see also the related discussion in Sanso, 1978), it is the author's opinion that the spatio-statistical formulation of collocation is a very powerful and complete tool as it is, without having the need to be considered as a 'supplement' to W-K theory for physical geodesy approximation problems. Furthermore, it is one of the main goals of this thesis to show that the use of 'stationary' signal CV functions, as in eq.(3.3), not only does not impose any assumption of uniform behaviour for the underlying unknown fields, but is actually closely related to one of the best mathematical tools available today for localized ('non-stationary') signal analysis and estimation.

From the above two important problems in collocation theory (stability/convergence and stationarity), the first is explored in this chapter whereas the second is discussed in Chapter 4. Our starting point will be the deterministic viewpoint in collocation, from

which we will gradually arrive at a spatio-statistical MR stable version, without the need of using the intermediate stochastic ‘burden’.

3.2 Linear Approximation in Hilbert Spaces

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

$$b_n = \langle f, g_n \rangle, \quad n \in \Gamma \quad (3.5)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in the Hilbert space H , and g_n are known elements from the same space. The index set Γ will be assumed, for the moment, as a finite subset of the integers (i.e. only a finite amount of observations is available). The problem is how to recover the unknown signal f using the observation values b_n and the observational representers g_n .

Let us denote by V the Hilbert subspace of H which is defined as the linear span of the finite sequence $\{g_n\}_{n \in \Gamma}$. The unknown function can now be uniquely decomposed into two orthogonal elements as follows:

$$f = f_V + f_{V^\perp} \quad (3.6)$$

where f_V is the orthogonal projection of f onto V , and f_{V^\perp} is the orthogonal projection of f onto the Hilbert subspace V^\perp (orthogonal complement of V in H , i.e. $V \oplus V^\perp = H$). In this way, the general observation equation (3.5) takes the following form:

$$\begin{aligned}
 b_n &= \langle f_V + f_{V^\perp}, g_n \rangle \\
 &= \langle f_V, g_n \rangle + \langle f_{V^\perp}, g_n \rangle \\
 &= \langle f_V, g_n \rangle
 \end{aligned} \tag{3.7}$$

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

$$\hat{f} = f_V = \sum_{n \in \Gamma} a_n g_n \tag{3.8}$$

where a_n are unknown coefficients with respect to the sequence that generates the actual **solution space** V of the linear approximation problem. Using (3.8), the observation equation (3.7) yields

$$b_n = \langle \sum_{k \in \Gamma} a_k g_k, g_n \rangle = \sum_{k \in \Gamma} a_k \langle g_k, g_n \rangle \quad , \quad n \in \Gamma \quad (3.9)$$

or, by using matrix notation

$$\begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} = \begin{bmatrix} \langle g_1, g_1 \rangle & \cdots & \langle g_1, g_N \rangle \\ \vdots & \ddots & \vdots \\ \langle g_N, g_1 \rangle & \cdots & \langle g_N, g_N \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} \quad (3.10a)$$

$$\mathbf{b} = \mathbf{G} \mathbf{a} \quad (3.10b)$$

where N denotes the finite number of the available observations. It is also useful to express the original observation equations in the following operatorial form:

$$\mathbf{b} = U P_V f = U f_V \quad (3.11)$$

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

$$U : V \rightarrow l^2(\Gamma) \quad , \quad f_V \rightarrow \langle f_V, g_n \rangle \quad \forall n \in \Gamma \quad (3.12)$$

The role of the above operator is to take every function from the subspace V and to compute its inner products with respect to the sequence $\{g_n\}_{n \in \Gamma}$. Since the latter corresponds to a finite set of functions, the collection of all these inner products (for

every $f_V \in V$) can be considered as an individual element \mathbf{b} of the Hilbert space $l^2(\Gamma)$. Note that the Hilbert space of real square-summable sequences over the index set Γ is essentially identical with the classic Euclidean space \Re^N of N -dimensional real vectors, where N is the cardinality of Γ . Also, since the family $\{g_n\}_{n \in \Gamma}$ spans the whole subspace V , the operator U in eq.(3.12) will always be an **injective** (one-to-one) operator.

The computation of the recoverable part f_V of the unknown field requires the inversion of the operator U , or equivalently the inversion of the symmetric $N \times N$ Gramian matrix \mathbf{G} in eq.(3.10). In order to perform such an inversion, we will distinguish between the following two cases regarding the behaviour of the observational representers:

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

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

A geometrical interpretation of the above two cases is illustrated in Figure 3.2. The only difference between them is that the linear operator U is **surjective** (onto) in the first case, whereas in the second case it is not. The internal correlations existing between the observation values $\langle f, g_n \rangle$, when the set $\{g_n\}$ is linearly dependent, make the reproduction of *any* N -tuple of real numbers in the observation space impossible. The extra dots that are left unconnected in Figure 3.2(b) represent those elements in $l^2(\Gamma)$

which do not belong in the range of the ‘observational’ operator U for *Case 2*. Since we are only interested in the inversion of the operator U , we can also create the representative diagram shown in Figure 3.3.

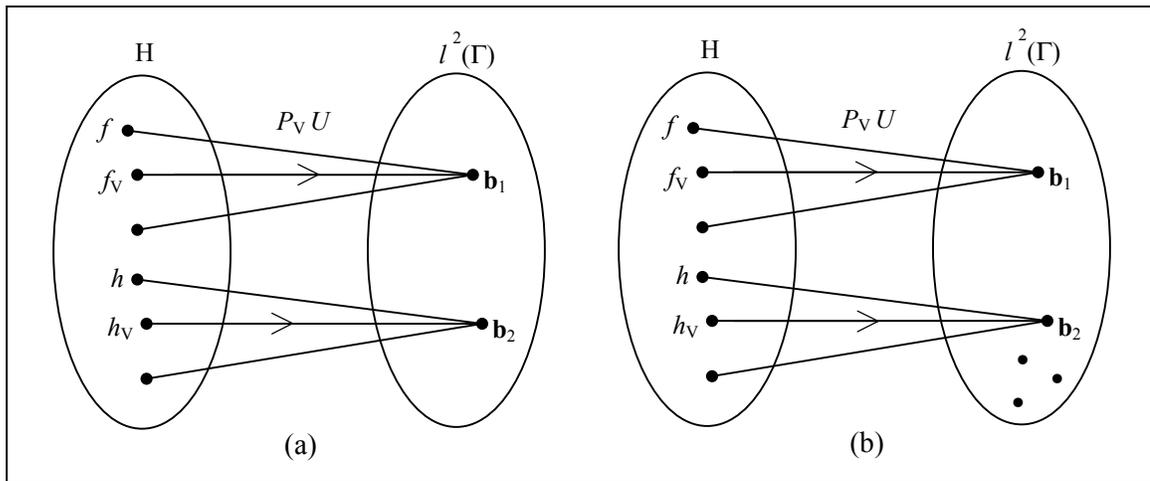


Figure 3.2 Mapping type of the observation equations for *Case 1* (a) and *Case 2* (b)

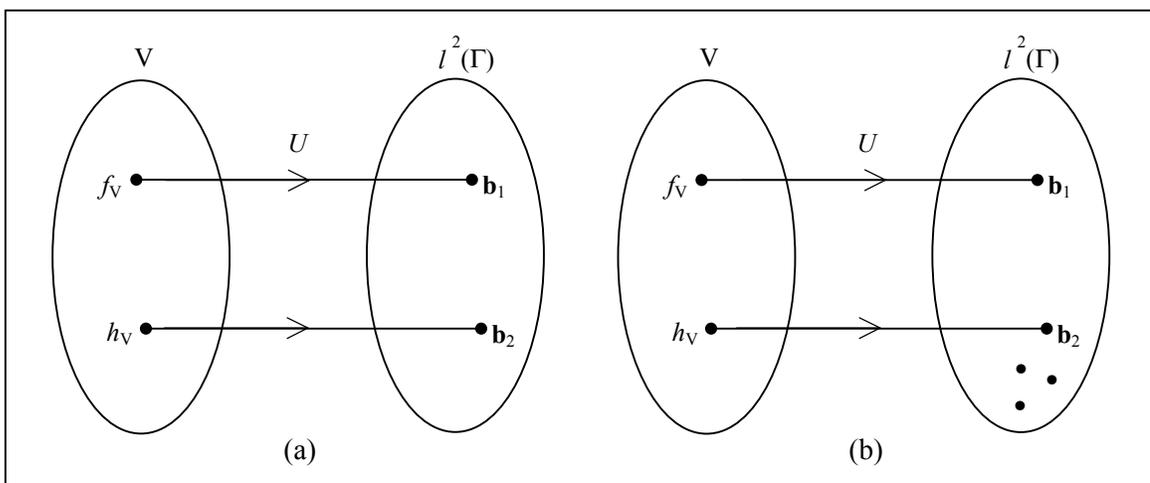


Figure 3.3 The geometry of the linear operator U for *Case 1* (a) and *Case 2* (b)

In both cases there will exist a unique pseudo-inverse operator \tilde{U}^{-1} , which can be defined on the image space $\mathbf{Im}U$ of the linear operator U (i.e. $\mathbf{Im}U$ is the actual observation space of the linear approximation problem). The basic property of the pseudo-inverse is that, when applied to elements of $l^2(\Gamma)$ that do not belong in $\mathbf{Im}U$ (*Case 2*), it becomes a zero operator (Mallat, 1998b; pp. 130-131). In the sequel, the inversion scheme is analyzed for each case separately. The properties of the linear approximation \hat{f} for the unknown field will also be discussed and explained in detail.

3.2.1 Inversion Scheme for Case 1

In the case of linear independence for the observational representers, the pseudo-inverse \tilde{U}^{-1} corresponds to the usual inverse operator U^{-1} . The latter is well defined, since U is always a bijective (surjective + injective) operator.

The numerical computation of the linear approximation for the unknown field is achieved by simply inverting the Grammian symmetric matrix \mathbf{G} in (3.10), from which we can finally determine the (unique) values of the unknown coefficients a_n that are needed in (3.8), i.e.

$$\mathbf{a} = \mathbf{G}^{-1} \mathbf{b} \tag{3.13}$$

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

3.2.2 Inversion Scheme for Case 2

In the case of linear dependence for the observational representers, the pseudo-inverse operator \tilde{U}^{-1} will not correspond to U^{-1} in the usual sense, because the latter does not exist in this case (U is not surjective). However, since U is always an injective linear operator, we can uniquely compute one of its infinite left-inverses which will also be identical with its pseudo-inverse (Mallat, 1998b; p. 130). The pseudo-inverse operator of U in this case will have the form

$$\tilde{U}^{-1} = (U^*U)^{-1} U^* \quad (3.14)$$

where U^* denotes the adjoint of U . Since the Grammian matrix \mathbf{G} is singular in this case, the numerical computation of \hat{f} can be achieved by a minimum-norm solution for the singular linear system of eq.(3.10). Such a solution is obtained by simply computing the unique pseudo-inverse (Moore-Penrose inverse) \mathbf{G}^+ of the $N \times N$ matrix \mathbf{G} , i.e.

$$\hat{\mathbf{a}} = \mathbf{G}^+ \mathbf{b} \quad (3.15)$$

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

$$\mathbf{G}^+ = (\mathbf{G} + \mathbf{E}^T \mathbf{E})^{-1} - \mathbf{E}^T (\mathbf{E} \mathbf{E}^T)^{-2} \mathbf{E} \quad (3.16)$$

where \mathbf{E} is an $r \times N$ matrix whose rows are some linearly independent solutions of the singular homogeneous system $\mathbf{G}\mathbf{a} = \mathbf{0}$, and r corresponds to the rank deficiency of \mathbf{G} .

This matrix is not uniquely defined and it always satisfies the condition $\mathbf{G}\mathbf{E}^T = \mathbf{0}$. A particularly interesting and useful formula, that has been extensively used by various routines for the numerical computation of \mathbf{G}^+ , is the following (Albert, 1972):

$$\mathbf{G}^+ = \lim_{\lambda \rightarrow 0} (\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{I})^{-1} \mathbf{G}^T = \lim_{\lambda \rightarrow 0} \mathbf{G}^T (\mathbf{G} \mathbf{G}^T + \lambda^2 \mathbf{I})^{-1} \quad (3.17)$$

The observational representers $\{g_n\}_{n \in \Gamma}$ in *Case 2* provide a **redundant** system of base functions for the solution space V , which is not a basis. The signal solution $\hat{f} = f_V$ can thus be expressed in an infinite number of ways with respect to the known family $\{g_n\}$.

The use of the pseudo-inverse operator \tilde{U}^{-1} , through eq.(3.15), gives just a single set $\hat{\mathbf{a}}$ among all admissible coefficients \mathbf{a} satisfying the system (3.10), which has the minimum-norm property, i.e.

$$\|\hat{\mathbf{a}}\| = \sqrt{\hat{\mathbf{a}}^T \hat{\mathbf{a}}} \leq \|\mathbf{a}\| = \sqrt{\mathbf{a}^T \mathbf{a}} \quad (3.18)$$

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

$$\forall \mathbf{b} \in (\mathbf{Im}U)^\perp, \mathbf{Im}U \oplus (\mathbf{Im}U)^\perp = l^2(\Gamma) \quad \text{then} \quad \tilde{U}^{-1} \mathbf{b} = \mathbf{0} \quad \text{or} \quad \mathbf{G}^+ \mathbf{b} = \mathbf{0} \quad (3.19)$$

For more details, see Mallat (1998b, Chapter 5). Note again that the Hilbert space $l^2(\Gamma)$ of real square-summable sequences over the index set Γ is essentially identical with the classic Euclidean space \Re^N of N -dimensional real vectors, where N is the cardinality of Γ .

3.2.3 Comments

The general solution of the linear approximation problem, given by eq.(3.8), obeys the **minimum norm principle** (deterministic collocation) in the sense that

$$\|\hat{f}\| \leq \|f\| \quad (3.20)$$

for every other function $f \in H$ that satisfies the given observation equations in eq.(3.5).

In the last equation, the symbol $\|\cdot\|$ denotes the norm in the Hilbert space H . This property is trivial to prove by simply taking into account the decomposition formula (3.6)

and the generalized Pythagorean theorem in Hilbert spaces. In fact, if we had started the development of our solution procedure by imposing a-priori the following optimality criterion for the estimated field \hat{f} :

$$\|\hat{f}\| = \min \tag{3.21}$$

then we would have derived the exact same system of normal equations as in (3.10). The unique solution $\hat{f} \in H$ that satisfies both the minimum norm principle (3.21) and the observation equations (3.5) will always belong to the Hilbert subspace $V \subset H$, and it will thus have the general expression of eq.(3.8); see Tscherning (1986). However, when the observational representers are linearly dependent (**Case 2**) then the expansion coefficients of this expression are not uniquely defined, and an additional minimum-norm solution for them should be computed according to (3.15). It should be emphasized that the *uniqueness of the linear approximation* $\hat{f} = P_V f$ is always assured, regardless of the linear dependence/independence of the finite set of the observational representers.

It is quite interesting to realize that the minimum norm principle (3.21), which has been used extensively in geodesy up to date, is more of a by-product of the Hilbertian ‘geometrical’ setting for the linear approximation problem, rather than an arbitrary optimal estimation criterion of maximum smoothness for the unknown field in a given topology. Also, in the geodetic literature the problem of linear approximation for gravity

field signals in Hilbert spaces is exclusively treated for the *Case 1* only (see, e.g. Tscherning, 1986; Moritz, 1980; Meissl, 1976). The *Case 2*, where the observational representers are linearly dependent, can also be easily included in the Hilbertian approximation framework by simply computing the pseudo-inverse of the singular Grammian matrix \mathbf{G} , as was explained in the previous paragraphs. Numerical stability problems that may occur during the computation of the matrices \mathbf{G}^{-1} or \mathbf{G}^+ and other related issues are discussed in section 3.3.

3.2.4 Modelling Considerations

The previous presentation reveals in an interesting way one aspect of the *modelling choice problem*, which emerges when we use the deterministic collocation concept in actual practical applications. The starting point in such cases is just a given N -tuple of real numbers $\{b_n\}_{n \in \Gamma}$, corresponding to the discrete observations obtained from the unknown field. As a second step, a certain modelling choice for f will enable us to construct linear observation equations of the form (3.5), which are then solved according to the previous methodology. In order for this methodology to work, however, we have to ensure that the given N -tuple of observations belongs to the image space $\mathbf{Im}U \subseteq l^2(\Gamma)$ of the linear operator U . This, in turn, depends solely on our modelling choice, i.e. the form of the inner product $\langle \cdot, \cdot \rangle$ and the type of the observational representers g_n .

Assuming that the observations b_n correspond to point values of the unknown field itself, we can then express the observational representers in the following general form:

$$g_n(P) = K(P, Q_n)$$

where $K(P, Q)$ is the reproducing kernel of the Hilbert space H chosen to model the unknown field, and $\{Q_n\}_{n \in \Gamma}$ is the data point configuration. In this way, the interpolation modelling is exclusively based on the selection of a positive-definite, symmetric, bivariate function. If the set $K(P, Q_n)_{n \in \Gamma}$ is linearly independent, then our observation sequence (regardless of its numerical values b_n) will certainly belong in the image space $\mathbf{Im}U \equiv l^2(\Gamma)$. The linear estimation algorithm will always produce a proper numerical result, but our modelling choice can neither be questioned, nor validated in this case!

However, if we repeat the same modelling choice using denser data point configurations $\{Q'_n\}_{n \in \Gamma'}$, we should expect that the new set of the observational representers $K(P, Q'_n)_{n \in \Gamma'}$ starts to become linearly dependent. In order for the denser sequence of observation values $\{b'_n\}_{n \in \Gamma'}$ to remain in $\mathbf{Im}U' \subset l^2(\Gamma')$, the unknown field must exhibit a certain ‘correlated’ behaviour induced by our interpolating modelling choice. In simple words, the unknown field f *has to* belong in the Hilbert space with the specific

reproducing kernel. For a given approximation model $K(P, Q)$, the level of the imposed correlation in the observation values will generally increase as the data point density (resolution) increases.

Hence, an essential factor in the modelling procedure becomes the minimum spatial resolution of data points $\{Q_n\}$ above which the unknown field should start to exhibit an increasing correlation in its values. The answer to this question (if any) will still leave a relative freedom in choosing a specific model $K(P, Q)$ for measuring this correlation, whose unique selection requires the incorporation of additional criteria.

3.3 Numerical Stability and the Role of Frames

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

$$\hat{f} = f_V = \tilde{U}^{-1} \mathbf{b} \quad (3.22)$$

Note that there is no need for the observational representers to be linearly independent in order to have a stable approximation scheme, since the continuity of the operator \tilde{U}^{-1} does not depend on the linear dependence/independence of the family $\{g_n\}_{n \in \Gamma}$, as it will be explained in the following sections. Think, for example, of the case of three unit vectors lying on the same plane and forming angles of 120° between each other. At least our intuition should tell us that this linearly dependent set corresponds to a very stable and robust planar coordinate system, which may be more preferable than a coordinate system formed by two linearly independent vectors with an angle of 20° between each other.

3.3.1 General Frame Theory

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

- (i) every $f \in V$ is uniquely determined by its projections $\langle f, \phi_j \rangle$
- (ii) the reconstruction of f from the values $\langle f, \phi_j \rangle$ is a stable algorithmic process

The rigorous mathematical definition of a frame $\{\phi_j\}$ in a Hilbert space V is based on the following formula:

$$0 < A \|f\|^2 \leq \sum_j |\langle f, \phi_j \rangle|^2 \leq B \|f\|^2 < +\infty \quad (3.23)$$

which should be satisfied by every $f \in V$. Note that (3.23) provides a simple generalization of the well known Plancherel-Parseval property for orthonormal bases. The symbols A and B are some strictly positive constants, independent of f , and they are called **frame bounds**. Frame theory was originally developed by Duffin and Schaeffer (1952), in the context of nonharmonic Fourier series for the reconstruction of band-limited signals from their irregularly spaced samples. Some very good introductory references on frame theory can be found in the books by Young (1980), Hernandez and Weiss (1996) and Mallat (1998b). For more advanced details on frame theory and related applications, see Daubechies et al. (1986), Daubechies (1990), Heil and Walnut (1994) and Strohmer (1995).

Every frame in a Hilbert space V is associated with a corresponding **frame operator** F , which is defined as follows:

$$F : V \rightarrow l^2(\mathbb{Z}) , f \rightarrow \langle f, \phi_j \rangle \quad \forall j \in \mathbb{Z} \quad (3.24)$$

where it is assumed, for generality, that the number of the frame components is infinite. The definition of a frame system, according to the double inequality (3.23), ensures that the linear operator F is always: (i) injective, and (ii) bounded (continuous). Furthermore, it can be shown that (3.23) is a necessary and sufficient condition guaranteeing that F is invertible on its image, with a bounded (continuous) inverse; for a proof, see Mallat (1998b, p. 129). If the frame components are all linearly independent elements of V , then the frame operator is also surjective and its image space is the whole Hilbert space $l^2(Z)$. In this special case, the frame is called Riesz basis (see Chapter 2).

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

The reconstruction of an arbitrary signal $f \in V$ from its frame ‘coordinates’ $\langle f, \phi_j \rangle$ is computed with the help of a unique *dual frame* $\{\tilde{\phi}_j\}$, according to the equation (Daubechies et al., 1986)

$$f = \sum_j \langle f, \phi_j \rangle \tilde{\phi}_j = \sum_j \langle f, \tilde{\phi}_j \rangle \phi_j \quad (3.25)$$

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

$$\tilde{\phi}_j = (F^*F)^{-1} \phi_j \quad (3.26)$$

where F^* is the adjoint of F . A proof for the invertibility of the operator F^*F can be found in Mallat (1998b, pp. 130-131). The signal reconstruction formula in eq.(3.25) can also be written in the equivalent operatorial form

$$f = (F^*F)^{-1} F^* F f = \tilde{F}^{-1} F f \quad (3.27a)$$

It can easily be shown that the pseudo-inverse operator $\tilde{F}^{-1} = (F^*F)^{-1} F^*$ corresponds to the usual inverse of F , when the latter is restricted on its image space. Furthermore, \tilde{F}^{-1} is always bounded (continuous) and it has the minimum supremum norm among all possible generalized inverses of F . Its condition number $k(\tilde{F}^{-1})$ satisfies the relation

$$k(\tilde{F}^{-1}) = \|F\|_{sup} \|\tilde{F}^{-1}\|_{sup} = \sqrt{\frac{B}{A}} \quad (3.27b)$$

For more details and proofs, see Daubechies et al. (1986). In simple words, therefore, the fundamental property of a redundant frame is to provide a unique and stable expansion for every signal $f \in V$

$$f = \sum_j a_j \phi_j \quad (3.28)$$

using, among all infinite possibilities for the sets of coefficients a_j , the one that has the smallest $l^2(Z)$ norm. Note that, in general, the frame components will not constitute a *basis* in the technical sense, although their closed linear span is all of V . This is so because $\{\phi_j\}$ need not be independent. On the other hand, a basis for V , which always assumes the expansion form (3.28), is not necessarily stable and, therefore, not a frame.

A number of simple examples can be used to clarify the situation. Let $\{e_j\}_{j=1}^{\infty}$ be an *orthonormal* basis for an infinite-dimensional Hilbert space V . Then

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

Hence, in infinite-dimensional Hilbert spaces a family of vectors may be complete and not yield a stable signal representation. For more examples and explanations, see Strohmer (1995) and Heil and Walnut (1994). Also, one can easily verify that a finite set of signals $\{\phi_j\}_{1 \leq j \leq N} \in V$ is *always* a frame for the Hilbert subspace generated by their linear combinations, regardless of their linear independence/dependence (Mallat, 1998b). This last fact will be used later on in section 3.3.3.

3.3.2 Gabor and Affine Frames

In order to recover a signal from its frame coefficients $\langle f, \phi_j \rangle$, we should first analytically pre-compute the dual frame components $\tilde{\phi}_j$ according to eq.(3.26), and then we can reconstruct f using the sum

$$f = \sum_j \langle f, \phi_j \rangle \tilde{\phi}_j \quad (3.29)$$

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

$$\tilde{\phi}_j = \frac{1}{A} \phi_j \quad (3.30)$$

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

$$\phi_{n,k}(x) = w(x - nu_o) e^{ik\xi_o x} \quad (3.31)$$

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

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

$$\phi_{n,k}(x) = a^{-k/2} w\left(\frac{x}{a^k} - nu_o\right) \quad (3.32)$$

where again the parameters u_o and a depend on the space-frequency localization of the selected window $w(x)$. More details on wavelet frames can be found in Daubechies et al. (1986), Daubechies (1990, 1992) and Benedetto (1994).

Note that the generating window function, as well as the discretization parameters (u_o, ξ_o, a) , should satisfy certain conditions in order for the functions $\phi_{n,k}(x)$ to generate either a Gabor or a wavelet tight frame, and they cannot be selected arbitrarily. In Daubechies et al. (1986) and Daubechies (1990, 1992), sufficient conditions are given to build window functions that generate either a Gabor or a wavelet tight frame in L^2 Hilbert spaces. These conditions allow us to construct generating functions which are: (i) smooth (differentiable), and (ii) well localized in both space and frequency domain sense. The first property is important when we try to study the Gabor or wavelet frame spectrum, since the use of a discontinuous $w(x)$ may introduce artificial high-frequency content in the spectrum and it can distort the image of the analyzed signal. Also, the localization properties of the frame components $\{\phi_{n,k}\}$ allow us to get a closer look at the fine signal details, avoiding a simple and vague ‘global averaging’ that is usually obtained when we use non-localized reconstruction components (e.g. infinite sinusoids in classic Fourier harmonic analysis).

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

In certain applications, the frame components may depend on the specific analyzed signal f , and they cannot be arbitrarily chosen from a pre-determined Gabor or wavelet frame ‘dictionary’. A classic example is the frames associated with irregular sampling problems (Duffin and Schaeffer, 1952; Feichtinger and Grochenig, 1994; Strohmer, 1995), where every frame component depends on the specific location of each sample value $\langle f, \phi_j \rangle$. If the irregular sampling grid varies from signal to signal this modifies the frame vectors, and it is then highly inefficient to compute the dual frame for each new signal. An overview of various numerically efficient, iterative algorithms that can be applied in such cases is given in Mallat (1998b, p. 135).

One of the main practical advantages of overcomplete frame systems is that their redundancy becomes very useful in reducing the effect of possible additive noise in the frame coefficients $\langle f, \phi_j \rangle$, and it significantly increases the robustness of the signal reconstruction algorithm (compared to ‘minimal basis’ reconstruction schemes). These issues have been studied in detail mainly in the signal and image processing community,

where frames are often used for high precision analog-to-digital conversion based on oversampling. For more details, see Benedetto (1998) and Cvetkovic and Vetterli (1998).

3.3.3 Frames and Linear Approximation in Hilbert Spaces

After the preceding brief overview on frame theory, let us now see its relevance with respect to the initial approximation problem of section 3.2. Since a finite set of elements in a Hilbert space is always a frame for the Hilbert subspace generated by their linear combinations, the set of the observational representers $\{g_n\}_{n \in \Gamma}$ will also be a frame for the Hilbert subspace $V \subset H$ (solution space). The linear estimate of the unknown field, according to eq.(3.8), will thus be just an expansion with respect to the ‘observational frame’ $\{g_n\}$. Furthermore, the unknown coefficients of this expansion, which are obtained from the solution of the consistent linear system in (3.10), will correspond to the optimal/stable ones implied by frame theory.

In order to see this better, let us denote by \mathbf{g} an $N \times 1$ vector containing all the observational representers as follows:

$$\mathbf{g} = [g_1 \cdots g_j \cdots g_N]^T \quad (3.33)$$

Each component of the observational frame can be expressed with respect to the unique dual frame $\{\tilde{\mathbf{g}}_n\}$ according to the formula (see section 3.3.1)

$$\mathbf{g}_j = \sum_{n \in \Gamma} \langle \mathbf{g}_j, \mathbf{g}_n \rangle \tilde{\mathbf{g}}_n \quad \forall j \in \Gamma \quad (3.34)$$

or, by using matrix notation to combine all the frame components,

$$\mathbf{g}^T = \tilde{\mathbf{g}}^T \mathbf{G} \quad (3.35)$$

where

$$\tilde{\mathbf{g}} = [\tilde{\mathbf{g}}_1 \dots \tilde{\mathbf{g}}_j \dots \tilde{\mathbf{g}}_N]^T \quad (3.36)$$

and \mathbf{G} is the symmetric $N \times N$ Grammian matrix given in (3.10). The general linear approximation algorithm in (3.8) can be written in the equivalent vector form

$$\hat{f} = f_V = \mathbf{g}^T \mathbf{a} \quad (3.37)$$

which, using (3.35), takes the form

$$\hat{f} = f_V = \tilde{\mathbf{g}}^T \mathbf{G} \mathbf{a} \quad (3.38)$$

Now, if we substitute the product $\mathbf{G}\mathbf{a}$ according to the system in (3.10), we finally get

$$\begin{aligned}
\hat{f} = f_V = \tilde{\mathbf{g}}^T \mathbf{b} &= \sum_{n \in \Gamma} \langle f, g_n \rangle \tilde{g}_n \\
&= \sum_{n \in \Gamma} \langle f_V, g_n \rangle \tilde{g}_n \\
&= \sum_{n \in \Gamma} \langle f_V, \tilde{g}_n \rangle g_n
\end{aligned} \tag{3.39}$$

Hence, the optimal linear approximation of the unknown field is reduced to a simple **frame expansion** for its recoverable part in the finite-dimensional Hilbert subspace $V \subset H$. The linear operator U [see eq.(3.12)] will be a frame operator for the solution space V , and its pseudo-inverse \tilde{U}^{-1} is thus always continuous (stable).

There are two different, but equivalent, ways to view the numerical computation of the signal estimation \hat{f} , which both require the calculation of the pseudo-inverse matrix \mathbf{G}^+ . In particular, we can either compute the projections (frame coefficients) of f with respect to the dual frame $\{\tilde{g}_n\}$ of the observational representers $\{g_n\}$, according to the scheme

$$\mathbf{G}^+ \mathbf{b} \quad \rightarrow \quad \langle f, \tilde{g}_j \rangle = \langle f_V, \tilde{g}_j \rangle \tag{3.40a}$$

or we can directly compute the dual frame components \tilde{g}_j using the product

$$\mathbf{g}^T \mathbf{G}^+ \rightarrow \tilde{g}_j \quad (3.40b)$$

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

$$\hat{f} = \mathbf{g}^T \mathbf{G}^+ \mathbf{b} \quad (3.40c)$$

Of course, when the observational representers are all linearly independent then the Gramian matrix \mathbf{G} will be non-singular, and the last three equations are simplified accordingly by using $\mathbf{G}^+ = \mathbf{G}^{-1}$. Furthermore, in this special case the sets $\{g_n\}$ and $\{\tilde{g}_n\}$ will constitute a pair of biorthogonal Riesz bases for the solution space V .

The use of frame theory shows that the stability of the linear approximation algorithm in an arbitrary Hilbert space H is always guaranteed, as long as we deal with a finite amount of data $\langle f, g_n \rangle$. Although this result may seem satisfactory for practical purposes, it actually raises new interesting questions related to the convergence of the linear approximation \hat{f} towards the true field f , as well as to the original setup of the estimation problem itself (i.e. selection of the Hilbert space H and its associated inner product, type of the observational representers g_n , invariance properties of the solution, etc.). In particular, the convergence problem should not be viewed just as an advanced theoretical ‘detail’, since the highly increasing flow of various observations for the Earth’s gravity field that takes place today makes convergence considerations, within the

linear approximation context, especially important in both theoretical and practical/computational sense.

Regardless of the type of the Hilbert space H in which the signal estimation takes place, a necessary condition for convergence is that the closed linear span V of the observational representers tends to H , as the amount of the discrete data $\langle f, g_n \rangle$ increases. However, the structure of H and/or the structure of the measurement modelling (i.e. form of the various representers g_n) may be such that the observational frame becomes less and less stable as the **data resolution** increases. Although the system of the observational representers will constantly be a frame (for its corresponding linear span = solution space of the linear approximation problem) for every finite set of observations, its bound values A and B may converge rapidly towards 0 and/or ∞ , respectively, for high data point density. This can cause serious numerical problems in calculations using the matrices \mathbf{G}^+ or \mathbf{G}^{-1} , making the whole estimation procedure highly ill-conditioned. More importantly, in the limit where we consider infinitely dense data, the upper frame bound will definitely reach the value $B = \infty$, whereas the lower frame bound A may still converge to a finite positive value. Such a situation will make the recovery of f from infinite-resolution data sets an ill-posed problem! This does not mean, however, that the true field $f \in H$ cannot be uniquely defined from its ‘very dense values’ $\langle f, g_n \rangle$. It just means that the linear approximation algorithm of eq.(3.40c) cannot be reduced to a well defined (stable) signal description when the data resolution increases beyond limit,

and thus convergence cannot be secured in this case. Further discussion on the interplay between stability, convergence and data resolution, within the deterministic collocation framework, will be given in section 3.5.

3.3.4 A Note on Ill-Posed Problems

From the discussion given in the previous sections, it seems that an arbitrary selection of the Hilbert space H (within which we choose to model our unknown signals) is not enough to ensure a well-conditioned linear approximation scheme for any data resolution and, also, it does not necessarily provide convergence to a stable signal expression in the case of infinitely dense data. Although the selected space H may be perfectly suitable to describe some physical system, its ‘structure’ could pose certain limitations on the admissible data point configurations/topologies that can be used for a reasonably stable signal estimation procedure. Generalizing the notion of an ill-posed problem according to Hadamard (1923), we can say that some additional type of ill-posedness may be hidden in the original approximation problem of section 3.2, which is related to the stability behaviour of its solution *for different data resolutions*, as well as to its convergence properties for infinitely dense data.

The general linear procedure that was followed in section 3.2 basically corresponds to a simple Tikhonov-type regularization scheme, which is needed to overcome the non-uniqueness aspect of the solution of the underdetermined estimation problem of eq.(3.5).

Many authors choose to call this minimum-norm regularization methodology (within a certain infinite-dimensional Hilbert space H) the *least-squares solution* for the signal approximation problem. The formalism of frame theory ensured that this projective regularization scheme will always lead to a stable solution algorithm for a given configuration of finite data. But when it comes to considering different possible data configurations with increasing spatial density and related convergence issues, an additional regularization of the initial approximation problem may be feasible. Such specific procedures will be discussed later on in this chapter. For some general alternatives, see Nashed (1976).

3.4 The Hilbert Space Choice Problem

In this section, we will study more closely the general modelling problem for linear signal approximation in Hilbert spaces, which was briefly discussed in the previous paragraphs. Note that the modelling aspects now do not include the problems mentioned at the end of section 3.2.4 (i.e. it is assumed a-priori that the unknown field does indeed belong to the chosen Hilbert space). In particular, we just try to establish a set of conditions satisfied by a global Hilbert space H enclosing all of our unknown signals, which are sufficient to ensure certain desirable properties for the solution of the linear approximation problem in both theoretical and practical sense. Such properties are:

- (i) convergence of the linear approximation \hat{f} to the true unknown field $f \in H$ in the case of infinitely dense data;
- (ii) a numerically stable algorithm for the computation of \hat{f} ; and
- (iii) the level of stability should be independent of the available data point configuration, i.e. the condition number of the pseudo-inverse approximation operator \tilde{U}^{-1} should not worsen as the data resolution increases.

It will be seen that such an ‘optimization’ attempt not only provides significant insight into the limitations of the classic deterministic collocation method but, most importantly, it directs us to a certain *multiresolution* reformulation of the linear estimation problem, through which we can achieve the previous three essential properties.

An additional important property is the invariance of the solution with respect to arbitrary rigid transformations of the reference system. Essentially, this will guarantee the independence of the estimated field \hat{f} from the origin (and orientation) of the reference system used to describe the spatial position of the data points. For 1D approximation problems, these transformations correspond to simple translations of the reference system with respect to the data point configuration, whereas for higher dimensions we should include possible rotations of the reference system as well. In order to achieve such

invariance properties for the solution of the linear approximation problem, the norm in the global Hilbert space H should satisfy a corresponding invariance condition^{*}, i.e.

$$\forall f \in H, \quad \|f(\bullet)\| = \|f(T\bullet)\| \quad (3.41)$$

where T denotes a general rigid transformation in the domain of f . In the 1D case, the last equation takes the form

$$\forall f \in H, \quad \|f(x)\| = \|f(x-\tau)\| \quad (3.42)$$

with τ being an arbitrary real number.

It would seem only natural to require that the linear estimation scheme should also satisfy some kind of *scale-invariance*. For example, we would want the relative approximation

error $\frac{\|f - \hat{f}\|}{\|f\|} = \frac{\|e\|}{\|f\|}$ to be independent of the scale of the used reference system. For

this type of invariance, a reasonable condition on the norm of the Hilbert space H could be, e.g., for the 1D case,

$$\forall f \in H, \quad \|f(x)\| = \frac{1}{\sqrt{|a|}} \left\| f\left(\frac{x}{a}\right) \right\| \quad (3.43)$$

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

where a is some non-zero scaling factor. In the sequel, we will assume that the norm in the global Hilbert space H satisfies both invariance conditions (3.42) and (3.43). The derivation of their multi-dimensional counterparts is straightforward and it will not be given here.

3.4.1 Data Type and Configuration

In the two previous sections, 3.2 and 3.3, the available discrete data from the unknown field $f \in H$ were assumed to have the general inner product form $b_n = \langle f, g_n \rangle$, where g_n are known elements of the global Hilbert space H . This is the usual modelling approach for the observation equations when we consider linear estimation problems in infinite-dimensional Hilbert spaces. The measurements b_n correspond to the values of certain bounded linear functionals $L_n f$ that are applied to the unknown field, and the elements $g_n \in H$ are exactly the *representers* of these functionals according to the well known **Riesz representation theorem** (Moritz, 1980). In order to study the structure of the Hilbert space H , in view of the desired solution properties (i)-(iii), a more specific physical meaning for both b_n and g_n is required. In this way, expressions like ‘infinitely dense data’, ‘increasing data density’, or ‘data point configuration’ could be unambiguously formulated in proper mathematical terms.

The most straightforward case arises when the discrete data represent point values of the unknown field itself (i.e. interpolation problem). In such cases, the observational representers correspond to the reproducing kernel that is associated with the global Hilbert space. A value of the unknown field $f \in H$, at a point Q , can always be expressed in the linear form

$$f(Q) = \langle f(P), K(P, Q) \rangle_P \quad (3.44)$$

where $K(P, Q)$ is the reproducing kernel (r.k) of the Hilbert space H , and the subscript P means that the inner product is calculated using the point P as the independent variable. According to the dimensionality of the problem, the points P and Q will depend on one, two, or more coordinates. In order to keep the notation simple, we will work in the sequel only for the 1D case, where the domain of the signals will be assumed as the whole real line. The extension into higher dimensions and/or compact domains does not produce major conceptual complications, and it will not restrict the generality of the subsequent results.

The consideration of different spatial configurations for the point data values $f(Q_n)$ is also greatly simplified if we work with regularly gridded data, i.e. $f(Q_n) = f(n\Delta x)$. In this way, cases of increasing data resolution, as well as limiting cases of infinitely dense data, can be simply considered by changing the value of the sampling parameter Δx .

Hence, in the following we will deal exclusively with discrete data of an unknown field $f \in H$ that have the gridded form

$$b_n = f(n\Delta x) = \langle f(x), K(x, n\Delta x) \rangle \quad (3.45)$$

where the range (n) of the available values will be generally assumed infinite. Such an assumption basically implies that we have sampled the unknown field, at the given resolution level Δx , over its entire support. In cases of signals with finite support, the observation sequence b_n will consist of a finite number of non-zero values padded with zeros. Note that the domain of the signals is assumed to be the whole real line.

3.4.2 Stable and Convergent Deterministic Collocation (Interpolation Problem)

Let us assume that we have a sequence of gridded point data $f(n\Delta x)_{n \in Z}$ for an unknown field f belonging in an infinite-dimensional Hilbert space H , with reproducing kernel $K(x, y)$. Following section 3.2 and taking into account eq.(3.45), we know that the optimal solution space of the linear approximation problem (for the given data resolution Δx) is the closed linear span $V_{\Delta x} \subset H$ of the set $K(x, n\Delta x)_{n \in Z}$. Since the solution space is a closed linear subspace of the global RKHS H , it will also be a RKHS itself with its reproducing kernel being generally different from $K(x, y)$; see Aronszajn (1950). The

norm and inner product associated with both $V_{\Delta x}$ and H , however, will always be the same.

In order to have a stable linear approximation scheme, we know that the translates $K(x, n\Delta x)_{n \in Z}$ of the reproducing kernel should constitute a frame for their closed linear span. This simply means that every function in the solution space $V_{\Delta x}$ must satisfy the relationship (see section 3.3)

$$0 < A \|g\|^2 \leq \sum_n |\langle g(x), K(x, n\Delta x) \rangle|^2 \leq B \|g\|^2 < +\infty, \quad \forall g(x) \in V_{\Delta x} \quad (3.46a)$$

for some finite frame bounds A and B . We can further simplify the above relationship using the reproducing property of the kernel $K(x, y)$. In this way, for a unique and stable optimal linear approximation from the gridded signal samples, all the functions in the solution space $V_{\Delta x}$ should satisfy the relation

$$0 < A \|g\|^2 \leq \sum_n |g(n\Delta x)|^2 \leq B \|g\|^2 < +\infty, \quad \forall g(x) \in V_{\Delta x} \quad (3.46b)$$

The left part of the above inequality implies that the signals of the solution space, which take zero values at all data points, should all be topologically equal to the zero function. As a result, for any reasonable norm choice in H , the solution space should not contain functions like the one shown in Figure 3.4.

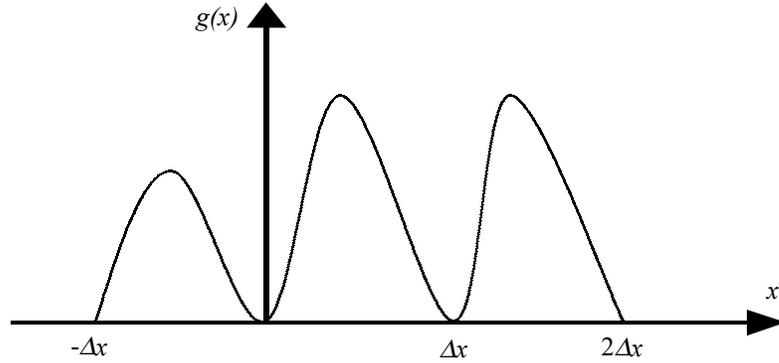


Figure 3.4 Unacceptable function for the solution space $V_{\Delta x}$ in the case of stable optimal linear approximation using discrete samples with resolution Δx

As the data configuration changes (i.e. varying sampling interval Δx), the optimal solution space for the linear approximation problem changes accordingly. Let us denote by $V_{\Delta x'}$ the closed linear span of the set $K(x, n\Delta x')_{n \in \mathbb{Z}}$, which is the optimal solution space corresponding to a new data sequence obtained from the unknown field with sampling resolution $\Delta x' \neq \Delta x$. Obviously, in order to continue having a unique and stable signal solution from the new data set $f(n\Delta x')_{n \in \mathbb{Z}}$, all the functions in the new solution space should satisfy the following relationship:

$$0 < A' \|h\|^2 \leq \sum_n |h(n\Delta x')|^2 \leq B' \|h\|^2 < +\infty, \quad \forall h(x) \in V_{\Delta x'} \quad (3.47)$$

where A' and B' are some new frame bounds, generally different from A and B given in eq.(3.46b). Both frame conditions in the last two equations are associated with two corresponding frame operators $U_{\Delta x}$ and $U_{\Delta x'}$, where the underlying frame components

are assumed to be $K(x, n\Delta x)_{n \in Z}$ and $K(x, n\Delta x')_{n \in Z}$, respectively. In each case, the optimal linear estimation (minimum-norm solution) of the unknown field $f \in H$ from its discrete samples is determined through the continuous pseudo-inverses $\tilde{U}_{\Delta x}^{-1}$ and $\tilde{U}_{\Delta x'}^{-1}$, according to eq.(3.22).

The condition number of the linear approximation operators $\tilde{U}_{\Delta x}^{-1}$ and $\tilde{U}_{\Delta x'}^{-1}$, gives a measure of the sensitivity of the optimal solution \hat{f} with respect to small perturbations in the discrete data values. A large condition number implies a numerically ill-conditioned problem, which may have a strong effect on the accuracy of the solution. The condition number of each of these operators is given by the following formulas:

$$\begin{aligned}
 k(\tilde{U}_{\Delta x}^{-1}) &= \|U_{\Delta x}\|_{sup} \|\tilde{U}_{\Delta x}^{-1}\|_{sup} \\
 &= sup \frac{\left(\sum_n |g(n\Delta x)|^2 \right)^{1/2}}{\|g(x)\|} sup \frac{\|g(x)\|}{\left(\sum_n |g(n\Delta x)|^2 \right)^{1/2}} \quad (3.48a)
 \end{aligned}$$

$$\forall g(x) \in V_{\Delta x}, \|g(x)\| \neq 0$$

and

$$\begin{aligned}
k(\tilde{U}_{\Delta x'}^{-1}) &= \|U_{\Delta x'}\|_{sup} \|\tilde{U}_{\Delta x'}^{-1}\|_{sup} \\
&= \sup \frac{\left(\sum_n |h(n\Delta x')|^2 \right)^{1/2}}{\|h(x)\|} \sup \frac{\|h(x)\|}{\left(\sum_n |h(n\Delta x')|^2 \right)^{1/2}} \\
&\quad \forall h(x) \in V_{\Delta x'} , \|h(x)\| \neq 0
\end{aligned} \tag{3.48b}$$

where the symbol $\|\|_{sup}$ denotes the supremum operator norm (Naylor and Sell, 1982).

The norm symbol $\|\|$ without a subscript denotes, of course, the norm associated with the global Hilbert space H.

In order to keep the stability of the linear estimation problem independent of the data resolution, the two above condition numbers, corresponding to two arbitrary data sampling intervals (Δx and $\Delta x'$), should be equal. An easy and straightforward way to achieve this result is to require that the two corresponding solution spaces are related through a simple **isometric scaling**, i.e.

$$g(x) \in V_{\Delta x} \Leftrightarrow h(x) = \sqrt{\frac{\Delta x}{\Delta x'}} g\left(\frac{\Delta x}{\Delta x'} x\right) \in V_{\Delta x'} \tag{3.49}$$

Indeed, if the last relationship is true, then the two condition numbers in eqs.(3.48a) and (3.48b) become equal by taking into account the isometric scaling property of the norm in the Hilbert space H; see eq.(3.43).

A non-varying stability level for the linear signal approximation requires, therefore, that all the different solution spaces $V_{\Delta x}$ (corresponding to different data resolutions Δx) are just scaled isometric versions of each other, according to the general formula (3.49). Under the initial assumption that the global space H is a proper RKHS, each of these closed subspaces $V_{\Delta x} \subset H$ is a RKHS itself, with its reproducing kernel $K_{\Delta x}(x, y)$ depending on the data resolution. The previous scaling condition between the various solution spaces implies an equivalent scaling condition between their corresponding reproducing kernels. In this way, the r.k of every solution space $V_{\Delta x}$ should have the following form for non-varying stability in the optimal approximation procedure:

$$K_{\Delta x}(x, y) = \frac{1}{\Delta x} K_o\left(\frac{x}{\Delta x}, \frac{y}{\Delta x}\right) \quad (3.50)$$

where $K_o(x, y)$ is the r.k corresponding to the solution space for a normalized data resolution $\Delta x = 1$. Due to the scale-invariance property of the norm in H, imposed by eq.(3.43), all the different reproducing kernels will satisfy the condition

$$\|K_{\Delta x}(x, y)\| = \|K_o(x, y)\| = \text{constant} \quad (3.51)$$

where the constant on the right hand side is independent of the data density. In order to establish this last property, the norm operator should be applied twice to the reproducing kernel $K_{\Delta x}(x, y)$, first considered as a function of x only and then as a function of y . The final norm equivalence in eq.(3.51) should be understood either with respect to the x coordinate (keeping y fixed), or with respect to the y coordinate (keeping x fixed).

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

$$\lim_{\Delta x \rightarrow 0} V_{\Delta x} = H \quad (3.52)$$

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

$$\lim_{\Delta x \rightarrow 0} K_{\Delta x}(x, y) = K(x, y) \quad (3.53)$$

Using eq.(3.50), the previous relation can be finally expressed as follows:

$$K(x, y) = \lim_{\Delta x \rightarrow 0} \frac{1}{\Delta x} K_o\left(\frac{x}{\Delta x}, \frac{y}{\Delta x}\right) \quad (3.54)$$

The last condition should be satisfied by the reproducing kernel of the global Hilbert space H in order for the solution of the linear approximation problem (according to the

deterministic collocation approach - section 3.2) to satisfy the three fundamental properties stated in the beginning of section 3.4. The derivation of this condition was based on a certain scale-invariance property that has been imposed for the norm in H . Note that other types of scale-invariant norms, not necessarily the same as in eq.(3.43), may be introduced and used for the modelling of the Hilbert space H . In such a case, we should modify accordingly the isometric scaling (3.49) for the solution spaces, as well as the condition (3.50) for their reproducing kernels. An analogous change will also occur in the final condition (3.54).

It is seen, however, that this ‘optimal’ reproducing kernel will not be a proper function in the usual sense. As a matter of fact, $K(x, y)$ seems to have all the basic characteristics of the Dirac delta distribution $\delta(x - y)$. This type of ‘function’ is a rather complicated notion that is defined as a limiting process on well behaved kernels (e.g. Gaussian, Dirichlet, Fejer, etc.), and it is particularly useful for its informal reproducing property under the usual L^2 inner product (Gelfand and Shilov, 1964), i.e.

$$\forall f(x) \in L^2(\mathfrak{R}) \quad , \quad \int f(x) \delta(x - y) dx = f(y) \quad (3.55)$$

Of course, in such cases the usual formalism of deterministic collocation cannot be applied anymore (due to the lack of a proper reproducing kernel for H), but instead we have to develop a new methodology for constructing the solution subspaces $V_{\Delta x} \subset H$

and obtaining a stable solution for the linear approximation problem at every data resolution level. More details and explanations will be given in section 3.6.

3.5 Trade-Off Between Data and Model Resolution

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

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

Although both stability and convergence are very important in the deterministic collocation framework, in this section we will restrict our attention mostly on the convergence issue. The stability problem has already been discussed in the previous sections using the tool of frame theory, and it was actually shown that it is reduced to a convergence question in the case of infinitely dense data (see the discussion at the end of

section 3.3.3). For some interesting studies on the stability of geodetic estimation methods, see Schwarz (1979), Rummel et al. (1979) and Gerstl and Rummel (1981).

The convergence problem in collocation theory has been investigated by many authors; see Moritz (1976a), Tscherning (1978a), Sanso and Tscherning (1980) and Barzaghi and Sanso (1986). Most of the studies that have used Krarup's deterministic framework rely upon rather strong assumptions, which create new (and even more difficult) questions that need to be answered for a complete treatment of the problem. In particular, the first standard assumption is that the observational representers in the Hilbert space (within which the convergence problem is studied) are always linearly independent, so that the Grammian matrix \mathbf{G} in eq.(3.10) is always non-singular. Such a restriction is imposed even for very dense, yet discrete, data point configurations (i.e. ε -net point sets, see Moritz, 1976a), which are considered sufficient to ensure the *full* recovery of the unknown field through the linear estimation algorithm. In brief, the observational representers associated with such sufficiently dense data topologies (e.g. the set of the representers of the evaluation functional over an ε -net point network) are assumed to provide a complete system of independent base functions for the whole Hilbert space in which we model our unknown signals (Tscherning, 1978a). Based on this setting, a fundamental result of strong convergence for the collocation algorithm has already been established by Moritz (1976a) and it was subsequently extended by Tscherning (1978a).

The previous basic assumption for the convergence of deterministic collocation implies, in a way, that there should exist a close connection between the adopted approximation model (i.e. the RKHS H) and the overall ‘resolution properties’ of the signals that this model/space can provide. In simple words, for a given Hilbert space H with reproducing kernel $K(P, Q)$, we should be able to identify a set of points $\{Q_n\}$ such that the functions $K(P, Q_n)$ form a basis for H (Tscherning, 1978a). Any data point configuration (finite or infinite) $\{Q'_n\}$ that violates the linear independence condition for $K(P, Q'_n)$ is considered unacceptable and it cannot be incorporated in the linear approximation framework. But is it reasonable to use an estimation model based on selective ‘sampling scenarios’ for the unknown field, excluding other possible data configurations with varying spatial resolution that may arise in practice? And, most importantly, how are we going to a-priori select or determine the resolution properties of our unknown signals that should be implicitly associated with the form of their modelling reproducing kernel?

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

$$\forall f(P) \in H, \quad f(P) = \lim_{N \rightarrow \infty} \sum_{n=1}^N a_n K(P, Q_n) \quad (3.56)$$

where the coefficients a_n are obtained from the solution of the following non-singular linear system:

$$\begin{bmatrix} K(Q_1, Q_1) & \cdots & K(Q_1, Q_N) \\ \vdots & \ddots & \vdots \\ K(Q_N, Q_1) & \cdots & K(Q_N, Q_N) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} f(Q_1) \\ \vdots \\ f(Q_N) \end{bmatrix} \quad (3.57)$$

Thus, the critical issue that will control the convergence of the linear estimation algorithm becomes the spatial point configuration $\{Q_n\}$ for which (3.56) and (3.57) are satisfied. It is a rather challenging and relatively unexplored mathematical problem to try finding (if it exists) for a given reproducing kernel $K(P, Q)$ its associated point distribution $\{Q_n\}$ that can provide completeness and linear independence in the corresponding Hilbert space H , according to the two previous equations. For some special cases, see Nashed and Walter (1991) and Yao (1967).

It is worth mentioning that the convergence setting in deterministic collocation, as expressed through the last two equations, clearly corresponds to the search of a *sampling theorem* in the underlying Hilbert space, where the associated sampling functions will be given by the general formula (Moritz, 1976b)

$$S(P, Q_m) = \lim_{N \rightarrow \infty} \sum_{n=1}^N K(P, Q_n) \mathbf{G}^{-1}[n, m] \quad (3.58)$$

and $\mathbf{G}^{-1}[n, m]$ are the elements of the inverse of the $N \times N$ Gramian matrix shown in eq.(3.57).

It should be intuitively obvious that the use of a slowly decaying reproducing kernel (with a sparse sampling point distribution $\{Q_n\}$ associated with it) will limit our modelling capabilities to correspondingly smooth signals, which may not be compatible with relatively dense data sets. If we want to increase the resolution content ('details') of the signals in the adopted Hilbert space model H , we have to choose a faster decaying reproducing kernel with a denser sampling point distribution $\{Q_n\}$ associated with it. It is less obvious, however, that the use of such 'higher-resolution' reproducing kernels in conjunction with low-resolution data sets (relative to the spread of $K(P, Q)$ or the spatial density of its associated sampling point set $\{Q_n\}$) can result in minimum-norm signal approximations that may be completely erroneous. An example of such situation is illustrated in Figure 3.5, where 13 point gravity anomaly values along a certain profile, with a normalized sampling interval $\Delta x = 1$, are used to determine optimal linear approximations of the underlying signal through two different reproducing kernels.

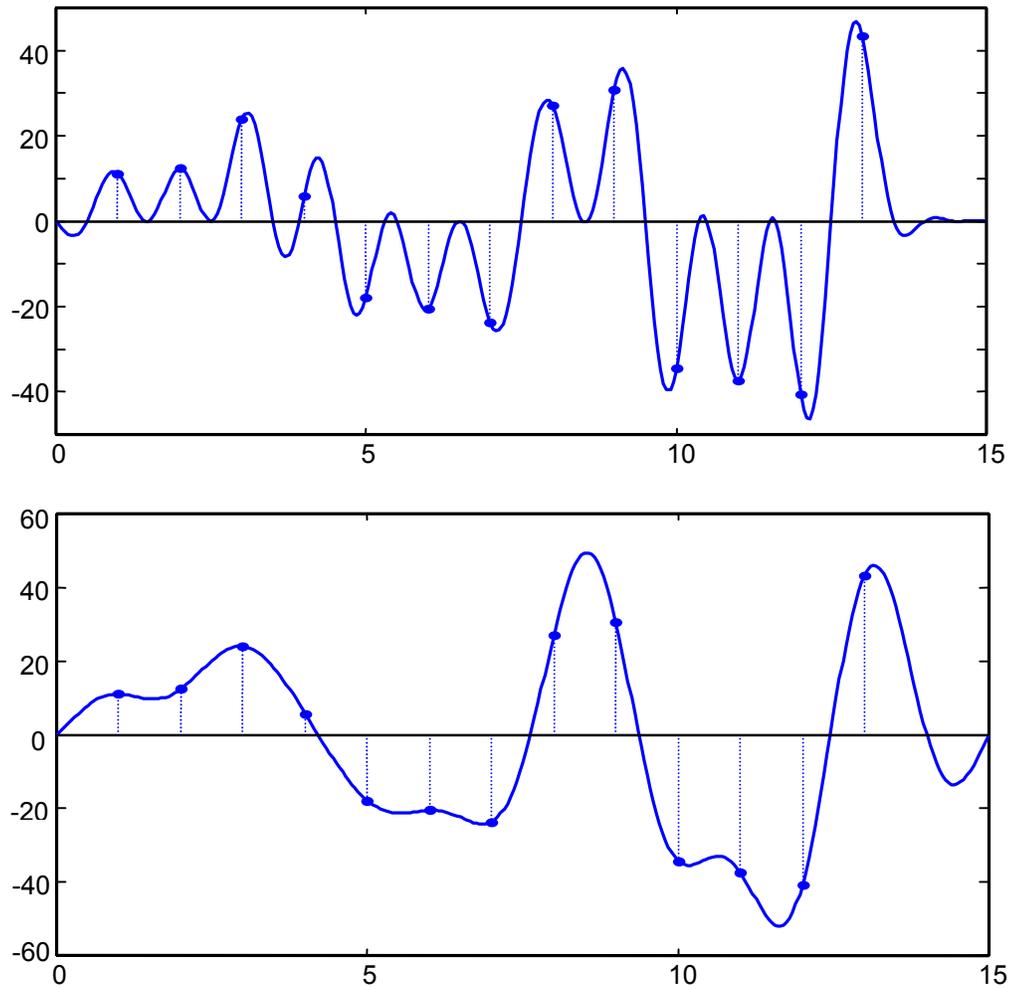


Figure 3.5 Minimum-norm collocation using a ‘low-resolution’ r.k (lower graph) and a ‘higher-resolution’ r.k (upper graph). The dots represent the data values.

In the one case (upper graph) the used reproducing kernel corresponds to the Hilbert space of $L^2(\mathfrak{R})$ band-limited functions within the frequency interval $[-2\pi, 2\pi]$, whereas in the second case (lower graph) the minimum-norm interpolation takes place in the Hilbert space of $L^2(\mathfrak{R})$ band-limited functions within the frequency interval $[-\pi, \pi]$. Although the first RKHS is capable of providing more detailed signals than the second

RKHS, its use with a sparse data point distribution produces a linear estimation that contains artificial details which cannot be justified from the resolution of the available data. On the other hand, the use of a lower-resolution reproducing kernel in the second case results in a more reasonable approximation scheme, which ‘merges’ the spatial resolution of the available data points with the model resolution implied in the corresponding RKHS in a certain optimal fashion (in this specific example, we just have an application of Shannon’s sampling formula). Similar examples can be easily constructed for many other pairs of ‘lower’ and ‘higher’ resolution reproducing kernels.

In every convergence problem in signal approximation methods the central role is always played by the notion of a *correct* or *true* solution for the unknown field that we try to estimate. Unlike the philosophy of pure mathematics, in applied physical sciences (like physical geodesy) the correct solution is just a model obtained by simplifying in some way physical reality. In Sanso’s (1987) words: ‘*this is what we can do at most*’. When we consider the deterministic collocation concept for our approximation purposes, the modelling takes the form of a specific Hilbert space with an associated reproducing kernel. As it was explained in the previous paragraphs, however, there will always exist a certain trade-off among: (i) the overall signal details that this model can provide, (ii) the quality of the signal approximation for sparse data point configurations (relative to the implied model spatial resolution $\{Q_n\}$), and (iii) the stability/existence of the signal approximation for data sets denser than the implied model resolution (see also Rummel et

al., 1979). This would suggest that we should change our modelling choice for every new data point configuration in order to construct reasonable and stable signal approximations that balance between the data and model resolution. But in this way the convergence problem becomes meaningless since the global RKHS will be different for every new data set!

3.6 The Connection with the MRA Concept

The final result in section 3.4 may seem a bit puzzling and even contradictory with the original setting of the problems discussed therein. Essentially, we have shown that in order to achieve

- a constantly stable linear signal approximation for increasing data resolution, and
- convergence for infinitely dense data,

then the global Hilbert space H , within which we model our unknown signals, should not possess a reproducing kernel! That is because $K(x, y)$, as determined by the limiting procedure in section 3.4.2 under the two previous conditions, could not definitely belong in any reasonable space of natural signals and, therefore, it is not a reproducing kernel in the strict sense. Actually, a more involved mathematical analysis is required to establish the non-existence of a proper reproducing kernel for H in this case. For the purpose of this thesis, it is sufficient to conclude that the method of deterministic collocation cannot

satisfy the two previous properties, since its fundamental ingredient (the reproducing kernel) is reduced to a delta-type kernel which is useless for our signal estimation applications. Recall that the validity of this result depends on a certain scale-invariance condition that had been imposed for the norm in H , according to eq.(3.43).

Although we have only considered the special case where the observed data are gridded point values of the unknown field itself (interpolation problem), the same result can be obtained by using other types of gridded linear functionals as well. Also, if we had chosen to truncate the sequence of the stable solution spaces $V_{\Delta x}$ at some ‘ideal’ finite resolution Δx_{min} (thus requiring only convergence of the linear approximation scheme for data sets approaching a certain finite resolution level), then we can establish the existence of a proper reproducing kernel for the global Hilbert space H , but we would still be facing the important trade-off problems between model and data resolution that were discussed in section 3.5.

On the other hand, this peculiar behaviour of the ‘optimal’ $K(x, y)$ according to eq.(3.54) corresponds to a perfectly acceptable mathematical structure that lies in the vast field of generalized functions and distributions (Sobolev, 1964). What is the important characteristic of all function spaces admitting such a distributional delta form as their reproducing kernel*? Basically, such a Hilbert space H should be interpreted as an

* The term ‘reproducing kernel’ is now used in a non-rigorous sense.

infinite-resolution signal space. Regardless of its topological structure (i.e. the specific type of its scale-invariant norm $\| \cdot \|$ according to eq.(3.43)), the value of an arbitrary signal in H , at a point Q , cannot be generally determined by its local behaviour in the neighborhood of Q . That is because no well behaved kernel $K(x,y) \in H$ exists which, when applied to an arbitrary $f(x) \in H$, could recover the exact same signal. However, this does not mean that H will not contain smooth functions, whose value at a specific point could be predicted arbitrarily well from a set of values at a dense network of adjacent points.

An infinite-resolution Hilbert space provides an ideal choice to model signals with highly irregular patterns, where abrupt changes may be expected between nearby points. In view of the rapid increase in gravity field data resolution that takes place today, the recovery of such local irregularities in the gravity field signals not only seems possible from an observational point of view, but it is also extremely useful in certain types of applications (e.g. geophysical inverse problems, geodynamical studies, cm-level geoid determination, etc.). Thus, it would seem only natural to use a Hilbert space setting that is actually capable of providing signals with such ‘erratic’ behaviour. Extreme cases of signals with singularities should not be excluded from the approximation framework, since such situations may very well arise in various signal processing applications (Mallat and Hwang, 1992).

A very important problem, however, still remains. If the ‘optimal’ global Hilbert space should not possess a proper reproducing kernel, how are we going to actually construct the various solution subspaces $V_{\Delta x} \subset H$ for the linear estimation problem at each data resolution level? In such cases, the observational representers

$$g_n(x) = K(x, n\Delta x) \quad (3.59)$$

are not usual functions, and linear expressions of the form of eq.(3.40c) cannot be properly defined and used for the numerical computation of the estimated field. A possible solution to this problem is discussed below.

3.6.1 Linear Estimation as a Multiresolution Approximation

At first, it would be useful to understand in simple terms why the minimum-norm solution of the linear estimation problem will fail to give a reasonable answer when the selected signal Hilbert space has the infinite-resolution characteristics discussed above (i.e. H does not possess a proper reproducing kernel). In this case, the *evaluation functional* in H is not bounded (continuous), which means that there is no real number $C > 0$ such that the inequality

$$|f(P)| \leq C \|f\| \quad (3.60)$$

will hold for every $f \in H$ and for every point P in the domain of f (Moritz, 1980). The invalidity of (3.60) basically means that there will exist functions in H with $\|f\| = 0$, but at the same time there may be points in their domain for which $f(P) \neq 0$. In other words, isolated singularities are acceptable in infinite-resolution Hilbert spaces and they are topologically equal to the zero function. Therefore, the minimum-norm function that assumes given values at a discrete network of data points will be the zero function, and that makes minimum-norm interpolating solutions for the linear estimation problem useless, both theoretically and computationally. As was explained in the previous paragraphs, knowledge of f at discrete points cannot generally provide any information about its behaviour in the neighborhood of these points, and this fact is immediately reflected in the inadequacy of eq.(3.40c) to yield a proper numerical result. A more rigorous mathematical treatment requires the incorporation of concepts from *measure theory* (Halmos, 1991) which will be avoided here.

The above situation should not be perceived as a ‘deficiency’ of the infinite-resolution Hilbert space H , but it is merely a modelling requirement. In order to compute a reasonable linear estimation for an unknown field from its discrete values, we have to develop a different methodology which will not employ the generalized reproducing kernel of H and it will satisfy the crucial properties of constant stability (for increasing data resolution) and convergence (for infinitely dense data). Such a constructive methodology will now be briefly described.

The final result of any linear estimation procedure in a Hilbert space H can always be viewed as a bijective linear mapping (T) from an observation space (Λ) to a solution subspace ($V \subset H$). The observation space contains all possible data sets obtained from different numerical realizations of an unknown field, which is now modelled through an infinite-resolution Hilbert space H with a scale-invariant norm according to eq.(3.43). All these data sets are associated with a *specific* geometrical/spatial configuration, which in this chapter has been assumed as a sequence of gridded sampled values at a certain resolution level. In order to have a stable bijective mapping T , the following condition should be satisfied:

$$A \| \mathbf{b} \|_{\Lambda} \leq \| \hat{f} \|_{H} \leq B \| \mathbf{b} \|_{\Lambda} \quad \forall \hat{f} \in V, \mathbf{b} \in \Lambda \quad (3.61)$$

where \mathbf{b} denotes the gridded data set obtained from an unknown field $f \in H$, $\hat{f} = T \mathbf{b}$ denotes its linear approximation that is determined via the estimation operator T , and A , B are strictly positive constants. The formula (3.61) is just an alternative definition for an **isomorphic mapping** between the solution space and the observation space (Holschneider, 1995; p. 183). For all practical purposes, it is convenient to choose the norm in Λ as the standard l^2 norm for discrete sequences, and thus the observation space can always be viewed as the $l^2(Z)$ Hilbert space of square-summable sequences. This should not raise any major objections, since the available discrete data are always

measurable through such a norm in almost every scientific discipline. The situation is illustrated in Figure 3.6.

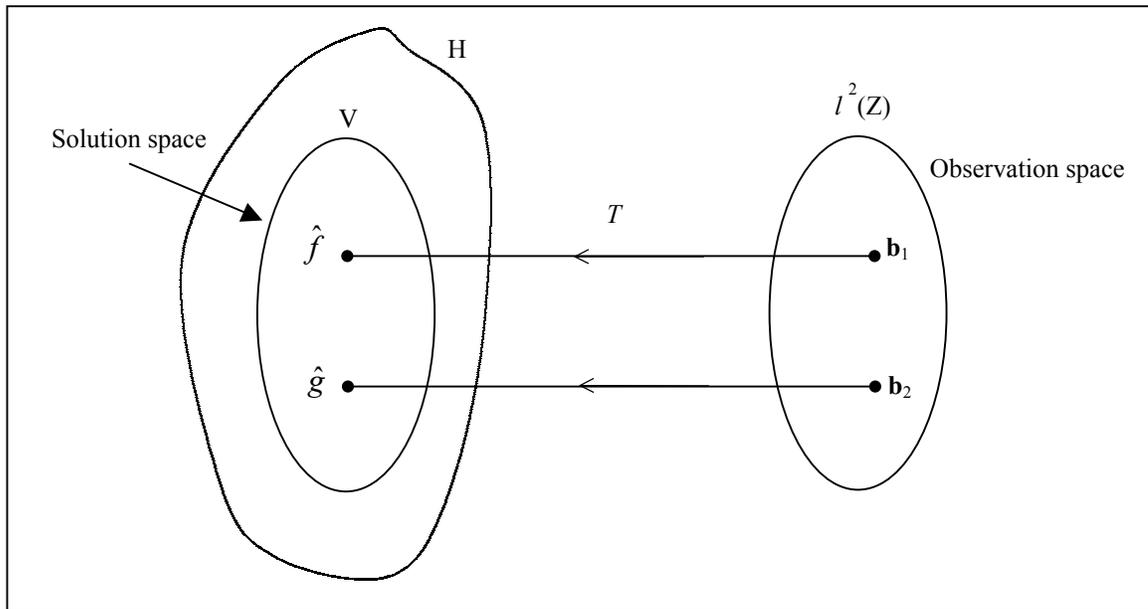


Figure 3.6 Linear estimation as an isomorphic mapping T

So far we have identified a stable linear estimation procedure (in an infinite-resolution Hilbert space H) as an isomorphic mapping T between a solution space $V \subset H$ and the observation space $l^2(Z)$, which contains regularly sampled values of the unknown signal that we want to approximate. If we further require that T is a translation-invariant estimation operator, then the solution space should be generated by a stable (Riesz) basis of the form $\varphi_{\Delta x}(x - n\Delta x)_{n \in \mathbb{Z}}$, where Δx is the data resolution level associated with the observation space. The proof is very simple and it is omitted (see Mallat, 1989b; Walter, 1992).

After having established the structure of a single solution space (corresponding to a specific data resolution level), let us now see what additional properties we should require when the data density changes. For a new resolution level $\Delta x'$, we should again be able to obtain a similar isomorphic mapping T' between a new solution space $V' \subset H$ and the observation space $l^2(Z)$. Note that the observation space again contains all square-summable data sequences, which are obtained though with a different sampling rate from the unknown signal. In order to maintain the same stability level in the new isomorphism T' as in T , the two bijective estimation operators should have equal condition numbers. It is easy to prove that this will be true (under the scale-invariance condition (3.43) for the norm in the global Hilbert space H) when V and V' are related through the isometric scaling of eq.(3.49). If we combine this last property with the requirement of translation-invariance for the new isomorphism T' , then the new solution space should, too, be generated by a stable (Riesz) basis of the form $\varphi_{\Delta x'}(x - n\Delta x')_{n \in Z}$, such that

$$\varphi_{\Delta x'}(x) = \varphi_{\Delta x}\left(\frac{\Delta x}{\Delta x'}x\right) \quad (3.62)$$

In essence, the linear signal estimation algorithm should always employ the ‘same’ isomorphic mapping at different spatial scales (depending on the resolution of the available sampled data), thus adding a scale-invariance property to our framework as well. The underlying approximation model of the procedure described above can be expressed in the general linear form

$$\hat{f}(x) = \sum_n f(n\Delta x) \varphi\left(\frac{x}{\Delta x} - n\right) \quad (3.63)$$

where $\varphi(x) \in H$ is a basic kernel, whose scaled translates $\varphi\left(\frac{x}{\Delta x} - n\right)_{n \in Z}$ should always constitute a Riesz basis for their closed linear span $V \subset H$. The choice of a specific kernel $\varphi(x)$ must be based on certain optimality criteria for the estimation procedure, which will normally try to minimize some functional of the signal approximation error $e(x) = f(x) - \hat{f}(x)$.

It is also logical to impose a certain causality principle with respect to the signal information that we are going to extract from the available discrete data through the various isomorphic estimation mappings. Such a principle can be stated as follows:

$$V \subset V' \quad \text{for} \quad \Delta x > \Delta x' \quad (3.64)$$

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

$$\lim_{\Delta x \rightarrow 0} V = H \quad (3.65)$$

If we recall from Chapter 2 the basic properties that characterize a multiresolution sequence of subspaces (MRA), then we see from the previous discussion that a constantly stable, convergent and translation-invariant scheme for the signal estimation problem in an infinite-resolution Hilbert space H requires, basically, the introduction of such a MR structure within H . The solution at each data resolution level Δx can then be obtained through an isomorphic mapping of the form (3.63), where the scaling kernel $\varphi(x)$ is the one that generates the MRA-type sequence of solution subspaces. In a way, we have ‘re-invented’ Mallat’s MRA concept as a useful regularization tool for linear estimation problems in infinite-resolution Hilbert spaces, using discrete gridded data at varying resolutions.

It is interesting to observe the basic modelling difference with respect to the deterministic collocation estimation method. Here, we a-priori defined the observation space to be the whole $l^2(\mathbb{Z})$ Hilbert space, irrespectively of the actual data resolution level. The unknown field $f(x) \in H$ and its observed gridded values are not restricted by any specific smoothing conditions, apart from the fact that the sampling procedure should always result in square-summable data sequences. The necessary smoothing that is required to obtain a unique and stable solution for the linear estimation problem is applied entirely on the approximated signal $\hat{f}(x)$ through an isomorphic mapping which ‘adapts’ to the current data resolution level. In deterministic collocation, on the other hand, the behaviour of the unknown field is a-priori restricted according to a certain

model (i.e. reproducing kernel) and the admissible observation space may change, for different data resolution levels, due to this restriction (see the discussion in section 3.2.4).

3.6.2 Summary – Final Remarks

The main purpose of this rather long chapter was to reveal the drawbacks of the classic deterministic collocation method, and to adopt an MRA-type estimation framework as a useful alternative for dealing with signal interpolation problems in Hilbert spaces using gridded data. A descriptive, rather than strictly mathematical, approach has been followed with the emphasis mostly put on the modelling aspects of the linear approximation problem.

The basic drawback behind deterministic collocation is that it depends exclusively on a *fixed* model/waveform (i.e. the reproducing kernel) for describing the signal behaviour, without really taking into account the resolution of the available discrete data. This affects directly the quality of the approximated signal (for low-resolution data sets), as well as the numerical stability of the solution algorithm (for high-resolution data sets). In section 3.4.2, it was shown that a constantly stable and convergent (in the sense of infinitely dense data) linear estimation scheme in a Hilbert space H is generally incompatible with the existence of a proper reproducing kernel for H . In particular, when the norm in H satisfies a certain scale-invariance condition, then the reproducing kernel should have a generalized form similar to the Dirac delta function in order to maintain

stability and convergence for increasing data density. This suggests an interesting alternative formulation for the linear estimation problem, which will not ‘connect’ anymore the data values and the unknown signal via inner product observation equations in a finite-resolution Hilbert space with a proper reproducing kernel.

Such an alternative methodology considers the unknown field as a single element of an infinite-resolution Hilbert space H with a scale-invariant norm according to eq.(3.43), and it uses a sequence of nested multiresolution subspaces $\{V_j\}$ within H , similar to the dyadic MRAs that were discussed in Chapter 2. In this way, rather than trying to invert an increasingly ill-conditioned frame operator (deterministic collocation approach), the solution of the linear estimation problem at each data resolution level Δx_j is obtained through constantly stable isomorphic mappings between the observation space $l^2(Z)$ and a corresponding signal subspace $V_j \subset H$, according to the translation-invariant scheme of eq.(3.63). The sequence of the solution subspaces will depend on a generating kernel whose scaled translates $\varphi(\frac{x}{\Delta x_j} - n)_{n \in Z}$ should provide a Riesz (stable) basis for their closed linear span, as well as on a specific rule according to which the data resolution level Δx_j changes from one solution space V_j to the next V_{j+1} . In a way, the problem of selecting a specific scaling approximation kernel has now replaced the reproducing kernel choice problem that existed in the deterministic collocation approach.

It should be noted that in such an MR interpolatory framework, the signal estimation from a (finite or infinite) data set with resolution Δx_j always takes place in an infinite-dimensional solution subspace V_j . This is in contrast to deterministic collocation, where the minimum-norm signal solution from a finite data set is always a member of a finite-dimensional Hilbert subspace spanned by the observational representers. What remains ‘limited’ in the MR approximation framework is not the dimensionality of the estimation space, but rather the spatial resolution of the recovered field $\hat{f}(x) \in V_j$ within a certain hierarchical sequence of nested Hilbert subspaces $\{V_j\}$ of increasing resolution. This maximum recoverable field resolution is dictated by the sampling density of the available data points. The latter also determines the spread of the basic scaling kernel $\varphi(x)$, thus obtaining the desired balance between model and data resolution that was discussed in section 3.5.

Note that we have not even established that such an MR subspace structure $\{V_j\}$ generally exists. From Mallat’s developments, we definitely know that it exists when we identify the infinite-resolution space H as the $L^2(\mathfrak{R})$ Hilbert space, and when the data resolution level has the dyadic form $\Delta x_j = 2^{-j}$ (Mallat, 1989b). From Walter’s developments, we also know that in most dyadic MRAs there do exist cardinal scaling kernels $\varphi(x)$, which can be used to solve the linear signal interpolation problem at various resolution levels in a constantly stable and convergent way (Walter, 1992).

Multi-dimensional extensions, as well as extensions in compact signal domains, are also possible in these cases. In the following chapter, we will actually generalize Mallat's MRA framework for non-dyadic schemes, and we will also solve the problem of selecting an optimal scaling kernel based on the spatio-statistical collocation concept.

Chapter 4

OPTIMAL MULTIREOLUTION APPROXIMATION

Following the suggestions given in the last sections of the previous chapter, we will now study the problem of finding an *optimal* scaling kernel $\varphi(x)$ for the linear translation-invariant approximation of an unknown field from discrete gridded data at different resolution levels. The basic setting will remain relatively simple by considering only 1D signals, and the available discrete data will still be assumed as noiseless gridded values of the unknown field itself. The latter will be modelled as an individual element of the infinite-resolution Hilbert space $L^2(\mathfrak{R})$, which permits the use of the (always very useful) Fourier transform formalism. Certain modifications can be applied to the following formulations of this chapter to accommodate more complicated estimation problems in higher dimensions, using various types of gridded linear functionals. Nevertheless, even for the simplest of all cases that we are going to analyze here, the following methodology remains especially important since it will open a new and less restrictive viewpoint to the classic MR approximation concept that was originally developed by Mallat. Using a

spatio-statistical collocation approach and a mean square error optimal criterion, a new constructive (frequency-domain based) framework for building generalized MR analyses in $L^2(\mathfrak{R})$ will be presented, without the need of the usual dyadic restriction that exists in standard wavelet approximation theory.

4.1 Basic Aspects of the Multiresolution Model

The basic form of the linear approximation model that we are going to consider in this chapter is

$$\hat{g}(x) = \sum_n g(n\Delta x_j) \varphi\left(\frac{x}{\Delta x_j} - n\right) \quad (4.1)$$

where $\varphi(x) \in L^2(\mathfrak{R})$ is an unknown scaling kernel that should be determined in some optimal sense. In the previous chapter, the use of linear models of the form given by eq.(4.1) was introduced as an effective alternative to deterministic collocation, which has certain limitations for dealing with discrete data at varying resolutions. When $\varphi(x)$ is a fixed kernel (i.e. independent of the actual data resolution) that satisfies the Riesz condition of eq.(2.8), then the previous linear approximation formula can easily be identified as a continuously stable isomorphic mapping from the observation space $l^2(Z)$ to a model-solution signal space $V_{\Delta x_j} \subset L^2(\mathfrak{R})$, for every resolution value Δx_j . If, apart

from stability in the estimation algorithm, we additionally require causality (i.e. denser data provide new information about the unknown field, whereas increasingly sparse data contribute very little knowledge that becomes *no knowledge* \equiv *zero function* in the limit) and convergence to the true field $g(x) \in L^2(\mathfrak{R})$, then the sequence of the solution subspaces $\{V_{\Delta x_j}\}_{j \in \mathbb{Z}}$ should create an MRA-type structure within the infinite-resolution Hilbert space $L^2(\mathfrak{R})$.

The use of convolution-based estimation models of the form of eq.(4.1) has been extensively studied by many researchers, especially in the signal and image processing community. Depending on the choice of their basic kernel, they provide a wide variety of corresponding interpolating (or quasi-interpolating, since $\varphi(x)$ need not be necessarily cardinal) signal spaces with certain associated optimal properties, whose study goes back to the famous *Strang-Fix theory* on error bounds for n th-order approximations of finite-energy signals (Strang and Fix, 1971). Such type of linear estimation models can be identified as a natural generalization of Shannon's interpolation formula for band-limited signals, which is often avoided in practice in favor of shorter kernel methods such as bilinear interpolation (Pratt, 1978), cubic convolution (Park and Showengerdt, 1983; Keys, 1981), or polynomial spline approximation (Unser et al., 1991; Hou and Andrews, 1978). The latter approximation models usually outperform *sinc*-based interpolation for finite-support $L^2(\mathfrak{R})$ signals and they are much more efficient to implement, especially

for higher dimensions (Parker et al., 1983; Unser, 1999). Some interesting results on the asymptotic equivalence between Shannon's sampling theory and the generalized interpolation formula in eq.(4.1) have been obtained for the case where $\varphi(x)$ is a B-spline kernel of order n (Unser et al., 1992), as well as for more general cases (Aldroubi and Unser, 1994). A comprehensive treatment of such generalized linear interpolation (or quasi-interpolation) models, including a detailed pointwise, asymptotic, and L^2 error analysis of their performance with respect to the data resolution level, can be found in the papers by Blu and Unser (1999a) and Unser and Daubechies (1997). In Unser and Zerubia (1997, 1998), the well known Papoulis sampling theorem for multi-sensor/multi-channel data (Papoulis, 1977) was appropriately extended for non band-limited cases using a combination of expressions of the form given in eq.(4.1), and it was studied with respect to various issues such as stability and aliasing error performance.

The multiresolution aspects of these generalized interpolation models, in connection with the MRA and wavelet theory in the $L^2(\mathfrak{R})$ Hilbert space, have also been explored in detail with many important achievements, such as: sampling theorems for wavelet multiscale subspaces (Walter, 1992, 1994; Zayed, 1993), use of the Zak transform for general sampling theorems in MRA subspaces and for studying their stability at non-zero sampling phase values (Janssen, 1993), construction of various orthonormal cardinal scaling functions (Xia and Zhang, 1993), development of interpolating scaling functions based on the autocorrelation function of an orthogonal scaling function (Saito and

Beylkin, 1993), spline-based interpolating scaling functions and development of fast B-spline transforms for continuous image representation and interpolation at various scale levels (Unser and Aldroubi, 1992; Unser et al., 1991), construction of interpolating wavelet transforms and symmetric interpolating wavelet functions (Aldroubi and Unser, 1993; Donoho, 1992), extension of Walter's sampling theory in MRA subspaces to include non-uniform sampling, derivative sampling, oversampling and local averages (Djokovic and Vaidyanathan, 1997), quantitative and qualitative Fourier analysis of the approximation error in generalized interpolation and quasi-interpolation scaling models (Blu and Unser, 1997, 1999b, 1999c), and study of families of multiresolution and wavelet spaces with certain optimal properties (Aldroubi and Unser, 1993), among many others. An excellent recent review paper on various aspects of linear sampling-type estimation models within a Hilbert space framework is also Unser (2000), which can be considered as an updated version of the classic review paper on sampling theory by Jerri (1977). At a considerably more advanced mathematical level are the papers by Higgins (1985) and Butzer et al. (1988), which provide a solid foundation on the basic properties of sampling estimation models and an exhaustive study of their approximation error.

For geodetic applications, the use of linear estimation models of the type given in eq.(4.1) (i.e. of their multi-dimensional counterparts) is of great importance, and not just for routine interpolation purposes. The multiscale character of these models provides a very natural framework to **reference** the discrete data values with respect to a continuous representation of the underlying unknown field, that is directly adapted to its sampling

resolution; see Sanso (1987) for an interesting discussion on this aspect. As it was mentioned in the previous chapter, this is in contrast to the usual deterministic collocation method where the approximation kernel does not ‘sense’ the actual data resolution level, causing (apart from numerical stability problems for high data point density) certain ‘fitting’ or ‘incompatibility’ problems between data and model resolution that automatically affect the quality of the final signal estimation (see section 3.5)*.

The need for a proper **data referencing model** is embedded in most problems of modern operational geodesy and its implementation is probably the most important and crucial step for their solution. For example, in order to solve Stokes’s BVP using a global grid of gravity anomaly measurements in a way that will allow us to rigorously study the stability, convergence and the behaviour of the estimation error of its solution, we should first decide upon a choice of base functions for approximating the continuous gravity anomaly signal using the original discrete data (data referencing). Then, we can apply the Stokes convolution integral to this discretized representation of the gravity anomaly field and study the result within the Hilbert space associated with the chosen system of base functions. This is exactly what the collocation framework offers when it is applied, for example, in geoid determination from discrete gravity values in a global (or local) sense.

* One might argue here that the classic collocation approximation kernel (i.e. the one given in terms of a fixed CV function or reproducing kernel, with a possible bounded linear functional applied on it depending on the type of data) does take into account the actual data resolution, especially when it is determined through an experimental procedure (i.e. empirical determination of the signal CV function). What we mean in the text, however, is that the *spread* of the usual collocation kernel is not adapted to the given data sampling resolution in any prescribed way, as it happens in the linear model of eq.(4.1), and most importantly it should (in principle) remain constant regardless of the data point density.

Since we usually prefer to employ linear and translation-invariant data referencing models, and most of the operators of interest in geodesy are also linear and shift-invariant, the study of geodetic formulas $f = S g$ using discrete input data is simply reduced to the determination of the operation $S \varphi$, where S is the operator under consideration and φ is the kernel of the chosen data referencing model. In the geodetic literature, this procedure is commonly known as the propagation law for reproducing kernels or CV functions within the collocation framework.

In this respect, the MR estimation formula in eq.(4.1) provides a general and stable background scheme, upon which we can base the study of geodetic linear operators in shift-invariant interpolation spaces of increasing resolution according to the scheme $\hat{f} = S \hat{g}$. The use of efficient FFT-based computational techniques is of course possible in cases where we deal with convolution operators S , and it requires the knowledge of the Fourier transform of the signal $S \varphi$. Note here that the traditional application of FFT methods in physical geodesy, for the numerical evaluation of convolution integrals, has been always based on the assumption that $\varphi(x) = \delta(x)$, which basically implies that the discrete input data are not referenced at all with respect to some continuous signal model. Apart from certain theoretical questions that such an assumption certainly raises, the incorporation of a proper reference kernel/filter for the discrete data in the FFT algorithms significantly improves the accuracy of the results, depending of course on the aliasing level in the discrete data as well as on the behaviour of the convolution kernel S

itself. Such issues, however, will not be treated in this thesis. For some relevant details, see section 5.1.4.

Of special interest is the case where the reference estimation kernel $\varphi(x)$, apart from satisfying the essential Riesz condition, also provides a scaling function for some MRA model in the $L^2(\mathfrak{R})$ Hilbert space. The main advantage in such cases is the computational efficiency that is achieved due to the sparse spectral representation of many types of linear operators (integral, differential, etc.) in wavelet bases, as well as the ability to apply a localized analysis of the effect of such operators at various signal scale levels; see Alpert (1993), Beylkin (1992) and Beylkin et al. (1991) for more details.

Such wavelet-based multiresolution computations with various geodetic operators, including cases where a stable and unique operator inversion is required (e.g. downward continuation, inverse gravimetric problem, computation of altimetrically derived gravity anomalies, etc.), have already started to receive considerable attention within the geodetic community; see Keller (1995, 1997), Ballani (1995), Barthelmes et al. (1995), Battha et al. (1995) and Benciolini and Zatelli (1998). More complicated signal estimation problems involving different types of input data from various sources and resolutions (e.g. altimetry-gravimetry BVP), can also be treated within a generalized MR framework as the one implied by the special case of eq.(4.1). This kind of multi-data estimation

problems, however, will not be discussed here. For relevant details, see Freedden and Schneider (1998), Freedden (1999) and Li (1996a,b).

4.2 Spectral Aspects of the Multiresolution Model

The main characteristic of the linear approximation model in eq.(4.1) is that its kernel directly adapts, through an appropriate dilation, to the resolution level of the regular data grid (see Figure 4.1). Although this ‘kernel tuning’ was initially introduced in Chapter 3 as a necessary regularization tool in order to achieve unperturbed stability for the linear interpolation scheme in the L^2 norm, its usefulness and implications go far beyond that.

Let us briefly describe what exactly these implications are. The unknown field $g(x)$, which we try to estimate through eq.(4.1), is initially modelled as an infinite-resolution $L^2(\mathfrak{R})$ signal. Such a modelling choice is very flexible and not restrictive at all, since most of the signals that we encounter in practical applications can be considered as finite-energy signals. Furthermore, the $L^2(\mathfrak{R})$ space provides us with some additional nice properties, such as: (i) a simple inner product in order to keep the same geometrical elegance for the linear approximation problem that existed in the deterministic collocation method, (ii) a very useful induced norm that can measure the signal estimation error in the familiar mean square sense, and (iii) the existence of a powerful

tool (Fourier transform) that simplifies and speeds up the numerical computations, as well as the theoretical analysis, associated with the estimation algorithm of eq.(4.1).

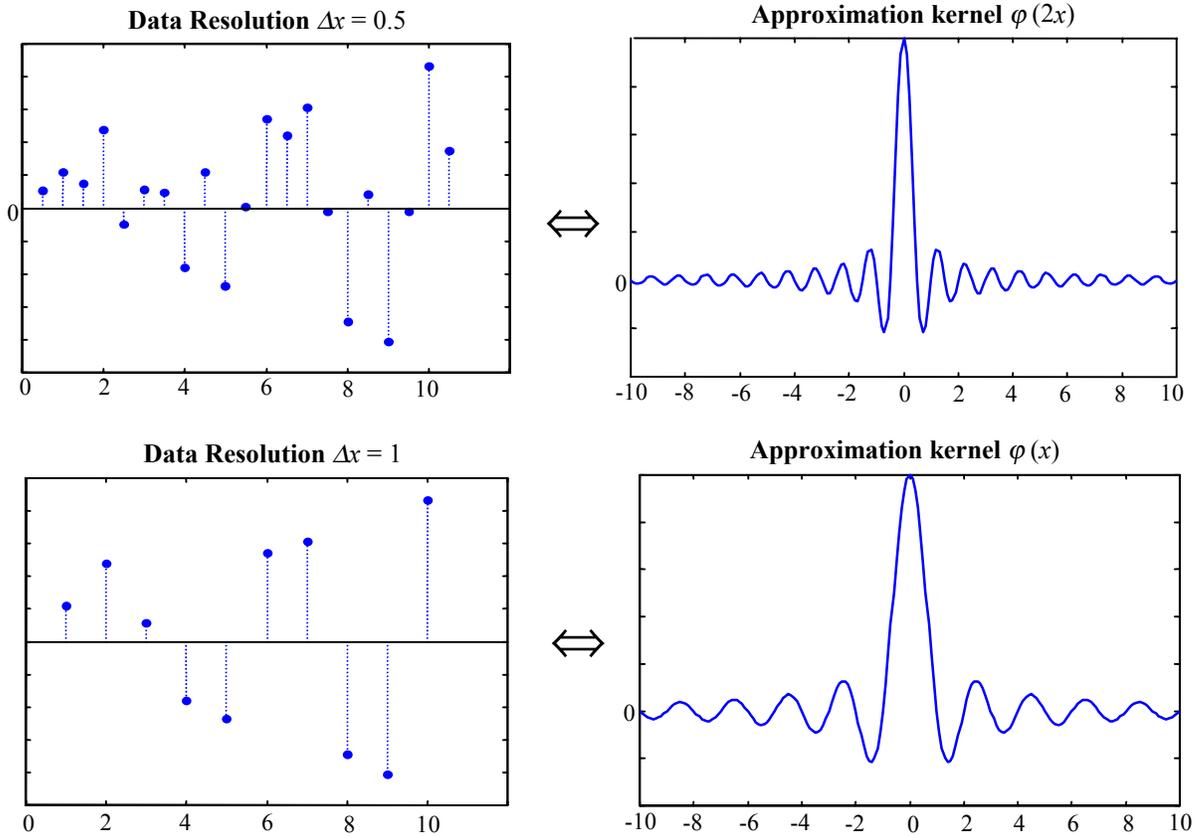


Figure 4.1 Adaptation of the approximation kernel to the data resolution level ($\varphi(x)$ is an interpolating B-spline of fifth order in this example)

There are of course many different methods for approximating our unknown field $g(x)$ with another simpler linear form $\hat{g}(x) \in L^2(\mathfrak{R})$ based on discrete gridded data, which all exhibit different levels of performance depending on the behaviour of the original signal and the grid density. Until recently, the only mathematical tool used by geodesists to

rigorously express and quantify the **physical limitations** (uncertainty) associated with the estimation of the gravity field from discrete measurements was the famous *Nyquist principle*. In terms of spectral content, discrete data were always linked with band-limited signal models which were considered the only possible realizations of the unknown fields under consideration. The interplay between data resolution and gravity field information was exclusively based on Fourier-based spectral concepts and measures like ‘maximum recoverable frequency’ or ‘minimum recoverable wavelength’ (Schwarz, 1984).

On the other hand, the various operational methods that are often employed in gravity field modelling do not necessarily imply a band-limited signal approximation. Consider, for example, the case where a Gaussian or a finite-support CV function is employed in the collocation estimation procedure, which clearly corresponds to the use of a non band-limited estimation kernel. In this way, spectral analysis issues are certainly affected by the used data referencing model. If we choose, for some reason, to approximate a continuous unknown field from its discrete values through a piecewise constant linear model (i.e. spline interpolation of zero order), it would not be very illuminating to apply a Fourier-based spectral analysis to the reconstructed field. In order to specify the limitations of our approximation models and to be able to spectrally analyze their results in a meaningful and consistent manner, a more general spectral system has to be used that will not necessarily use *frequency* as its spectral measure, and which could be directly ‘adapted’ to the chosen estimation kernel.

This is exactly the situation with the general approximation scheme in eq.(4.1), when it is viewed under an MRA hierarchical framework within the $L^2(\mathfrak{R})$ Hilbert space. If the basic estimation kernel $\varphi(x)$ corresponds to the scaling function of some nested multiresolution analysis model, then the approximated field $\hat{g}(x)$ will have a *resolution-limited* spectrum with respect to the associated wavelet basis (see Chapter 2). This provides a useful extension of the classic Nyquist principle, according to which discrete data can only recover a limited amount of spectral signal information. The central role now is not played by harmonic components of varying frequency, but by self-similar localized building blocks (wavelet components) according to a zoom-in/zoom-out approach. The reconstructed field is not measured against sinusoidal frequency variations of infinite extent, but against localized scale-dependent variations with respect to a given wavelet model that is implied by our estimation kernel. The maximum recoverable scale level of the signal details, according to the adopted MRA model, is actually determined by the resolution of the given signal values. Note that under such a generalized spectral framework, the usual frequency-domain approach that associates a band-limited signal Fourier spectrum to a set of gridded data is not lost, but it just becomes a special MRA case for the kernel choice $\varphi(x) = \text{sinc}(x)$; see, e.g. Zayed (1993, p. 317).

The need for using scale-dependent local measures to describe the behaviour of gravity field signals was already identified by Sanso (1987), long before the developments in wavelet theory started to reach the geodetic community. The linear approximation model

of eq.(4.1), in conjunction with Walter's sampling theory for MRA subspaces and related wavelet signal expansions, provide very powerful tools towards this direction. At this point, however, one could naturally ask: which specific MRA model/kernel $\varphi(x)$ should we use to describe the signal details at the given data resolution level? Is there some way to determine an optimal scaling kernel for the given unknown field $g(x)$ with the given discrete data? Should we restrict ourselves only to dyadic sampling schemes, as Mallat's MRA theory requires? Furthermore, our discussions from the last sections of Chapter 3 up to this point have considered only the use of a *fixed* reference kernel $\varphi(x)$, which, in a way, implies a 'stationary' treatment with respect to the data resolution parameter. An alternative, and perhaps more attractive, approach could be to allow for linear approximation models in which not only the spread of their kernel is adapted to the given data resolution, but the kernel itself as well. The above important questions will be answered and discussed in the remaining sections of this chapter.

4.3 Optimal Linear Approximation and Data Resolution

In this section, the optimal linear estimation problem for an unknown deterministic field $g(x) \in L^2(\mathfrak{R})$ will be solved in such a way that the immediate connection between the approximated signal and the available data resolution will explicitly appear in the solution formula. In particular, the final optimal estimate $\hat{g}(x)$ will depend only on a basic kernel $\varphi(x) \in L^2(\mathfrak{R})$, which is appropriately scaled to match the given data resolution level.

This scaling property will not be a-priori assigned to the linear approximation model, but it will rather result from a certain optimization principle that is going to be imposed in the estimation procedure.

4.3.1 General Formulation

We will assume that the available 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 $L^2(\mathfrak{R}^n)$ Hilbert space) is just a straightforward extension of the following derivations.

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

$$\hat{g}(x) = \sum_n g(nh) \varphi_{n,h}(x) \quad (4.2)$$

where $\varphi_{n,h}(x)$ is a family of unknown base functions which should be optimally selected to approximate $g(x)$. The dependence of these base functions on the data resolution is initially 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 transformations of the reference system that include both translation and rotational parameters), then the family $\varphi_{n,h}(x)$ should be generated from a single kernel $\varphi_h(x)$, such that

$$\varphi_{n,h}(x) = \varphi_h(x-nh) \quad (4.3)$$

and eq.(4.2) becomes

$$\hat{g}(x) = \sum_n g(nh) \varphi_h(x-nh) \quad (4.4)$$

The estimation formula in eq.(4.4) can be illustrated in terms of the linear filtering procedure shown in Figure 4.2. Applying the Fourier transform to the above convolution equation, we get

$$\hat{G}(\omega) = \bar{G}_h(\omega) \Phi_h(\omega) \quad (4.5)$$

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

$$\bar{g}_h(x) = g(x) \sum_n \delta(x-nh) = \sum_n g(nh) \delta(x-nh) \quad (4.6a)$$

and it has the periodic form (Oppenheim and Schaffer, 1989)

$$\bar{G}_h(\omega) = \frac{1}{h} \sum_k G\left(\omega + \frac{2\pi k}{h}\right) = \sum_n g(nh) e^{-i\omega nh} \quad (4.6b)$$

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

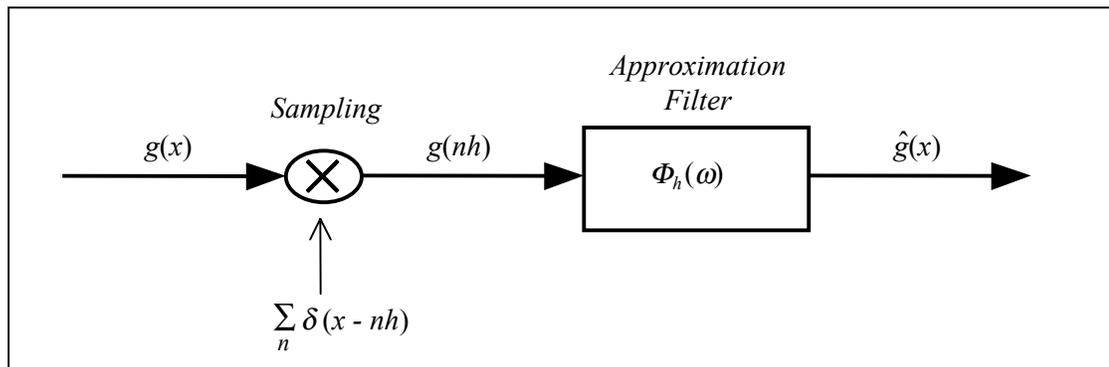


Figure 4.2 Filtering configuration of linear translation-invariant signal approximation using discrete samples

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 a limited part of the signal 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 support of the unknown field 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 approximation studies in spatial fields. It should also be emphasized that, even though $g(x)$ is assumed zero outside the given grid boundaries, its approximation $\hat{g}(x)$ by eq.(4.4) may exhibit a non-zero pattern outside the data grid. Of course, the theoretical case of infinitely extended 1D 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 that the available data sequence $g(nh)$ is always measurable in the following sense:

$$\sum_n |g(nh)|^2 < \infty \quad (4.7)$$

Indeed, under such a condition the Fourier transform $\overline{G}_h(\omega)$ of the data sequence in eq.(4.6b) will always converge to a finite periodic function of ω (Oppenheim and Schaffer, 1989; p. 48). Note that eq.(4.7) implies that the admissible observation space is the whole $l^2(\mathbb{Z})$ Hilbert space, which is in agreement with the MR refinement of the linear approximation problem that was suggested at the end of the previous chapter.

4.3.2 A Spatio-Statistical Optimal Principle

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

$$e(x) = g(x) - \hat{g}(x), \quad E(\omega) = G(\omega) - \hat{G}(\omega) \quad (4.8)$$

and its power spectrum can easily be derived by taking eq.(4.5) into account, i.e.

$$\begin{aligned} |E(\omega)|^2 = E(\omega) E^*(\omega) = G(\omega) G^*(\omega) - \Phi_h^*(\omega) G(\omega) \bar{G}_h^*(\omega) - \\ - \Phi_h(\omega) \bar{G}_h(\omega) G^*(\omega) + \Phi_h(\omega) \Phi_h^*(\omega) \bar{G}_h(\omega) \bar{G}_h^*(\omega) \end{aligned} \quad (4.9)$$

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. If we shift the sampler (or impulse train) $\sum_n \delta(x-nh)$ by an amount x_o , an infinite number of different data sequences can be obtained, which all represent different sampling schemes for the same unknown signal at the same resolution. The situation is illustrated in Figure 4.3, from which we can see that (at a specific resolution 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 $-h/2$ and $h/2$.

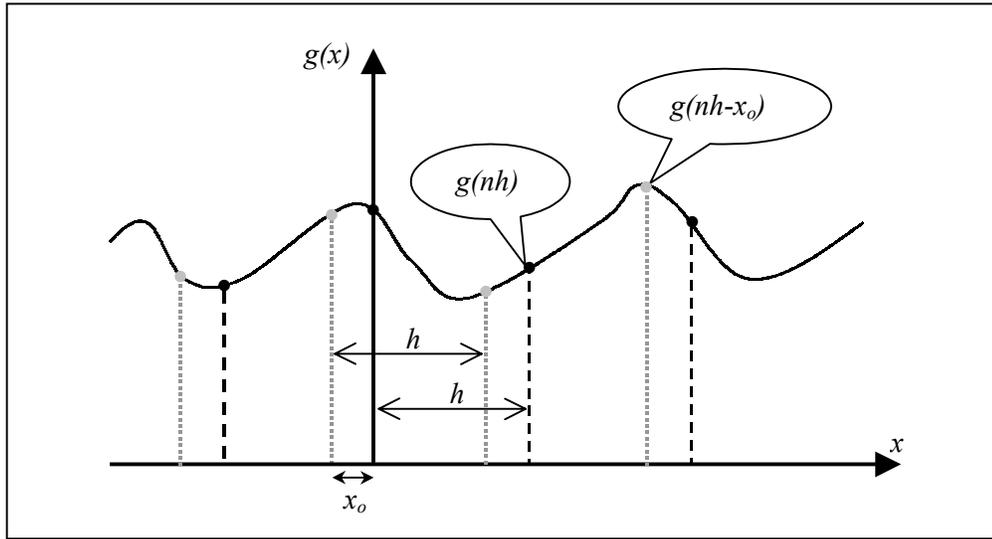


Figure 4.3 Different signal sampling configurations at a given resolution level h

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

$$\hat{g}(x, x_0) = \sum_n g(nh - x_0) \varphi_h(x + x_0 - nh) \quad (4.10)$$

The Fourier transform of eq.(4.10), considered as a function of x only, yields

$$\hat{G}(\omega, x_0) = \frac{1}{h} \Phi_h(\omega) \sum_k G(\omega + \frac{2\pi k}{h}) e^{-i\frac{2\pi k}{h}x_0} \quad (4.11)$$

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.(4.7). Hence, for each different

sampling phase value x_o , we will have a correspondingly different approximation error

$e(x, x_o)$, i.e.

$$e(x, x_o) = g(x) - \hat{g}(x, x_o) \quad (4.12a)$$

whose Fourier transform is

$$E(\omega, x_o) = G(\omega) - \frac{1}{h} \Phi_h(\omega) \sum_k G(\omega + \frac{2\pi k}{h}) e^{-i\frac{2\pi k}{h}x_o} \quad (4.12b)$$

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

$$P_e(\omega) = \frac{1}{h} \int_{-h/2}^{h/2} |E(\omega, x_o)|^2 dx_o = \min \quad (4.13)$$

The above equation represents a minimum mean square error (MMSE) principle, expressed in the frequency domain. The quantity $P_e(\omega)$ is nothing other than the **mean error power spectrum**. Note that the term ‘mean’ is not used in a probabilistic sense (as in Wiener’s linear prediction theory), but it has a rather spatio-statistical meaning. In other words, the optimization of the linear estimation algorithm does not employ the usual expectation operator considering different ‘experiment repetitions’, but it is based on the average error over all possible sampling configurations for the given data resolution level. This is exactly the logic behind the concept of statistical collocation that

was briefly discussed in section 3.1.3; see also Sanso (1978). In Appendix A, it is proven that

$$\int_{-h/2}^{h/2} |E(\omega, x_o)|^2 dx_o = h C(\omega) - \Phi_h^*(\omega) C(\omega) - \Phi_h(\omega) C(\omega) + \Phi_h(\omega) \Phi_h^*(\omega) \bar{C}_h(\omega) \quad (4.14)$$

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

$$c(x) = \int g(y) g(y+x) dy \quad \xleftrightarrow{\mathfrak{S}} \quad C(\omega) = G(\omega) G^*(\omega) = |G(\omega)|^2 \quad (4.15)$$

where the symbol \mathfrak{S} in the last equation denotes a Fourier transform pair. The term $C(\omega)$ is thus the usual signal power spectrum, and the term $\bar{C}_h(\omega)$ in eq.(4.14) denotes its following periodization (see Appendix A):

$$\bar{C}_h(\omega) = \frac{1}{h} \sum_k C(\omega + \frac{2\pi k}{h}) \quad (4.16)$$

Using equations (4.13) and (4.14), we can finally obtain the optimal estimation filter as follows:

$$\Phi_h(\omega) = \frac{C(\omega)}{\bar{C}_h(\omega)} = h \frac{C(\omega)}{\sum_k C(\omega + \frac{2\pi k}{h})} \quad (4.17)$$

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

$$\varphi_h(x) = \varphi\left(\frac{x}{h}\right) \quad (4.18)$$

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

$$\varphi(x) \xleftrightarrow{\mathfrak{F}} \Phi(\omega) = \frac{C\left(\frac{\omega}{h}\right)}{\sum_k C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right)} \quad (4.19)$$

The above result can easily be verified by taking into account the fundamental scaling property of the Fourier transform (Bracewell, 1986). Finally, if we combine eqs.(4.4) and (4.18), the optimal linear approximation formula for an unknown deterministic field $g(x)$ according to the MMSE principle (4.13), using its discrete samples on a uniform grid with resolution level h , will have the wavelet-like form

$$\hat{g}(x) = \sum_n g(nh) \varphi\left(\frac{x}{h} - n\right) \quad (4.20)$$

It is worth mentioning that the basic reconstructing kernel $\varphi(x)$ will always be a real *symmetric* function, since its Fourier transform in eq.(4.19) is always real-valued and symmetric (i.e. the signal power spectrum $C(\omega)$ is always a real-valued positive symmetric function).

4.3.3 Comments

The use of convolution-based linear estimation models of the form of eq.(4.20) is very common in many signal processing applications in the context of classical interpolation, quasi-interpolation and multiscale approximation through projections into MR subspaces (Keys, 1981; Unser and Daubechies, 1997; Blu and Unser, 1999a,b). In such cases, the selection of the basic kernel $\varphi(x)$ is usually made a-priori (e.g. sinc-based interpolation, polynomial spline interpolation, etc.), and its performance is evaluated according to an assumed behaviour for the unknown signal (e.g. bandlimitedness, spectrum decay rate, etc.) and/or certain theoretical error bounds that depend on the form of the used kernel (Strang-Fix conditions); for more details, see Unser and Daubechies (1997). Here, on the other hand, we have a-priori introduced a spatio-statistical error power spectrum as an accuracy measure for the linear estimation algorithm, which is then optimized in order to choose the best kernel $\varphi(x)$ for the given unknown signal $g(x)$. Furthermore, the

translation-invariance condition that was imposed in the estimation procedure makes this optimal kernel depend only on the spatial CV function (or, equivalently, on the power spectrum) of the unknown field under consideration, according to eq.(4.19). The additional dependence of $\varphi(x)$ on the data resolution level, as it is evident from eq.(4.19), will be discussed in detail in the next section.

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

$$\varphi(n) = \begin{cases} 1 & , n = 0 \\ 0 & , n = \pm 1, \pm 2, \pm 3, \dots \end{cases} \quad (4.21a)$$

Indeed, using eq.(4.19) we easily see that the Fourier transform $\Phi(\omega)$ of the optimal kernel satisfies the relation

$$\sum_n \Phi(\omega + 2\pi n) = \frac{\sum_n C\left(\frac{\omega + 2\pi n}{h}\right)}{\sum_k C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right)} = 1 \quad (4.21b)$$

which assures, through the well known Poisson summation formula, that the corresponding space-domain function $\varphi(x)$ is a sampling function. Some mild technical

conditions on the signal power spectrum $C(\omega)$, which are needed to ensure the validity of eq.(4.21b), will be discussed later in this chapter.

The use of the optimal kernel $\varphi_h(x) = \varphi(x/h)$ also ensures the convergence of the linear interpolation algorithm to the true field, as the data resolution increases. This can easily be seen in the frequency domain using the Fourier transform $\Phi_h(\omega)$ of the optimal approximation kernel, given in eq.(4.17). If we take into account eqs.(4.5) and (4.6b), we have

$$\hat{G}(\omega) = \Phi_h(\omega) \frac{1}{h} \sum_k G(\omega + \frac{2\pi k}{h}) \quad (4.21c)$$

and by using the optimal filter expression from eq.(4.17), we finally obtain

$$\hat{G}(\omega) = \frac{C(\omega)}{\sum_k C(\omega + \frac{2\pi k}{h})} \sum_k G(\omega + \frac{2\pi k}{h}) \quad (4.21d)$$

From the last equation, we can see that the estimated signal will converge to the true signal in the L^2 sense, as $h \rightarrow 0$.

An interesting similarity exists between the optimal filter in eq.(4.17) and the *Wiener filter* for noisy stationary random signals. According to Wiener's linear prediction theory,

the optimal estimation filter is defined as the ratio between the power spectral densities (PSDs) of the noiseless stochastic signal and the noisy input signal (Sideris, 1995). This is very similar to eq.(4.17), 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 spatial CV function of the ‘noisy’ input signal $\bar{g}_h(x)$; see eq.(4.6a). In our case, the noise takes the form of the lost information due to the discretization of the original true signal (aliasing error), shown in Figure 4.2.

It should be noted that, apart from the previous informal algorithmic similarity with the Wiener filter theory, no stochastic tools have been used in the present filtering formulation for the signal estimation problem. Its optimal solution has been based on entirely different concepts and assumptions from the ones found in linear prediction theory of random fields (Christakos, 1992). The term *covariance function*, that has been used throughout this section, should be understood in a purely spatial deterministic sense and not in any stochastic/probabilistic context under some stationarity and ergodicity assumption. This is especially important in view of the stationarity restriction problem which is believed to exist in the statistical collocation framework. Our present formulation can be considered as ‘stationary’ only in the sense that we use the same kernel at every data point to describe the behaviour of the estimated signal; see eq.(4.20). This results solely from the logical requirement of having a translation-invariant approximation scheme, i.e. independent of the origin of the reference system used to

describe the position of the data points; see also the related discussion given in Sanso (1978). However, this does not imply that the signal has (or should have) a uniform-stationary behaviour across its domain, and it certainly does not inhibit us from obtaining localized information for its varying behaviour. On the other hand, if we choose to use an algorithmically ‘non-stationary’ linear approximation model, where the associated kernel should change from data point to data point, then we automatically lose the invariance property under translations (or more general rigid transformations in the multi-dimensional case) of the spatial reference system.

4.4 The Multiresolution Character of Statistical Collocation

The final result of the previous section is quite general and it did not involve any particular concepts from Mallat’s multiresolution theory. It is interesting that the statistical collocation principle 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 data points), similar to the one encountered in wavelet theory. However, there is a significant difference between the collocation-based model of eq.(4.20) and the classic MRA-based approximation methodology, due to the fact that the optimal scaling kernel $\varphi(x)$ associated with the collocation case is now changing for every different data resolution level h , according to the frequency-domain form given in eq.(4.19).

The most appropriate way to describe the behaviour of the signal estimation model in eq.(4.20), with its associated kernel defined by eq.(4.19), 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 used to describe the spatial position of a given set of gridded data points, the estimated field according to the statistical collocation algorithm will always have the same form/shape.

Let us briefly demonstrate the scale-invariance aspect (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 the estimation problem is reduced to approximating a new unknown field $g'(x) = g(ax)$ using its point data values $g'(nh') = g'(nh/a) = g(nh)$. The application of the collocation formula in eq.(4.20) yields

$$\begin{aligned}\hat{g}'(x) &= \sum_n g'(nh') \varphi\left(\frac{x}{h'} - n\right) = \sum_n g'\left(n\frac{h}{a}\right) \varphi\left(\frac{x}{h/a} - n\right) \\ &= \sum_n g(nh) \varphi\left(\frac{ax}{h} - n\right) = \hat{g}(ax)\end{aligned}\tag{4.22}$$

which demonstrates the scale-invariance property of statistical collocation. Note that the sampling resolution of the unknown field 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, is figuratively illustrated in Figure 4.4.

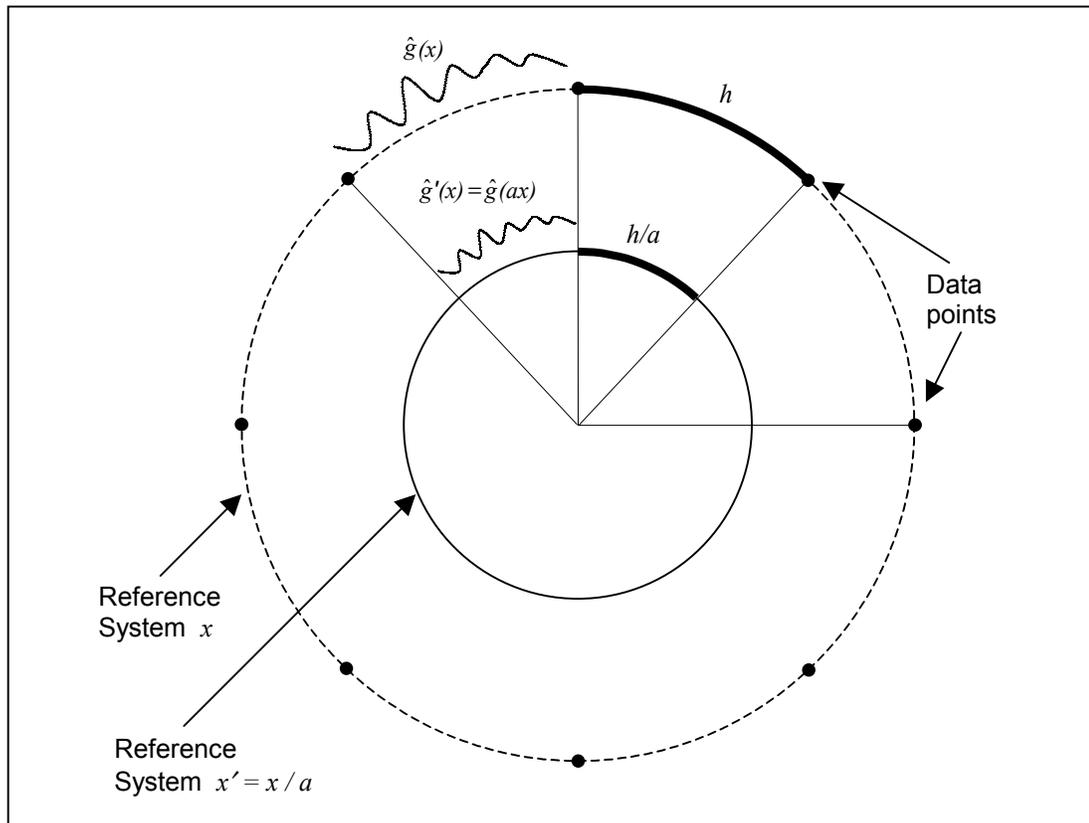


Figure 4.4 Scale-invariant signal approximation at a certain data resolution level h (the value of the scaling parameter is assumed $a > 1$)

The optimal kernel $\varphi(x)$ in the statistical collocation model of eq.(4.20) is appropriately scaled (shrunk or expanded) in order to match the resolution level 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 scheme very closely

obeys the classic multiresolution/wavelet spirit, it cannot really be identified as such because the actual form of $\varphi(x)$ is a function of the data resolution h itself. On the other hand, the standard MRA theory requires the use of a *fixed* scaling kernel, which is just tuned to the desired scale level of the signal approximation by proper dyadic dilations (see Chapter 2).

In order to better understand the above essential difference, we should express the optimal scaling kernel associated with the statistical collocation in the following parameterized form [see eq.(4.19)]:

$$\varphi(x, h) \xleftrightarrow{\mathfrak{F}} \Phi(\omega, h) = \frac{C(\frac{\omega}{h})}{\sum_k C(\frac{\omega}{h} + \frac{2\pi k}{h})} \quad (4.23a)$$

where the data resolution h plays the role of a constant parameter. According to the fundamental scaling property of the Fourier transform, the scaled version $\varphi(x/h) = \varphi_h(x)$ of the optimal kernel will have the following frequency-domain form:

$$\varphi(\frac{x}{h}, h) \xleftrightarrow{\mathfrak{F}} h\Phi(h\omega, h) = h \frac{C(\omega)}{\sum_k C(\omega + \frac{2\pi k}{h})} \quad (4.23b)$$

which is identical to the Wiener-like estimation filter $\Phi_h(\omega)$ that was determined in section 4.3.2. For each different value of the resolution parameter the optimal kernel in

eq.(4.23a) will assume a correspondingly different waveform, and thus the approximation model of eq.(4.20) will not employ scaled versions of the same $\varphi(x)$ for every data resolution level. Hence, we see that the statistical collocation concept not only produces a scale-invariant signal approximation, but it also forces the behaviour of its estimation kernel to be adapted to the sampling resolution of the unknown field in a certain optimal fashion, as suggested by eq.(4.23a). In contrast to the classic MRA methodology, this clearly provides a ‘non-stationary’ treatment of the linear interpolation problem with respect to the data resolution parameter, which may have a significant impact on the accuracy of the signal approximation (see Chapter 5 for some simulated comparisons). Finally, it is very important to note that, regardless of the actual value of h , the function $\varphi(x, h)$ will always correspond to a cardinal (sampling) kernel, as was explained in section 4.3.3.

4.5 Optimal Multiresolution Approximation Kernels Using Synthetic Signal Power Spectra

In order to see how the optimal approximation kernel $\varphi(x, h)$ behaves at different data resolution levels h , some simple simulation examples will be given in this section. Four different models for the power spectrum of the underlying unknown signal $g(x)$ are used. In particular, we consider the following cases:

Gaussian Power Spectrum

$$C(\omega) = B e^{-\omega^2} \quad (4.24a)$$

Exponential Power Spectrum

$$C(\omega) = B e^{-|\omega|} \quad (4.24b)$$

 $O(\omega^{-2})$ Power Spectrum

$$C(\omega) = \frac{B}{1+\omega^2} \quad (4.24c)$$

'Experimental' Power Spectrum

$$C(\omega) = B \frac{|\cos(4\omega)| + |\sin(\omega/3)|}{1+\omega^2} \quad (4.24d)$$

where B denotes an arbitrary scale factor in all cases. The Gaussian power spectrum has the fastest decay rate of all four models. The third and the fourth models have basically the same slow asymptotic decay, with the 'experimental' power spectrum showing higher frequency variations that may be expected in many practical situations. The exponential model exhibits an intermediate decaying pattern which, initially, is faster than the Gaussian (up to $\omega = 1$). After that point, it starts to decay much slower than the Gaussian model but still a bit faster than the other two models. All four models for the signal power spectrum $C(\omega)$ are illustrated in Figure 4.5.

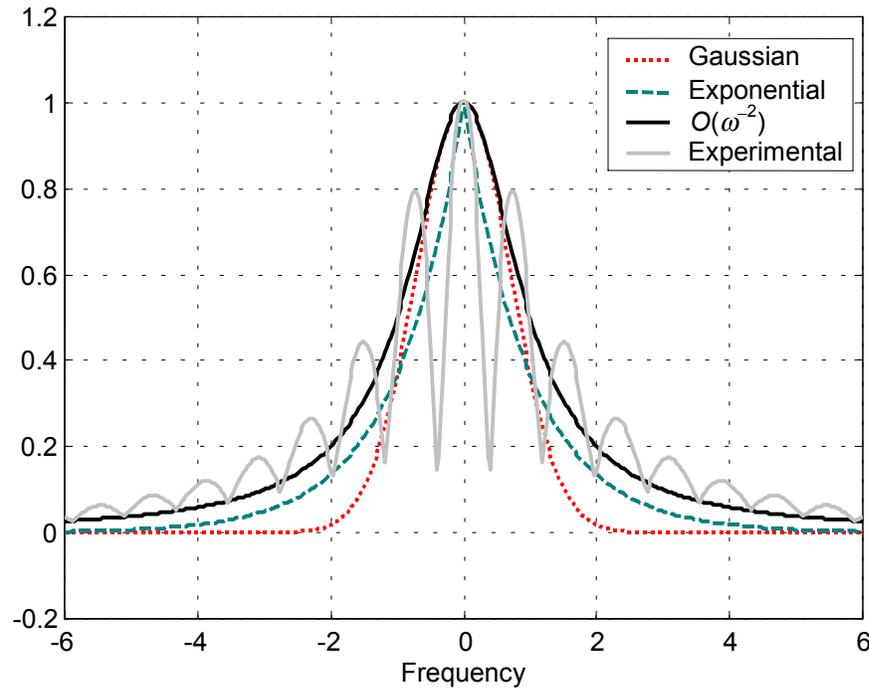


Figure 4.5 Various models for the signal power spectrum $C(\omega)$

The varying behaviour of the optimal interpolation filter according to eq.(4.23a), for some selected data resolution levels, is shown in Figures 4.6 and 4.7. In particular, Figure 4.6 shows the Fourier transform $\Phi(\omega, h)$ for the case where the signal power spectrum follows either a Gaussian model (left column), or an exponential model (right column). Accordingly, the left column in Figure 4.7 illustrates the optimal interpolation filter for the case where $C(\omega)$ behaves as the model given in eq.(4.24c), whereas the right column in the same figure corresponds to the case where $C(\omega)$ behaves as the experimental model given in eq.(4.24d).

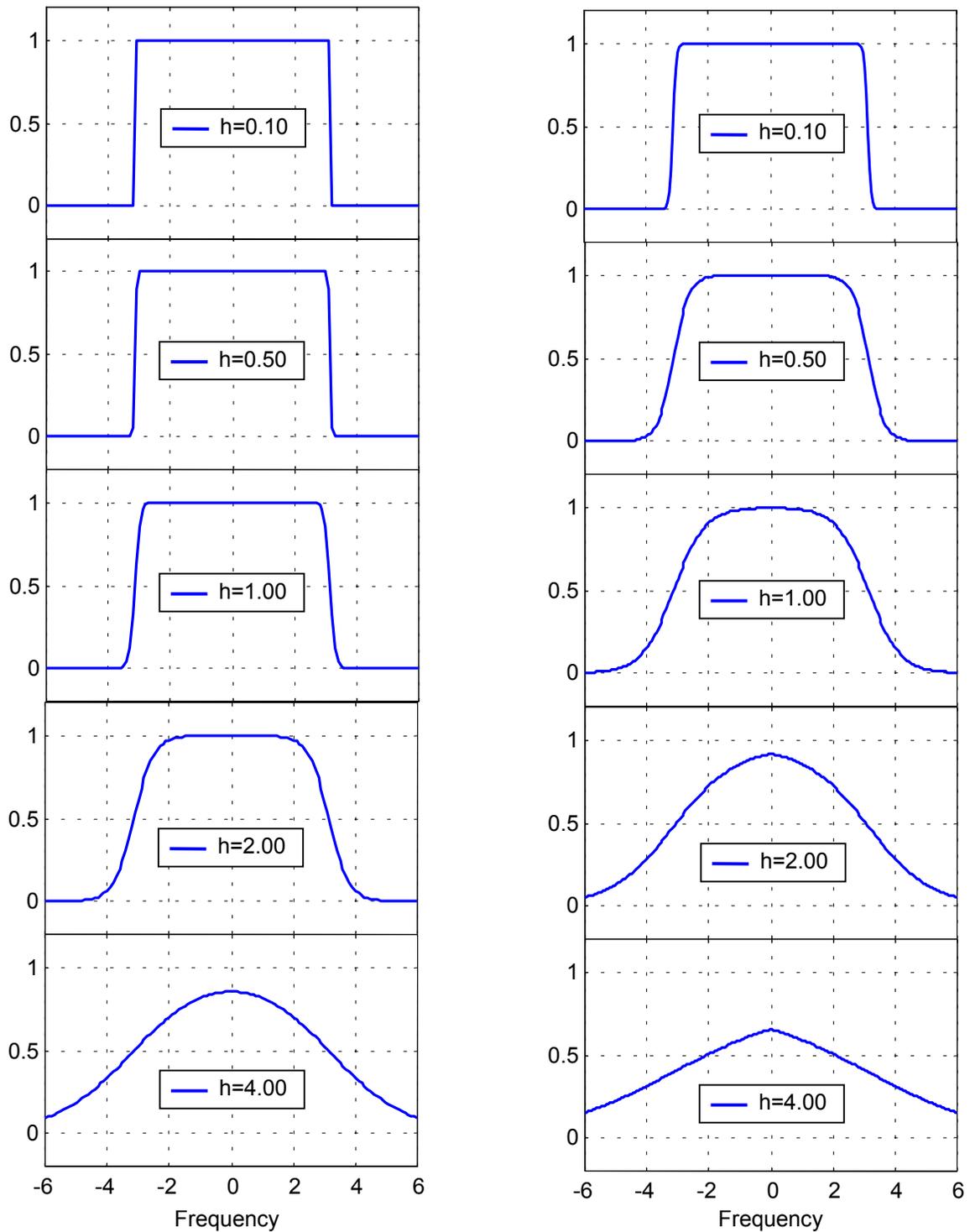


Figure 4.6 Fourier transform $\Phi(\omega, h)$ of the optimal approximation kernel for various data resolution levels h . The left column corresponds to a Gaussian model for the signal power spectrum, whereas the right column corresponds to an exponential model.

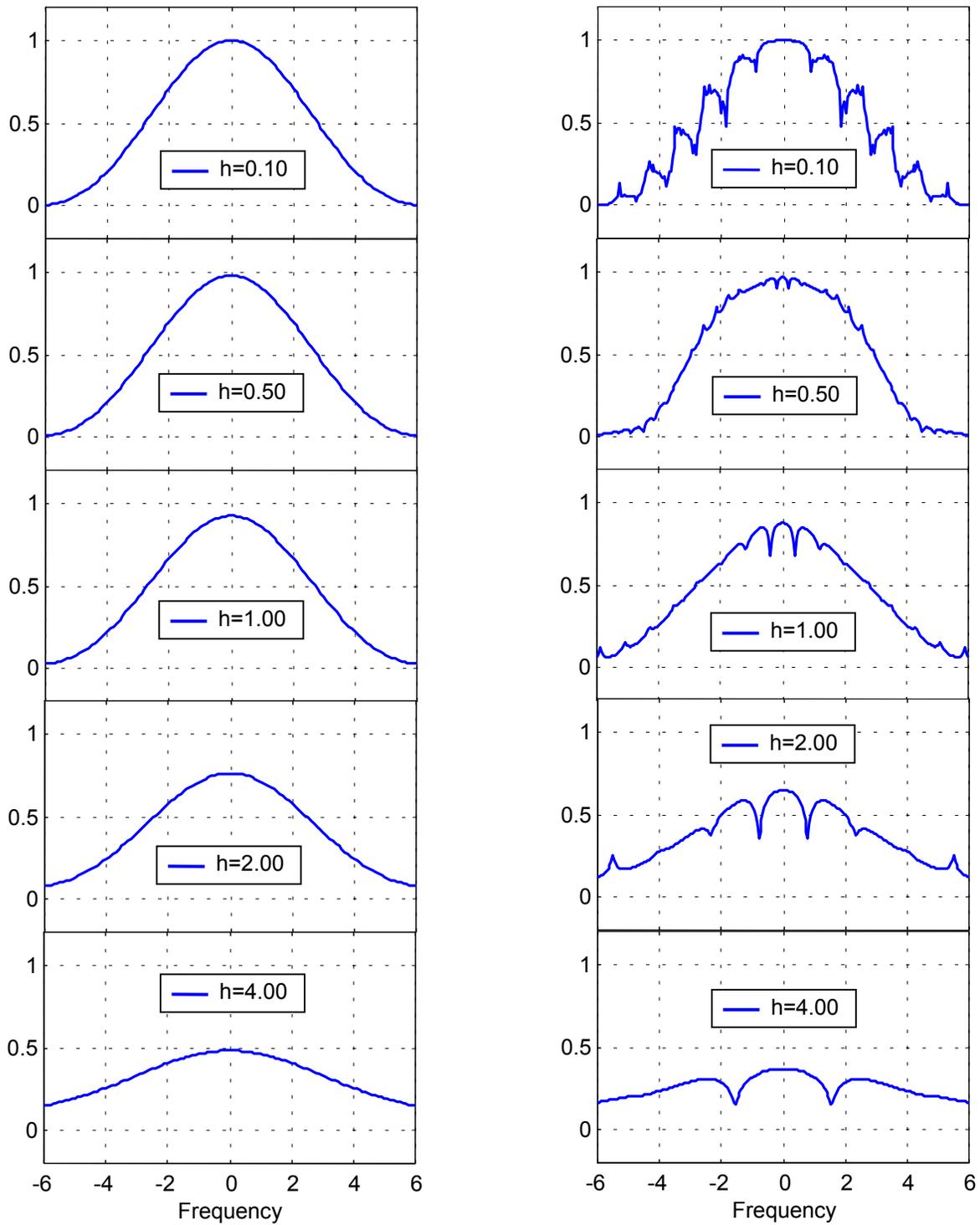


Figure 4.7 Fourier transform $\Phi(\omega, h)$ of the optimal approximation kernel for various data resolution levels h . The left column corresponds to the signal power spectrum model given by eq.(4.24c), whereas the right column corresponds to the model given by eq.(4.24d).

These plots help considerably in understanding the (somewhat peculiar) multiresolution behaviour of the optimal interpolation kernel in the statistical collocation framework. Under proper mild conditions on the signal power spectrum, the estimation kernel $\varphi(x, h)$ will asymptotically converge to an $L^2(\mathfrak{R})$ cardinal function as $h \rightarrow 0$. All the individual kernels of this convergent sequence will be $L^2(\mathfrak{R})$ sampling functions as well.

In the case of Figure 4.6, for example, it is obvious that the spatial expression $\varphi(x, h)$ of the optimal estimation filter will gradually converge to the *sinc* interpolator, since for high data density the Fourier transform $\Phi(\omega, h)$ tends to a perfect low-pass filter over the frequency band $[-\pi, \pi]$. This result is achieved for both the Gaussian and the exponential signal models, with the latter case showing a slower rate of kernel convergence due to the slower asymptotic decay of the exponential power spectrum over the Gaussian. Despite this interesting fact, it would be improper to conclude that band-limited signal interpolation will yield an almost optimal level of accuracy, at any data resolution level, for these two specific cases (signal models). As a matter of fact, *sinc*-based interpolation for Gaussian/exponential-type signals performs very poorly as h increases, compared to the actual optimal interpolation filters shown in Figure 4.6, as well as to other spline-based kernels (see Chapter 5 for such comparisons). In the case of Figure 4.7, where the implied signal power spectra decay even slower than the Gaussian and the exponential models, the asymptotic form of the optimal collocation filter will taper off less quickly than the perfect (Nyquist-based) low-pass estimation filter. Such a result is in agreement

with the short-kernel approximation methods that practitioners usually employ in signal processing applications (Unser and Daubechies, 1997; Keys, 1981), as well as with the general *wavelet spirit* that suggests good space localization as an effective kernel property for approximating the details of irregular signals with high frequency variations. Nevertheless, we should mention once more that the optimal collocation kernel is essentially different from the classic MRA scaling functions that appear in wavelet theory, since its behaviour depends explicitly on the actual data resolution level.

On the other hand, as the sampling resolution decreases ($h \rightarrow \infty$), the optimal kernel $\varphi(x, h)$ will gradually become the zero function in the $L^2(\mathfrak{R})$ sense. This is evident from the behaviour of its Fourier transform in both Figures 4.6 and 4.7. Such kind of behaviour is consistent with the general causality principle for the estimation procedure, which was mentioned in the beginning of this chapter. The rigorous mathematical proofs of the above statements, as well as the derivation of some necessary mild conditions on the signal power spectrum $C(\omega)$, are beyond the scope of this thesis and they will not be presented here. Some relevant details can be found in the next section.

4.6 Generalized Multiresolution Analysis

In this section, we will explore in more detail the connection between the statistical collocation model of eqs.(4.19) and (4.20), and the multiresolution approximation

framework which was presented in Chapter 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 conditions on the spatial CV function and the power spectrum of the unknown signal $g(x)$, the corresponding optimal estimation kernel $\varphi(x, h)$ produces a generalized MRA-type approximation scheme in the Hilbert space $L^2(\mathfrak{R})$.

4.6.1 MRA Properties of the Optimal Collocation Kernel

First, we need to establish that the optimal estimation kernel in statistical collocation, as given in eq.(4.19) or eq.(4.23a), is a well defined function in the $L^2(\mathfrak{R})$ Hilbert space for any positive value of the data resolution h . Using eq.(4.23a), the L^2 norm of the optimal kernel takes the following form:

$$\begin{aligned} \|\varphi(x, h)\|_{L^2}^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |\Phi(\omega, h)|^2 d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\left|C\left(\frac{\omega}{h}\right)\right|^2}{\left|\sum_k C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right)\right|^2} d\omega \\ &= \frac{1}{2\pi} \int_0^{2\pi} \frac{\sum_k \left|C\left(\frac{\omega + 2\pi k}{h}\right)\right|^2}{\left|\sum_k C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right)\right|^2} d\omega = \frac{1}{2\pi} \int_0^{2\pi} \frac{\sum_k \left(\frac{1}{h} C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right)\right)^2}{\left(\sum_k \frac{1}{h} C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right)\right)^2} d\omega = \end{aligned}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \bar{M}_{2\pi}(\omega) d\omega \quad (4.25)$$

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

$$\bar{M}_{2\pi}(\omega) = \frac{\sum_k \left(\frac{1}{h} C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right) \right)^2}{\left(\sum_k \frac{1}{h} C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right) \right)^2} = \frac{\sum_k a_k^2}{\left(\sum_k a_k \right)^2} \quad (4.26a)$$

and the discrete infinite sequence a_k is given by the general expression

$$a_k = \frac{1}{h} C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right) \quad , \quad k \in \mathbb{Z} \quad (4.26b)$$

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

$$b[n] = c(nh) \quad (4.27)$$

where $c(x)$ is the signal spatial CV function; see eq.(4.15). Therefore, if the 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 Schaffer 1989, p. 47). In this way, we will impose the following condition on the signal spatial CV function:

$$\text{CONDITION I: } \sum_n |c(nh)| < \infty, \quad \forall h > 0 \quad (4.28)$$

Note that the above condition is always satisfied when the underlying unknown field $g(x)$ has a finite support in the space domain. A simple example of a CV function with infinite support, for which the above condition is valid, is the Gaussian function. Under condition (4.28), the series $\sum_k a_k$ will converge uniformly for every value of ω and h , and since *all* its individual terms 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. It is also essential to ensure the validity of the following relationship:

$$\sum_k a_k = \sum_k \frac{1}{h} C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right) \neq 0, \quad \forall \omega \in \mathfrak{R}, h > 0 \quad (4.29)$$

There are various types of conditions that can be imposed on the signal power spectrum, in order for eq.(4.29) to be true. For the purpose of this thesis, we shall simply assume one of the following:

CONDITION II : (4.30)

- a. $C(\omega) = |G(\omega)|^2 > 0$, $\forall \omega \in \mathfrak{R}$ **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.(4.29), as long as these infinite points are not equidistant.

The justification of the previous restrictions on the signal power spectrum depends on the physical properties of the unknown field that we want to estimate. The case where the signal power spectrum vanishes in an unbounded frequency interval (i.e. the unknown field $g(x)$ is a *band-limited* signal) requires special consideration, and it will not be treated here.

If we further assume that the signal power spectrum is a continuous function, i.e.

CONDITION III : $C(\omega)$ is continuous for every $\omega \in \mathfrak{R}$ (4.31)

then, under the three previous conditions, the auxiliary term $\overline{M}_{2\pi}(\omega)$ in eq.(4.26a) will always converge to a finite, *strictly-positive, continuous* and *2 π -periodic* function, and therefore its integral in eq.(4.25) will always be a finite number. This makes the optimal

approximation kernel $\varphi(x, h)$ a proper $L^2(\mathfrak{R})$ function for any real positive value of the data resolution level h .

Finally, the condition that the optimal kernel 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.(4.23a), this is transformed to the following simple condition for the signal power spectrum:

$$\text{CONDITION IV : } C(\omega)_{\omega=0} \neq 0 \quad (4.32)$$

We are now in a position to consider a certain infinite sequence $\{V_j\}_{j \in \mathbb{Z}}$ of linear subspaces within the Hilbert space $L^2(\mathfrak{R})$. Each element of this sequence is defined as the closed linear span of the set $\varphi(\frac{x}{h_j} - n, h_j)_{n \in \mathbb{Z}}$, where $\varphi(x, h_j)$ is the optimal collocation kernel given by eq.(4.23a), and $h_j > 0$ denotes the data resolution level associated with each subspace V_j . We will further assume that

$$\text{CONDITION V : } h_j > h_{j+1} \quad , \quad \forall j \in \mathbb{Z} \quad (4.33)$$

which makes $\{V_j\}$ a subspace sequence of increasing resolution in $L^2(\mathfrak{R})$. Note that the scaling parameter is not restricted to dyadic values only (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.(2.3), section 2.1) for any possible form of the scaling parameter h_j , i.e.

$$f(x) \in V_j \Leftrightarrow f(x + nh_j) \in V_j \quad , \quad \forall \quad n \in \mathbb{Z} \quad (4.34)$$

In order for the specific sequence $\{V_j\}$ to satisfy the nesting property of an MRA, we have to impose some additional restrictions on the way that 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 $j \in \mathbb{Z}$

$$\text{CONDITION VI: } \frac{h_j}{h_{j+1}} = a_j \quad , \quad \text{where } a_j \in \mathbb{Z}^+ - \{1\} \quad (4.35)$$

The above condition implies that any two successive scaling parameters 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.(4.35) 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:

$$\frac{h_j}{h_k} = a_j a_{j+1} \dots a_{k-1}, \quad \forall j < k \in \mathbb{Z} \quad (4.36)$$

The special case where the scaling ratio in eq.(4.35) assumes a fixed positive integer value a (independent from the resolution index j) occurs if we restrict the data resolution level h_j to take the exponential form

$$h_j = a^{-j} \quad (4.37)$$

where a is now a fixed positive integer number, different from unity. Dyadic subspace schemes (as in the classic MRA framework) will arise if we set the value of a to be equal to 2. Nevertheless, the more general condition of eq.(4.35) is all that we actually need in order for the specific subspace sequence $\{V_j\}$ to be nested. The proof is straightforward and it can be found in Appendix B.

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

$$f(x) = \sum_n f(nh_j) \varphi\left(\frac{x}{h_j} - n, h_j\right), \quad \forall f(x) \in V_j \quad (4.38a)$$

Taking into account eq.(4.23a) and applying the Fourier transform to the last equation, we obtain the equivalent frequency-domain form of every signal belonging in an arbitrary subspace of the sequence $\{V_j\}$, i.e.

$$F(\omega) = \frac{C(\omega)}{\sum_k C(\omega + \frac{2\pi k}{h_j})} \sum_n F(\omega + \frac{2\pi n}{h_j}), \quad \forall f(x) \in V_j \quad (4.38b)$$

As the resolution index j increases, the data sampling level h_j associated with the corresponding subspace becomes smaller and smaller, according to the general condition imposed by eq.(4.33). Obviously, when h_j becomes infinitely small the last equation will be reduced to a simple identity, i.e.

$$F(\omega) = F(\omega) \quad (4.39a)$$

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

$$\lim_{j \rightarrow \infty} V_j = L^2(\mathfrak{R}) \quad (4.39b)$$

Lastly, we have to check if the family of scaled translates

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

of the optimal estimation kernel forms a *Riesz basis* for every element $V_j \subset L^2(\mathfrak{R})$ of the multiresolution subspace sequence that is spanned by this set. This final MRA property (see eq.(2.5), section 2.1) is especially important since it will ensure *stable* and *unique* signal approximation schemes from their discrete samples within every solution subspace V_j . A necessary and sufficient condition for this last property is

$$0 < A \leq \sum_k \left| \Phi(\omega + 2\pi k, h_j) \right|^2 \leq B < +\infty, \quad \forall \omega \in \mathfrak{R}, h_j > 0 \quad (4.40)$$

where A and B are some strictly positive bounds, and $\Phi(\omega, h_j)$ is the Fourier transform of the optimal kernel $\varphi(x, h_j)$ at data resolution level h_j . If we take into account eq.(4.23a), the above Riesz condition can be easily expressed as a function of the signal power spectrum in the following way:

$$0 < A \leq \frac{\sum_k \left(\frac{1}{h_j} C\left(\frac{\omega}{h_j} + \frac{2\pi k}{h_j}\right) \right)^2}{\left(\sum_k \frac{1}{h_j} C\left(\frac{\omega}{h_j} + \frac{2\pi k}{h_j}\right) \right)^2} = \overline{M}_{2\pi}(\omega) \leq B < +\infty \quad (4.41)$$

where $\overline{M}_{2\pi}(\omega)$ is the same 2π -periodic auxiliary function that was defined previously in eq.(4.26a). At the beginning of this section we established that (under conditions I, II and III) the term $\overline{M}_{2\pi}(\omega)$ will always converge to a *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 and the upper bound in the double inequality (4.41) is always guaranteed. Note that the actual numerical values of the two Riesz 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 collocation kernel will not necessarily be the same at each resolution level.

4.6.2 An Interesting Result

We have established the interesting result that: *the solution of the linear approximation problem for an unknown deterministic field $g(x) \in L^2(\mathfrak{R})$ from its discrete and regularly gridded samples, under the condition of translation-invariance and the spatio-statistical MMSE optimal principle (4.13), gives rise to a generalized MRA-type structure $\{V_j\}$ within the Hilbert space $L^2(\mathfrak{R})$.* The main difference between this multiresolution subspace structure and the classic MRAs according to Mallat (1989b) is that its scaling kernel does not have a fixed form, but it actually 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 generates the scaling function $\varphi(x, h_j)$ at each resolution value h_j , according to the optimal frequency-domain form given in eq.(4.23a). Certain conditions must also be satisfied by the spatial CV function and the power spectrum of the unknown signal, as discussed in the previous section.

The only traditional MRA property that will not necessarily be satisfied by the subspace sequence $\{V_j\}$ constructed through the optimal scaling kernel of statistical collocation is the ‘self-similar’ dyadic scaling condition between the individual subspaces, i.e.

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1} \quad (4.42)$$

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 sampling level) h_j decreases from one nested subspace V_j to the next V_{j+1} , based on the general formula of eq.(4.35).

Note that the optimal scaling kernel $\varphi(x, h_j)$ essentially generates not just a *single* nested sequence $\{V_j\}$ of dense MR subspaces in $L^2(\mathfrak{R})$ but an infinite number of such subspace sequences. Each one of them will depend on a specific formula that we choose to generate the various scale levels h_j , as well as on a specific *reference scale value* h_0 .

A list of such different alternatives is given in Table 4.1. The classic case where the

nested subspaces are associated with a fixed dyadic scaling parameter is shown in the last two columns of Table 4.1, for some selected reference scale values. Even for such dyadic scaling schemes, however, the self-similar property of eq.(4.42) will not necessarily be satisfied by the generalized MRA sequence $\{V_j\}$ associated with the optimal collocation kernel, unless we impose some further conditions on the signal power spectrum $C(\omega)$.

Table 4.1 Sample of scale level values h_j associated with different generalized MRA sequences $\{V_j\}$ that are produced from the same scaling kernel $\varphi(x, h_j)$

Individual Scale Levels	<i>SCALE LEVEL GENERATOR</i> [see eq.(4.35)]					
	$\frac{h_j}{h_{j+1}} = 2j^2 + 3$		$\frac{h_j}{h_{j+1}} = 2^{ j +1}$		$\frac{h_j}{h_{j+1}} = 2$	
	<i>REFERENCE SCALE VALUE</i>					
	$h_0 = 1$	$h_0 = 0.3$	$h_0 = 1$	$h_0 = 0.3$	$h_0 = 1$	$h_0 = 0.3$
h_3	1/165	1/550	1/64	3/640	1/8	0.0375
h_2	1/15	0.02	1/8	0.0375	1/4	0.075
h_1	1/3	0.1	0.5	0.15	1/2	0.15
h_0	1	0.3	1	0.3	1	0.3
h_{-1}	5	1.5	4	1.2	2	0.6
h_{-2}	55	16.5	32	9.6	4	1.2
h_{-3}	1155	346.5	512	153.6	8	2.4

It is worth mentioning that all the derivations in section 4.6.1 are valid even if the frequency-domain function $C(\omega)$ 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 (or

the signal CV function) in the construction of the estimation 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 estimate $\hat{g}(x)$ obtained by the statistical collocation algorithm will still converge in a stable way to the true field in the $L^2(\mathfrak{R})$ sense, as the data resolution increases ($h_j \rightarrow 0$). The optimal MMSE principle of eq.(4.13), however, will not be rigorously satisfied in such cases.

4.6.3 Stability of the Optimal Riesz Bases

A note should be made regarding the stability of the Riesz sampling bases which are constructed through the statistical collocation kernel $\varphi(x, h_j)$. In the usual dyadic MRA cases of Chapter 2, the stability level of a Riesz basis constructed from a certain scaling function $\varphi(x)$ remains the same within every multiresolution subspace V_j , and it is basically determined by the ratio (B/A) of the two bounds that appear in the frequency-domain Riesz condition of eq.(2.8). The ideal case occurs when the set $\varphi(x-n)_{n \in \mathbb{Z}}$ is also orthogonal (and thus $A = B$), which makes the reconstruction of a signal $f(x) \in V_j$ from its coefficients with respect to the corresponding orthogonal Riesz basis $\varphi(2^j x - n)_{n \in \mathbb{Z}}$ a perfectly stable linear process with condition number equal to one.

However, this situation of unperturbed stability (in the L^2 norm) does not apply in the generalized MRA structure that is constructed from the optimal collocation kernel. The reason is that its generating scaling function does not have a fixed form, but it varies for every nested subspace V_j according to eq.(4.23a). As a result, the two bounds in the generalized Riesz condition of eq.(4.40) will now depend on the scaling parameter h_j . In order to get an idea on how the condition number of the optimal Riesz bases changes within the generalized MRA $\{V_j\}$, the behaviour of the 2π -periodic function $\bar{M}_{2\pi}(\omega) = \sum_k \left| \Phi(\omega + 2\pi k, h_j) \right|^2$ in eq.(4.41) has been plotted for some selected scale levels h_j . We have used the same four synthetic models for the signal power spectrum that were considered in section 4.5. The results are shown in Figures 4.8 through 4.11 for the Gaussian, exponential, $O(\omega^{-2})$ and experimental signal model, respectively.

In all cases, the 2π -periodic function $\bar{M}_{2\pi}(\omega)$ does not exceed the value 1, which is actually expected due to the positivity of the signal power spectrum; see eq.(4.41). The condition number of the optimal Riesz bases $\varphi(x/h_j - n, h_j)_{n \in \mathbb{Z}}$, for every signal model and data resolution level h_j , is determined by the ratio of the maximum and minimum values in each corresponding curve shown in the previous figures. The value of this ratio will generally be larger than 1, and it will vary from one resolution level to the next for every signal model.

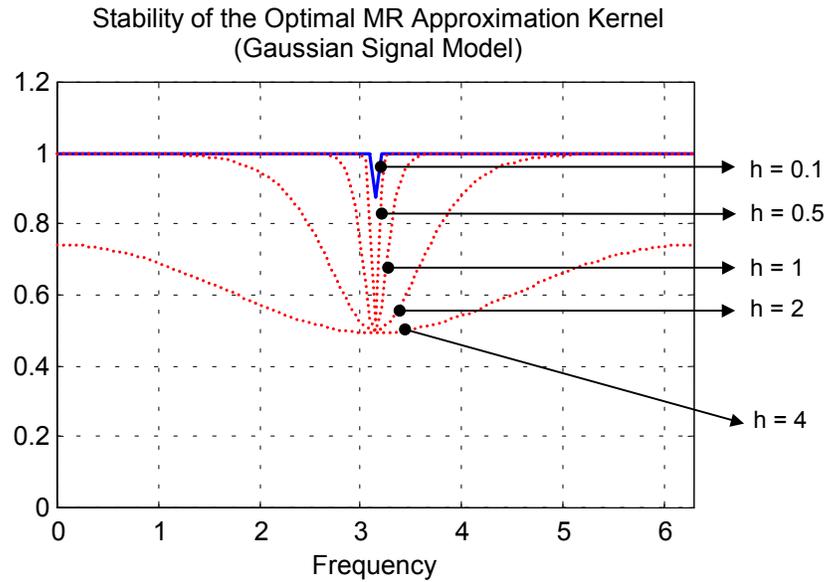


Figure 4.8 Plots of the 2π -periodic function $\bar{M}_{2\pi}(\omega)$, given in eq.(4.41), for various scale levels. The signal power spectrum $C(\omega)$ is assumed to follow a Gaussian model.

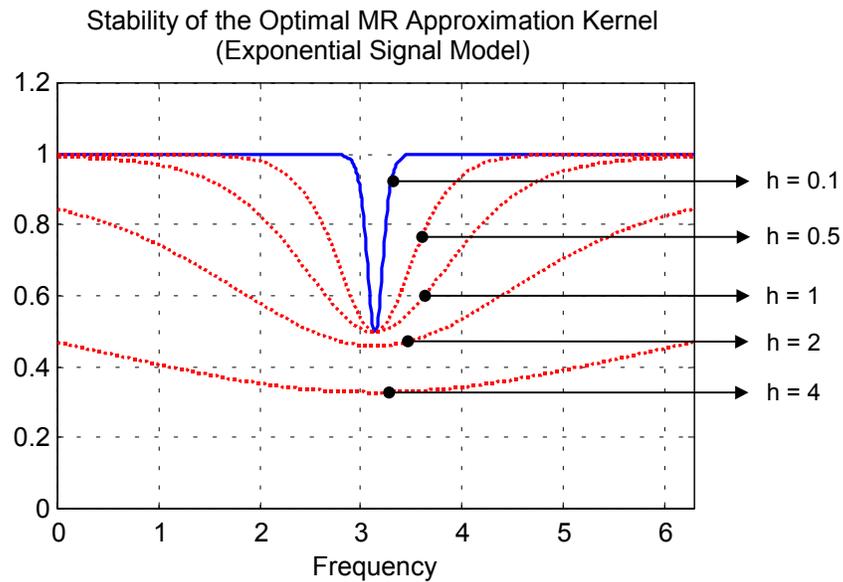


Figure 4.9 Plots of the 2π -periodic function $\bar{M}_{2\pi}(\omega)$, given in eq.(4.41), for various scale levels. The signal power spectrum $C(\omega)$ is assumed to follow an exponential model.

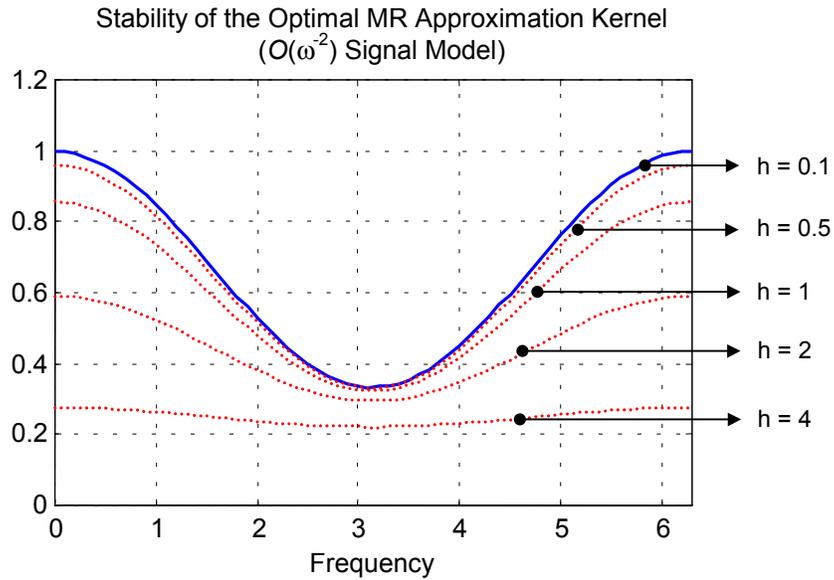


Figure 4.10 Plots of the 2π -periodic function $\overline{M}_{2\pi}(\omega)$, given in eq.(4.41), for various scale levels. The signal power spectrum $C(\omega)$ is assumed to follow the model given in eq.(4.24c).

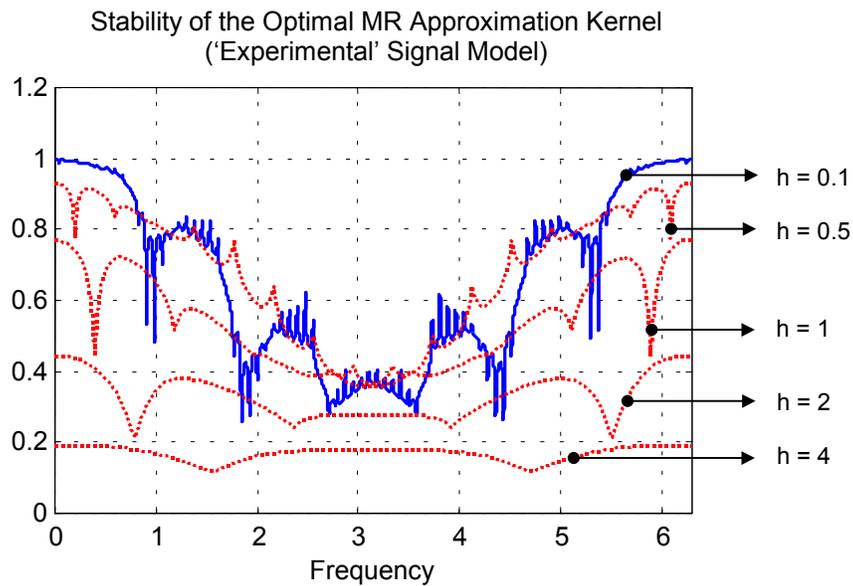


Figure 4.11 Plots of the 2π -periodic function $\overline{M}_{2\pi}(\omega)$, given in eq.(4.41), for various scale levels. The signal power spectrum $C(\omega)$ is assumed to follow the model given in eq.(4.24d).

As it is evidenced from the behaviour of these plots, the function $\overline{M}_{2\pi}(\omega)$ should converge to a certain strictly-positive bounded expression for increasing data density, that will determine the asymptotic stability behaviour of the optimal estimation algorithm in eq.(4.20). For the special case where the signal power spectrum $C(\omega)$ follows either a Gaussian or an exponential model, the asymptotic stability of the optimal Riesz basis becomes perfect ($A = B$), and the periodic function $\overline{M}_{2\pi}(\omega)$ will simply be reduced to a constant of value 1 (see Figures 4.8 and 4.9). This should be expected since in section 4.5 we had already seen that the optimal collocation kernel $\varphi(x, h_j)$ converges to the *sinc* interpolator in such cases, which of course creates an orthonormal set of base functions $\text{sinc}(x - n)_{n \in \mathbb{Z}}$. In general, however, the Riesz sampling basis that is constructed from the collocation kernel $\varphi(x, h_j)$ will not necessarily converge to an orthogonal system as seen, for example, in the cases of Figures 4.10 and 4.11.

It is interesting to observe that the stability of the optimal Riesz bases may not become worse as the data resolution level increases. This is obvious in the cases of the Gaussian and the exponential signal models, where the condition number of the linear approximation algorithm actually improves for high data density, approaching the ideal value of 1! A related important point, that requires a detailed mathematical analysis, is also the monotonic dependence of the stability of the optimal Riesz bases in the statistical collocation framework with respect to the data resolution level h_j . Such a conclusion,

however, cannot be supported from the limited numerical experiments that were performed in this section.

4.6.4 Final Remarks

The analysis performed in this chapter reveals an interesting viewpoint for the statistical collocation algorithm in eq.(4.20). Under certain conditions, the estimated (or referenced) field $\hat{g}(x)$ will always belong in a Hilbert subspace $V_j \subset L^2(\mathfrak{R})$ of a generalized MRA sequence, the scale level of which is dictated from the sampling resolution h_j of the discrete data. The collocation-based interpolation algorithm can be viewed as the application of a stable **sampling theorem** associated with this specific subspace, since the set of translates $\varphi(\frac{x}{h_j} - n, h_j)_{n \in \mathbb{Z}}$ of the optimal estimation kernel will constitute a Riesz sampling basis for V_j . This result closely adheres to similar mathematical studies, where it was shown that for (almost) every dyadic MRA in $L^2(\mathfrak{R})$ there exists a Riesz sampling basis in each of its nested subspaces (Walter, 1992). 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 estimation from discrete data; see Giacaglia and Lundquist (1972), Moritz (1976b, 1978b), Schmidt (1981), Sunkel (1981, 1984), Svensson (1983), Bjerhammar (1983) and Freeden (1983).

At this point, we should indicate the essential difference between the original MRA approximation concept according to Mallat (1989b) and the present collocation-based MR estimation scheme. The initial idea, as proposed by Mallat, was based on the orthogonal projection of the unknown signal $g(x) \in L^2(\mathfrak{R})$ onto a dyadic MRA subspace V_j . This approach requires access to the true continuous field $g(x)$, which is hardly known in most geodetic applications*. It combines a sequence of linear operations (i.e. prefiltering/analysis, sampling and postfiltering/synthesis) which are needed to determine the least-squares approximation of $g(x)$ within a multiresolution subspace $V_j \subset L^2(\mathfrak{R})$, and it is often used for efficient analog signal transmission, storage and compression in terms of a discretized representation at a certain scale level (Mallat, 1989a; Unser and Daubechies, 1997); see also section 2.4. Taking into account the basic properties that characterize a classic MRA structure, such an approximation scheme could be viewed as starting from the top of an inverted pyramid (i.e. MRA) and by successive orthogonal projections onto more and more detailed pyramid layers (i.e. MRA subspaces) we finally return to the top.

* Note that the usual *deterministic collocation* method in an arbitrary Hilbert space H also employs an orthogonal projective scheme to determine the optimal (smoothest) linear approximation $\hat{g}(x)$ of an unknown signal $g(x) \in H$. However, deterministic collocation does not require the full knowledge of the total field $g(x)$, but only the values of linear continuous functionals $L_n g$. The representers of these functionals then determine the linear subspace $V \subset H$ in which the orthogonal projective approximation takes place. Mallat's framework is essentially different from such a scheme, in that it does not use functional representers to define the solution subspace of the linear approximation problem. The Hilbert space H in Mallat's MRA method is identified as the infinite-resolution $L^2(\mathfrak{R})$ space, in which typical signal observables (such as discrete point values or derivatives) do not admit a bounded representation in terms of continuous linear functionals.

In geodetic operational problems, on the other hand, we usually start with a finite set of measurements from which we try to build a discretized representation of the underlying field based on some optimal estimation principles and certain information about its average spatial behaviour. The solution to this problem, according to the spatio-statistical collocation approach, can then be seen as starting from the bottom of a generalized MRA structure, and by obtaining denser and denser sampled values of the field (and correspondingly applying the sampling theorem associated with each subspace V_j of this generalized MRA) we finally reach the top. It can actually be shown that this type of **‘bottom-to-top’** multiresolution interpolatory scheme, through the use of a scaling cardinal kernel $\varphi(x, h_j)$, corresponds to an *oblique* projection of the original unknown signal $g(x) \in L^2(\mathfrak{R})$ onto a generalized MRA subspace V_j ; see Blu and Unser (1999a).

The derivation of the MR interpolation algorithm in section 4.3.2 was based only on a few simple principles (i.e. linearity, translation-invariance), as well as on the spatio-statistical MMSE optimal criterion. Properties such as stability and convergence, which motivated most of our discussions in Chapter 3, were not considered at all for arriving at the final result of eq.(4.20). Nevertheless, the optimal signal estimate $\hat{g}(x)$ according to this interpolation equation is both stable and convergent (in the L^2 norm) for increasing data resolution, as was explained in the previous sections. Hence, rather than trying to find an ‘optimal’ dyadic MRA that can overcome the stability, convergence and model-

versus-data resolution problems of deterministic collocation (as was suggested in Chapter 3), we followed a spatio-statistical collocation approach in this chapter that led to a similar, but more general, MRA-type approximation result. It is worth mentioning that similar attempts for refining the optimal interpolation procedures used in gravity field modelling have also been reported by the Swedish school of collocation (Svensson, 1983; Bjerhammar, 1983), which resulted in the so-called ‘inversion-free Bjerhammar predictors’; see also Bjerhammar (1987).

Concluding this chapter, we should briefly mention two major research topics which are natural extensions of the results presented herein. The first deals with the existence and construction of ‘generalized wavelet bases’ in the orthogonal complements $\{W_j\}$ of the generalized MRA sequence $\{V_j\}$ (i.e. $V_{j-1} \oplus W_{j-1} = V_j$) that is created by the optimal collocation kernel $\varphi(x, h_j)$. Such a spectral system would provide a powerful extension of the standard wavelet bases that are always associated with Mallat’s dyadic MRAs (Chapter 2). If such a step becomes successful, we can essentially link the interpolation algorithm of statistical collocation with ‘its own’ system of ‘non-stationary’ base functions. In this case, the estimated-referenced signal $\hat{g}(x)$ will always have a resolution-limited spectrum with respect to the corresponding wavelet-type spectral system (generalization of the Nyquist principle).

The second topic involves the application of a convolution operator to the optimally interpolated field $\hat{g}(x)$ of eq.(4.20), i.e.

$$\begin{aligned}\hat{f}(x) &= s(x) * \hat{g}(x) = \sum_n g(nh) \left[s(x) * \varphi\left(\frac{x}{h} - n, h\right) \right] \\ &= \sum_n g(nh) \varphi'\left(\frac{x}{h} - n, h\right)\end{aligned}\tag{4.43}$$

Such a case is of great importance in geodesy, since discrete gridded data (e.g. gravity anomalies, orthometric heights, topographic densities, etc.) are often used as input to many convolution algorithms for gravity field recovery. Given the optimal form of the interpolating kernel $\varphi(x, h)$ that was determined in section 4.3.2, it is easily realized that

$$\varphi'(x, h) \xleftrightarrow{\mathfrak{F}} \Phi'(\omega, h) = \frac{C\left(\frac{\omega}{h}\right) S\left(\frac{\omega}{h}\right)}{\sum_k C\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right)}\tag{4.44}$$

where $S(\omega)$ is the Fourier transform of the convolution kernel $s(x)$, and $C(\omega)$ corresponds to the power spectrum of the input field $g(x)$. The interest is now to study if (and under which conditions on the convolution operator s) the new synthetic scaling kernel $\varphi'(x, h)$ also creates its own generalized MRA sequence $\{V'_j\}$ that could be associated with the output estimated field $\hat{f}(x)$ of eq.(4.43). Both of the aforementioned

topics are very important and interesting, and they require extensive mathematical analysis that is not possible to be presented here.

Chapter 5

ALIASING ERROR AND NOISE FILTERING IN LINEAR MULTIRESOLUTION APPROXIMATION MODELS

Error analysis is one of the most critical issues in every signal estimation method. Meaningful and easy to compute error measures are very important for evaluating the quality of the final signal approximation from a given data set, for assessing data requirements (e.g. sampling resolution level) based on pre-selected signal error limits, as well as for comparing different estimation models/kernels for a given class of unknown signals. As it was mentioned earlier in section 3.1.1, the deterministic collocation methodology does not generally provide the means for a practically useful and straightforward error analysis, especially with respect to the data resolution parameter. The linear MR approximation model that was studied in the last chapter, on the other hand, is much better suited for a resolution-dependent analysis of the signal interpolation error. Furthermore, the topology (L^2 norm) that is associated with such MRA-type approximation schemes can easily produce a spatial average estimate of the overall signal error (RMS error), which is commonly used in many applications. In the following

sections, we will study in detail various types of signal error estimates that can be obtained when using a general linear MR approximation model with regularly gridded data. The main focus will be on the development of a simple algorithm for computing the mean L^2 estimation error as a function of the data resolution level, as well as on the study of aliasing propagation in convolution-type integral formulas. In addition to such deterministic error analysis that reflects only the effect of the finite data resolution, a linear noise filtering methodology is also presented for dealing with more realistic signal estimation problems where the discrete input data are influenced by additive (in general non-stationary) random noise.

5.1 Accuracy of Linear Multiresolution Approximation Models

In this section, we will present a certain spatio-statistical methodology for studying the behaviour of the signal error in linear multiresolution approximation models. The general estimation equation has the usual convolution-type form

$$\hat{g}(x) = \sum_n g(nh) \varphi\left(\frac{x}{h} - n\right) \quad (5.1)$$

where $\varphi(x)$ is an appropriate known kernel and h is the data resolution level. The justification for the use of the wavelet-like interpolating model of eq.(5.1), as well as criteria for an optimal determination of its basic scaling kernel, have already been

discussed in the previous chapters. Here, we concentrate our attention explicitly on the behaviour of the approximation error

$$e(x) = g(x) - \sum_n g(nh) \varphi\left(\frac{x}{h} - n\right) \quad (5.2)$$

where the used kernel $\varphi(x)$ does not necessarily have the optimal form that was derived in the last chapter. The term $e(x)$, as defined in the above equation, represents the *aliasing error* for the unknown signal at the given resolution. Of special importance is the development of a simple algorithm that computes the decay rate of some functional of this error with respect to the data resolution level h . Such an algorithm will be constructed and tested, using various kernels and simulated signals, in the following sections. It is always assumed that the signal is sampled over its entire (finite or infinite) support. Note also that no special properties/restrictions have been assigned to the unknown field (e.g. bandlimitedness, smoothness, etc.) besides the fact that it is a finite-energy signal with a properly defined Fourier transform, and that its sampling always results in a square-summable data sequence.

The only restrictions that we need to put on the kernel $\varphi(x)$ are those guaranteeing that the estimation model of eq.(5.1) provides: (i) an unambiguous (unique) signal description for any set of measurable data values $\{g(nh)\} \in l^2(Z)$, and (ii) a stable numerical algorithm. Since we would like to use the (always useful) Fourier transform formalism, it

is also important that the closed linear span of $\varphi(\frac{x}{h} - n)_{n \in \mathbb{Z}}$ be a well defined subspace of the Hilbert space $L^2(\mathfrak{R})$. In terms of error analysis, there is no particular reason to require that the approximation kernel should correspond to a scaling function of some MRA model in $L^2(\mathfrak{R})$. The three above properties are satisfied (for any finite data resolution level) if and only if the following condition is met by the Fourier transform of the kernel $\varphi(x)$:

$$0 < A \leq \sum_n |\Phi(\omega + 2\pi n)|^2 \leq B < \infty \quad (5.3)$$

For more details, see Aldroubi and Unser (1994) and Unser and Daubechies (1997). Note that the above Riesz condition is not restrictive at all and it is satisfied by virtually any approximation kernel used in practice. It is also necessarily satisfied by all scaling functions encountered in wavelet approximation theory, as well as by the optimal collocation kernel that was derived in the previous chapter under certain constraints on the spatial CV function and the power spectrum of the unknown signal.

5.1.1 Multi-Parameter Error Description – Error CV Functions

In deterministic signal estimation from discrete data it is not possible to obtain an *exact* expression for the actual error (or even the square error) as a function of the spatial point

position x . This is because such an expression requires a-priori knowledge of the unknown field $g(x)$, i.e.

$$e(x) = g(x) - \hat{g}(x) \quad (5.4)$$

Similarly, a spectral analysis for the pointwise estimation error also requires complete knowledge of the total unknown field. For practical applications, we need to develop alternative expressions/measures for describing the behaviour of the signal error, whose evaluation should be based on more accessible characteristics of the unknown field (e.g. its spatial CV function or its power spectrum).

In order to do that, we can initially express the signal approximation error as a function of three distinct spatial parameters, as follows:

$$e(x, x_o, h) = g(x) - \sum_n g(nh - x_o) \varphi\left(\frac{x + x_o}{h} - n\right) \quad (5.5)$$

The parameter x denotes the spatial point location where the error is evaluated, whereas the two additional parameters (x_o and h) correspond to the sampling phase and the data sampling resolution, respectively. The last two quantities are not completely independent and they always satisfy the relation $-h/2 \leq x_o < h/2$. The above error formula is valid in accordance with a translation-invariance condition for the multiresolution estimation algorithm. This multi-dimensional error description is illustrated in Figure 5.1. Note that,

even if we average the pointwise error $e(x, x_o, h)$ over the sampling phase parameter x_o , we would still need to know the complete field $g(x)$ in order to perform an accuracy evaluation at a certain resolution level.

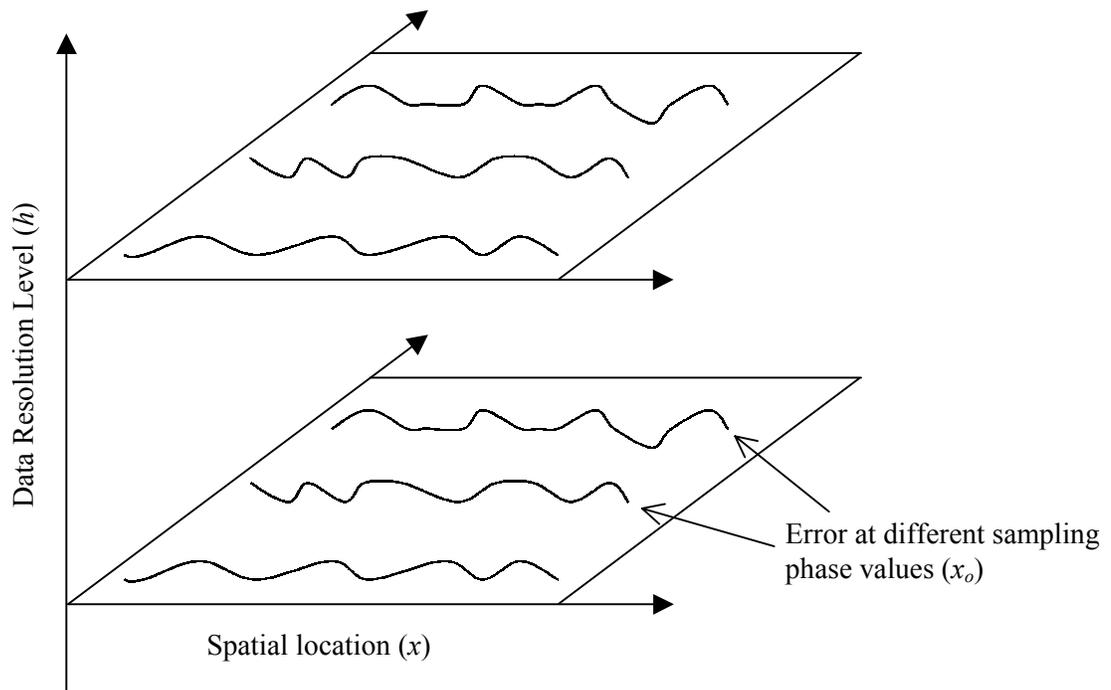


Figure 5.1 Multi-parameter description of the signal approximation error in linear translation-invariant MR models using 1D gridded data

Using eq.(5.5), we can define a **spatial error CV function** at a certain data resolution level (h) and sampling phase value (x_o). Such a CV function has the usual ‘stationary’ form

$$c_e(\xi, x_o, h) = \int e(x, x_o, h) e(x + \xi, x_o, h) dx \quad (5.6)$$

and its value at the origin ('error variance') corresponds to the square L^2 error norm at a specific x_o and h , i.e.

$$c_e(\xi, x_o, h)_{\xi=0} = \int e^2(x, x_o, h) dx = \|e(x, x_o, h)\|_{L^2}^2 \quad (5.7)$$

The study of the above error CV function, or equivalently the study of its Fourier transform (i.e. error power spectrum), can reveal valuable information about the average spatial behaviour of the pointwise error at specific h and x_o values. Unfortunately, such an error CV function cannot be computed without complete knowledge of the unknown field itself. In order to overcome this limitation, we can now define a **mean error CV function** over all possible sampling phase values. We will have

$$c_e^{aver}(\xi, h) = \frac{1}{h} \int_{-h/2}^{h/2} c_e(\xi, x_o, h) dx_o \quad (5.8)$$

It can be easily shown that the Fourier transform of the mean error CV function, considered as a function of ξ only, has the following integral form:

$$P_e(\omega, h) = \frac{1}{h} \int_{-h/2}^{h/2} |E(\omega, x_o, h)|^2 dx_o \quad (5.9)$$

where $E(\omega, x_o, h)$ is the Fourier transform of the pointwise error term in eq.(5.5) with respect to the spatial parameter x . In other words, the mean error CV function is just the inverse Fourier transform of the mean error power spectrum, where the ‘mean’ in both domains is meant in a spatio-statistical sense over all possible sampling phase values.

By analytically computing the Fourier transform of the pointwise error in eq.(5.5), we can finally obtain the following algebraic expression for the **mean error power spectrum at a certain resolution level**:

$$P_e(\omega, h) = C(\omega) - C(\omega)\Phi(h\omega) - C(\omega)\Phi^*(h\omega) + h \overline{C}_h(\omega) |\Phi(h\omega)|^2 \quad (5.10)$$

where $C(\omega)$ is the signal power spectrum, $\overline{C}_h(\omega)$ denotes its periodization according to eq.(4.16), and $\Phi(\omega)$ corresponds to the Fourier transform of the approximation kernel. The proof of the above equation is based on similar derivations as the ones given in Appendix A. This last formula defines an error measure which does not require complete knowledge of the unknown field, but only knowledge of its power spectrum (or its spatial CV function). Actually, the term $P_e(\omega, h)$ corresponds to the exact same spatio-statistical quantity that was used in the optimization procedure of Chapter 4, which resulted in the collocation filter according to eq.(4.19). If we substitute this optimal collocation filter in eq.(5.10), we get the following simple expression for the optimized mean error power spectrum:

$$P_e^{opt.}(\omega, h) = C(\omega) \left(1 - \frac{C(\omega)}{\sum_k C(\omega + \frac{2\pi k}{h})} \right) \quad (5.11)$$

Note that the error formula in eq.(5.10) is also valid for cases where the kernel of the linear MR approximation model does not have a fixed form, but it may depend on the data resolution itself, i.e. $\varphi(x) = \varphi(x, h)$ or $\Phi(\omega) = \Phi(\omega, h)$. This is actually the case with the **optimal collocation kernel**, as was explained in the last chapter. For simplicity, in all the following error formulas we will omit the possible dependence of the approximation kernel/filter on the data resolution level h .

Simulations were conducted in order to provide some examples for the behaviour of $P_e(\omega, h)$, and the results are shown in Figures 5.2 through 5.7. Four different interpolating kernels were used along with a few synthetic signal models. In particular, we used three of the signal models $C(\omega)$ that were introduced in section 4.5, namely Gaussian, $O(\omega^{-2})$ and ‘experimental’. The first three tested estimation filters $\Phi(\omega)$ correspond to fixed interpolating kernels (*sinc* function, linear interpolating spline, cubic interpolating spline), whereas the fourth estimation filter is resolution-dependent and it corresponds to the optimal collocation kernel for every signal model; see eq.(4.19). With the exception of the Gaussian signal model (where we already know that the optimal collocation kernel converges to the *sinc* interpolator for high data resolution), the use of a

perfect low-pass estimation filter $\Phi(\omega)$ shows by far the worst performance among all kernels. Even for Gaussian signals, the *sinc* kernel performs very poorly for relatively low data resolution, see Figure 5.2. It is also interesting to observe that, in almost all cases, the asymptotic frequency decay of $P_e(\omega, h)$ follows exactly the same pattern for every approximation kernel.

A final brief note should be made regarding the *consistency* of the linear estimation formula in eq.(5.1), as the data density increases. Using eq.(5.10), it can be shown that a necessary and sufficient condition in order for the mean error power spectrum $P_e(\omega, h)$ to vanish (as $h \rightarrow 0$) is

$$|\Phi(\omega)|_{\omega=0} = 1 \quad (5.12a)$$

In addition to the Riesz condition given in eq.(5.3), the last equation imposes a very mild extra restriction for the admissibility of the scaling kernels $\varphi(x)$ that should be used with such MR interpolating models. Note that for cases where the estimation filter $\Phi(\omega)$ does not have a fixed form, but it varies depending on the data resolution level, the condition of eq.(5.12a) takes the limiting form

$$\lim_{h \rightarrow 0} |\Phi(\omega, h)|_{\omega=0} = 1 \quad (5.12b)$$

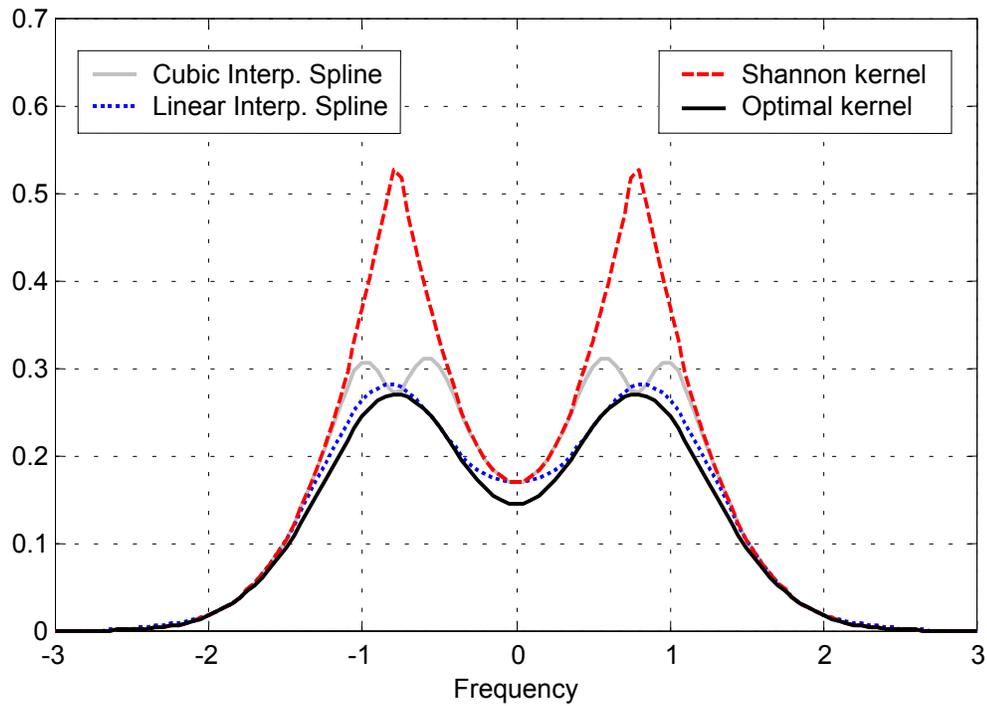


Figure 5.2 Mean error power spectrum for Gaussian signal model using various interpolating kernels (the data resolution level is $h=4$)

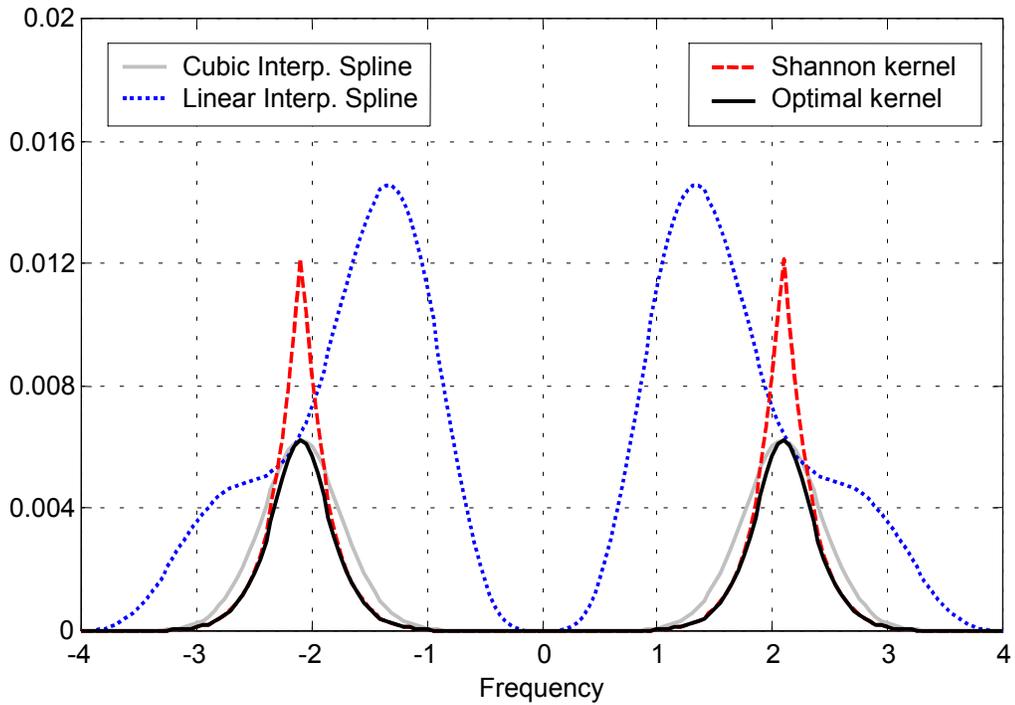


Figure 5.3 Mean error power spectrum for Gaussian signal model using various interpolating kernels (the data resolution level is $h=1.5$)

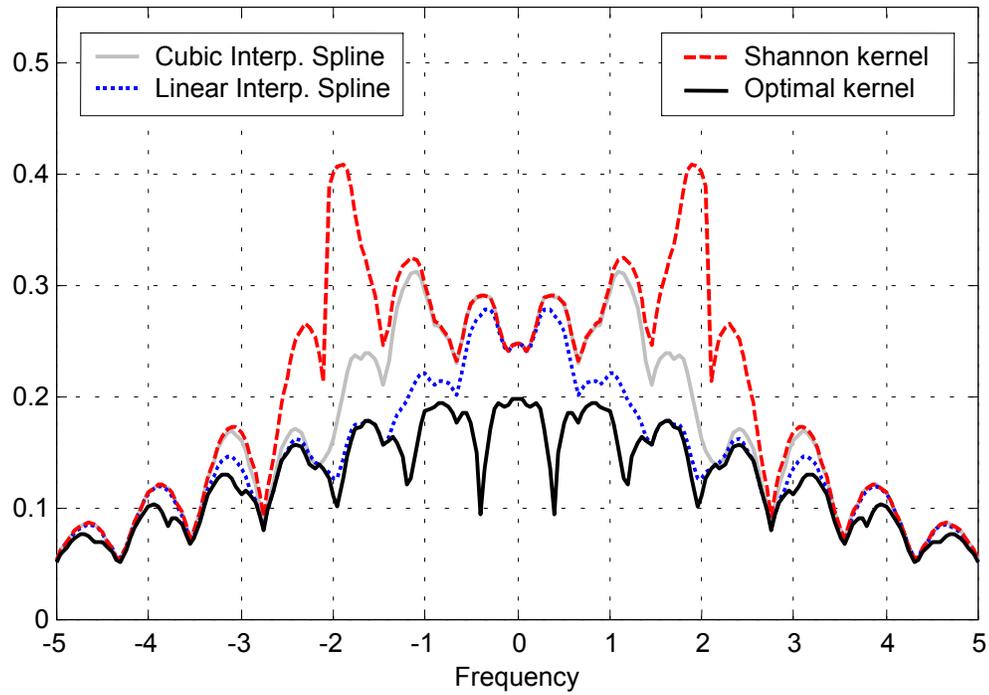


Figure 5.4 Mean error power spectrum for the ‘experimental’ signal model using various interpolating kernels (the data resolution level is $h=1.5$)

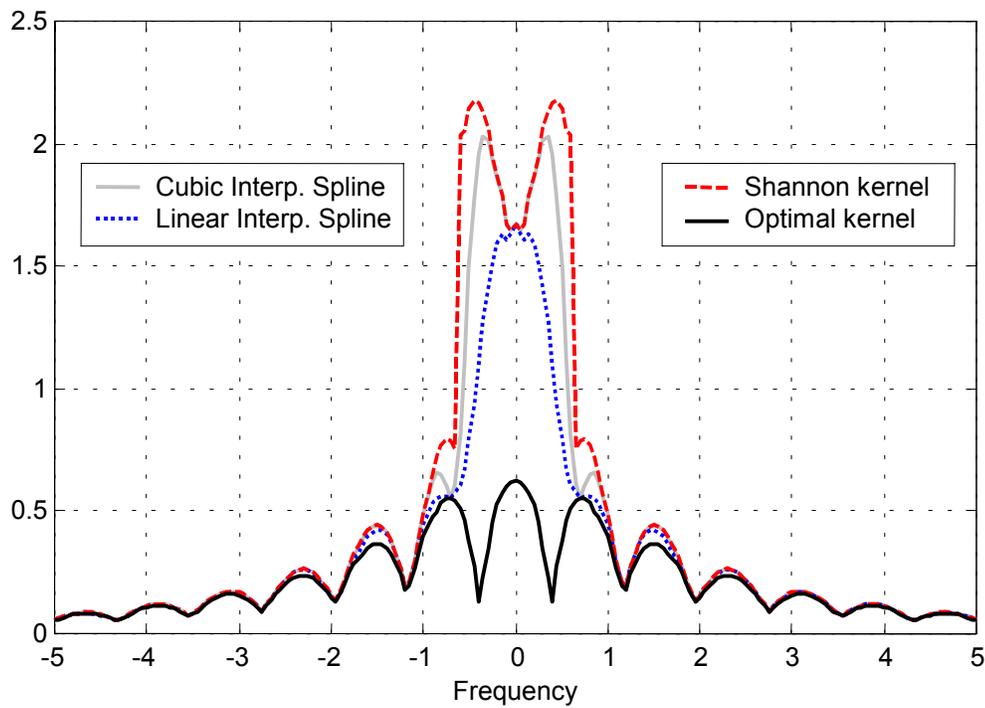


Figure 5.5 Mean error power spectrum for the ‘experimental’ signal model using various interpolating kernels (the data resolution level is $h=10$)

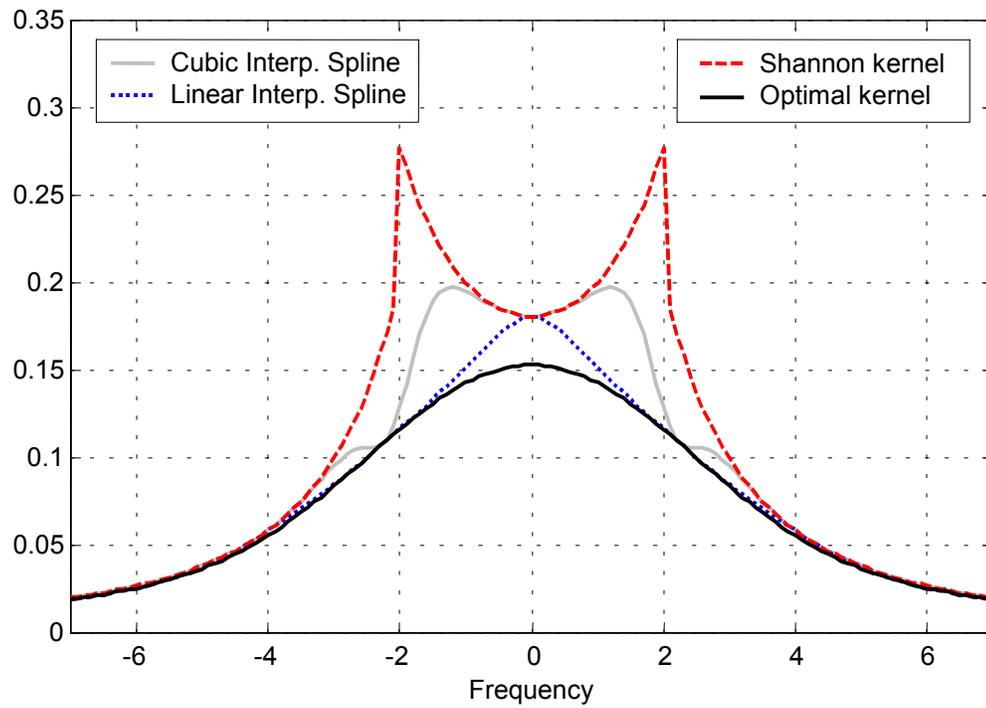


Figure 5.6 Mean error power spectrum for the $O(\omega^{-2})$ signal model using various interpolating kernels (the data resolution level is $h=1.5$)

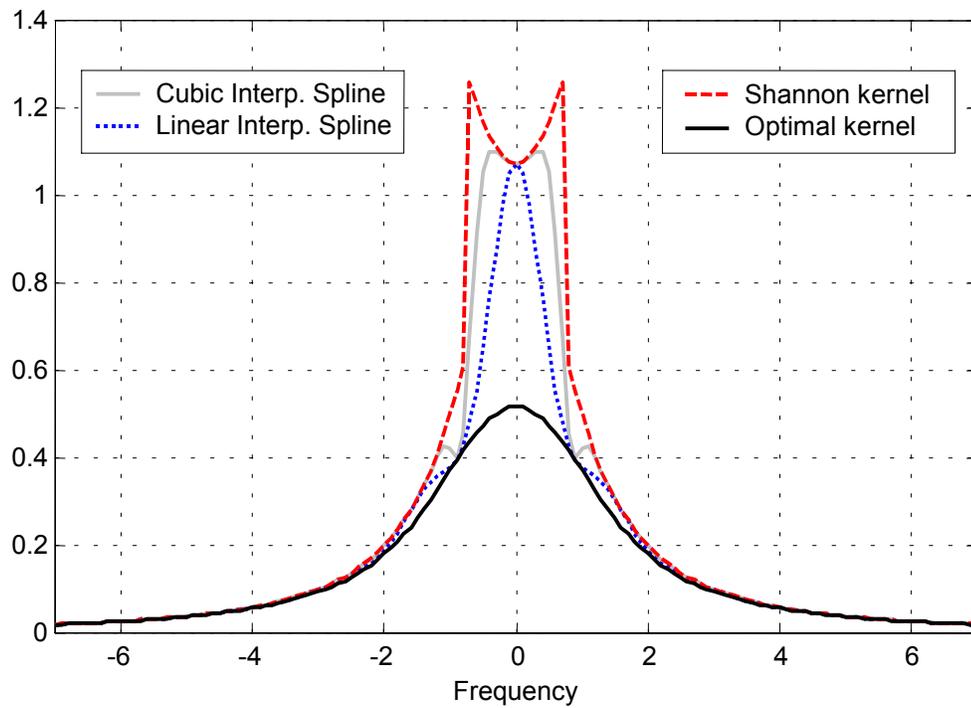


Figure 5.7 Mean error power spectrum for the $O(\omega^{-2})$ signal model using various interpolating kernels (the data resolution level is $h=4$)

which is of course satisfied by the optimal collocation filter that was determined in Chapter 4. It is worth mentioning that the admissible kernels do not need to be strictly interpolating in order to have a consistent MR estimation scheme (i.e. the data values $g(nh)$ do not need to be reproduced exactly by the signal approximation $\hat{g}(x)$). This offers great flexibility in cases where we want to work with smooth orthonormal approximation bases $\varphi(x-n)_{n \in \mathbb{Z}}$ having compact support, since the only interpolating orthonormal kernel $\varphi(x)$ with compact support is the discontinuous Haar kernel; for more details, see Xia and Zhang (1993).

5.1.2 Decay Rate of the Mean L^2 Approximation Error

As the data sampling step gets smaller, the error of eq.(5.1) decreases and eventually becomes negligible as h goes to zero. The rate of decrease of the approximation error as a function of h is very crucial in many signal processing applications (Unser and Daubechies, 1997; Blu and Unser, 1999a), as well as in classic approximation theory (Butzer et al., 1988; De Boor et al., 1994). Such information would permit, for example, an objective comparison between different estimation kernels for a given class of unknown signals. It is also very useful for identifying critical data resolution levels that can provide an overall mean square estimation error below certain threshold values. Therefore, it becomes very important to obtain a general expression that describes the

decaying pattern of the signal error of eq.(5.1) as a function of the data resolution h and the used kernel $\varphi(x)$.

In the previous section we mentioned that a pointwise approach for studying the behaviour of the approximation error requires the use of a-priori known (synthetic) fields. Here, on the other hand, we want to develop a resolution-dependent error description based on more easily modelled characteristics of the unknown field, such as the spatial signal CV function or the signal power spectrum. As a result, we should dismiss the concept of pointwise error description and replace it with a suitable error norm that characterizes the overall performance of the estimation algorithm at every data resolution level h . A convenient choice for such an error measure is the following quantity:

$$\sigma^2(h) = \frac{1}{h} \int_{-h/2}^{h/2} \|e(x, x_o, h)\|_{L^2}^2 dx_o \quad (5.13)$$

which corresponds to the **mean square L^2 error norm**, averaged over all sampling phase values at a specific data resolution. Such a spatio-statistical global error measure has also been considered in a similar recent study by Blu and Unser (1999a,b). Note that the value of $\sigma^2(h)$ is exactly equal to the ‘mean error variance’ derived from the mean error CV function in eq.(5.8), i.e.

$$\sigma^2(h) = c_e^{aver}(\xi, h)_{\xi=0} \quad (5.14)$$

A simple algorithm for the numerical computation of the mean error variance, at different resolution levels, can be easily constructed in the frequency domain. Using eq.(5.14) and one of the basic properties of the Fourier transform, we can express the mean error variance in terms of the following frequency-domain integral:

$$\sigma^2(h) = \frac{1}{2\pi} \int P_e(\omega, h) d\omega \quad (5.15)$$

where $P_e(\omega, h)$ is the mean error power spectrum which is given by the general formula in eq.(5.10). If we assume that the approximation kernel $\varphi(x)$ corresponds to a *symmetric* function (something that is often employed in signal analysis), then the above integral can be reduced to the simple form

$$\sigma^2(h) = \frac{1}{2\pi} \int C(\omega) K(h\omega) d\omega \quad (5.16)$$

where $C(\omega)$ is the power spectrum of the unknown signal under consideration. The key kernel $K(\omega)$ depends solely on the approximation filter and it is defined as follows:

$$K(\omega) = 1 - 2|\Phi(\omega)| + \sum_k |\Phi(\omega + 2\pi k)|^2 \quad (5.17)$$

where $\Phi(\omega)$ is the Fourier transform of the kernel $\varphi(x)$. The proof of the above algorithm can be found in Appendix C. If the approximation kernel is not a symmetric

function (e.g. Haar kernel), then the numerical value computed from the integral in eq.(5.16) will correspond to the maximum lower bound (infimum) for the mean error variance. Also, in the special case where the integer translates of $\varphi(x)$ constitute an orthonormal set under the L^2 inner product (e.g. orthonormal scaling functions encountered in wavelet theory), the auxiliary kernel $K(\omega)$ is reduced to the very simple form

$$K(\omega) = 2 \left(1 - |\Phi(\omega)| \right) \quad (5.18)$$

Note that the symbol $|\Phi(\omega)|$ denotes the magnitude of the Fourier transform, and not the absolute value of $\Phi(\omega)$. When the space-domain kernel is symmetric, we obviously have $\Phi(\omega) = |\Phi(\omega)|$.

Using the previous error algorithm, an interesting result can be obtained for the special case of *band-limited signal interpolation*. In such a case, the kernel $\varphi(x)$ corresponds to the usual *sinc* function and the associated approximation filter $\Phi(\omega)$ becomes a perfect low-pass filter over the frequency band $[-\pi, \pi]$. The application of the integral error formula in eq.(5.16) will then yield the result

$$\sigma^2(h) = \frac{1}{\pi} \int_{|\omega| > \pi/h} C(\omega) d\omega \quad (5.19)$$

which corresponds to twice the signal energy contained outside the Nyquist bandwidth. This is actually a reasonable result since applying a perfect low-pass filter to a discrete data set (in order to get a continuous field approximation) not only cuts off all signal information outside the Nyquist bandwidth, but it also keeps all the distorted (aliased) signal frequencies *within* the recovered bandwidth completely unfiltered; see Figure 5.8. It is quite remarkable that our error modelling methodology shows that the effect of this ‘spectrum folding’ to the mean square estimation error is exactly equal to the lost signal energy that lies outside the Nyquist bandwidth! If the original unknown field is already band-limited and the sampling resolution level h is below its Nyquist limit, then the mean error variance $\sigma^2(h)$ naturally becomes zero.

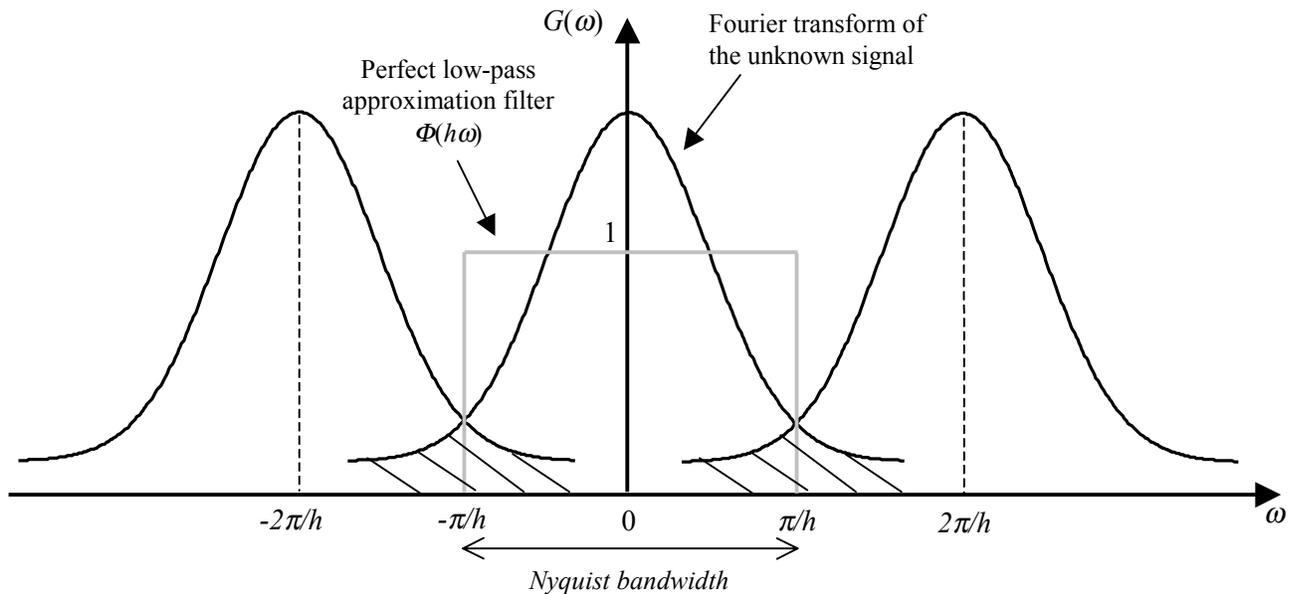


Figure 5.8 Filtering scheme of the linear MR estimation formula in eq.(5.1) for the case of band-limited signal interpolation

The procedure described in this section provides an original frequency-domain approach for studying the behaviour of the mean L^2 approximation error in linear MR models as a function of the data sampling level and the estimation kernel. Its implementation is straightforward and it is based on a simple integration of the signal power spectrum against an auxiliary kernel, which: (i) depends exclusively on the used estimation filter, and (ii) is scaled according to the given data resolution level. The computational algorithm becomes even simpler in the case where an orthonormal estimation filter is employed, which makes the use of orthonormal scaling kernels $\varphi(x)$ in the MR approximation framework of eq.(5.1) particularly attractive from an error analysis point of view.

The error formula in eq.(5.16) should be considered a refinement of a certain frequency-domain methodology that has often been applied for aliasing studies in gravity field signals (Forsberg, 1986; Kotsakis and Sideris, 1998, 1999). According to this methodology, the mean square aliasing error in a signal that is sampled at resolution level h and has a power spectrum $C(\omega)$ is estimated by the simple integral formula*

$$\hat{\sigma}^2(h) = \frac{1}{2\pi} \int_{|\omega| > \pi/h} C(\omega) d\omega \quad (5.20)$$

* In practice, a 2D generalization of this formula is used, as well as generalizations for signals defined on strictly compact domains and having purely discrete Fourier power spectra (Forsberg, 1986).

whose result corresponds to the signal energy contained outside the Nyquist bandwidth. However, the above error estimate is valid only for cases where a perfect low-pass **prefiltering** (over the Nyquist bandwidth) has been applied to the continuous signal, before the sampling procedure takes place. In gravity field approximation problems, on the other hand, it is usually impossible to apply such a prefiltering simply because we cannot access the original continuous signals but only their discrete values at a certain resolution level. Furthermore, it should be kept in mind that the aliasing error is a ‘relative’ concept, in the sense that it depends significantly on the chosen reconstruction kernel $\varphi(x)$, which may not necessarily be the *sinc* function as implied in the special case of eq.(5.20). Therefore, any methodology applied for signal error analysis should always incorporate the associated estimation kernel/filter, as in eq.(5.16).

5.1.3 Numerical Examples with Synthetic Signals

In order to test the behaviour of the mean error variance $\sigma^2(h)$ according to the frequency-domain formula (5.16), several numerical examples are presented herein. We will use three different models for the signal power spectrum $C(\omega)$, namely Gaussian, ‘experimental’ and $O(\omega^{-2})$ -type models; see section 4.5 for their analytical forms. The tested estimation filters $\Phi(\omega)$ for the computation of the mean error variance in every signal model correspond to the following kernels: (i) Shannon (*sinc*) kernel, (ii) Haar kernel, (iii) linear orthonormal B-spline (order 1), (iv) cubic orthonormal B-spline (order

3), and (v) optimal collocation kernel according to eq.(4.19). The first four scaling kernels $\varphi(x)$ are independent of the actual data resolution and they have a fixed waveform. Their Fourier transforms are shown in Figure 5.9. Analytical expressions for the two orthonormal spline kernels (as well as for their non-orthogonal and interpolating versions of any order) can be found in Mallat (1998b, p. 227) and Unser (1999). Note that the Haar function (linear orthonormal B-spline of zero order) is the only non-symmetric estimation kernel used, and thus the corresponding filter shown in Figure 5.9 illustrates only the magnitude of its Fourier transform. The optimal collocation kernel for every signal model is of course resolution-dependent, i.e. $\varphi(x) = \varphi(x, h)$, and its behaviour for various data sampling levels was demonstrated earlier in Chapter 4.

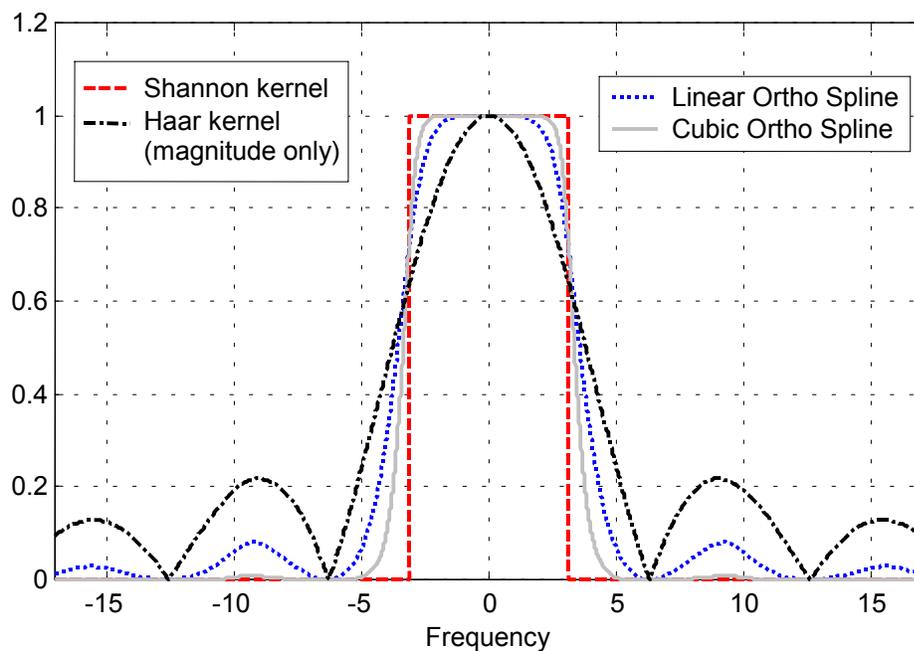


Figure 5.9 Tested MR approximation filters $\Phi(\omega)$ corresponding to different orthonormal scaling kernels $\varphi(x)$

The average performance of all five approximation filters $\Phi(\omega)$ for the various signal models, over a certain range of data resolution values, is shown in the Figures 5.10 through 5.12. Note that the computation of the mean error variance $\sigma^2(h)$ for the Haar kernel case was based on the integration of the mean error power spectrum $P_e(\omega, h)$ according to eq.(5.15), and not on the simplified integral formula of eq.(5.16) which is valid only for symmetric kernels.

The error curves shown in Figure 5.10 correspond to the case of a Gaussian signal model. From this figure we can easily verify the asymptotic convergence of the optimal collocation kernel to the *sinc* interpolator, for high data resolution (see also section 4.5). However, as the data sampling rate $1/h$ decreases, the performance of band-limited signal interpolation based on the use of the Shannon kernel worsens significantly, and the corresponding values of the mean error variance become even higher than the cases of spline-based signal approximation using the orthonormal B-spline kernels of degree 1 and 3. Note that the Shannon kernel is basically identical to a B-spline function of infinite order, as was recently shown by Aldroubi and Unser (1994).

A similar error trend is observed for the other two signal models (Figures 5.11 and 5.12), where the band-limited signal interpolation shows consistently larger error variances than the ‘shorter kernel’ approximation schemes based on the use of either a linear or a cubic orthonormal spline.

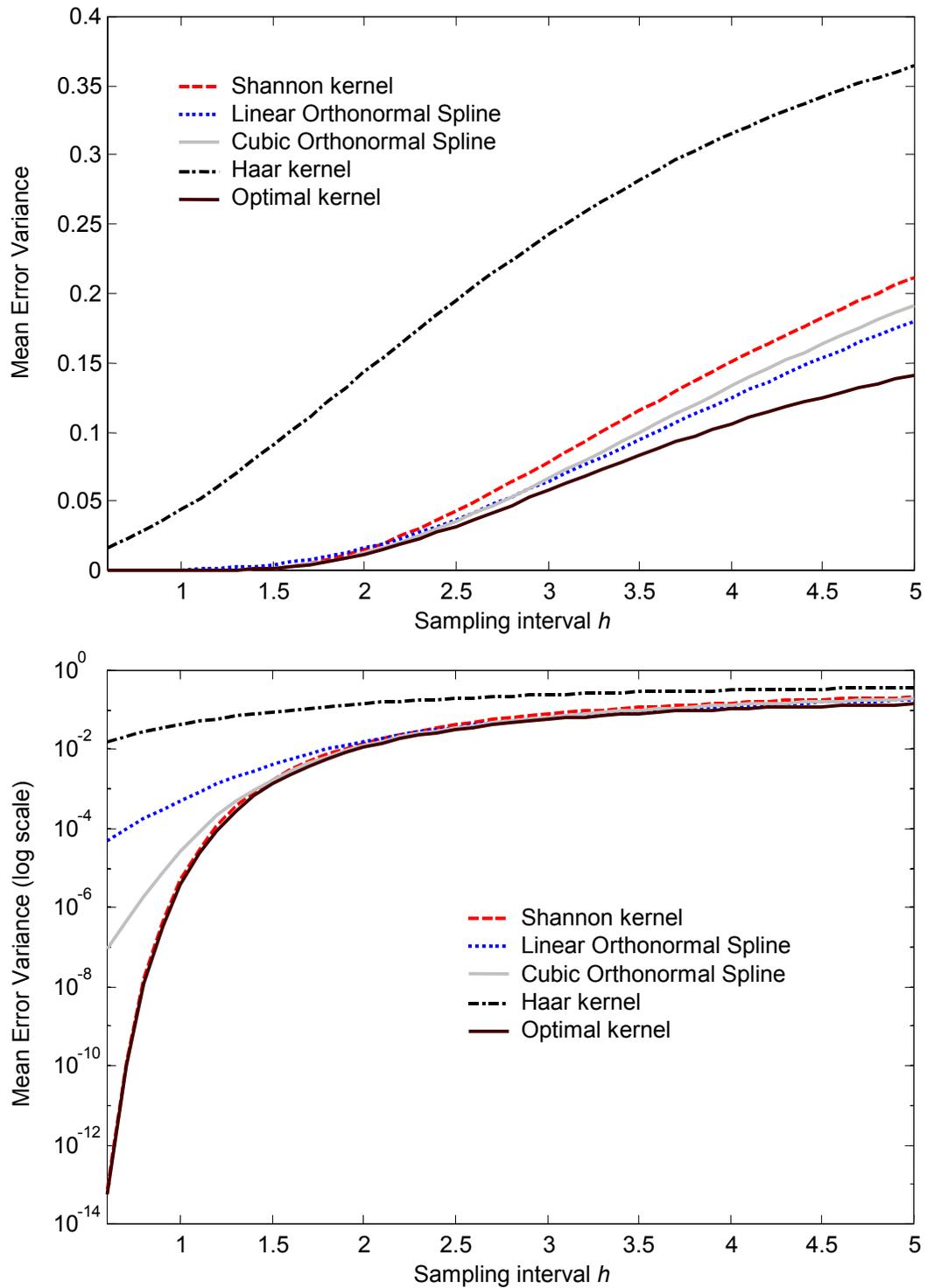


Figure 5.10 Decay rate of the mean error variance using various MR estimation kernels. The unknown signal follows a Gaussian power spectrum model.

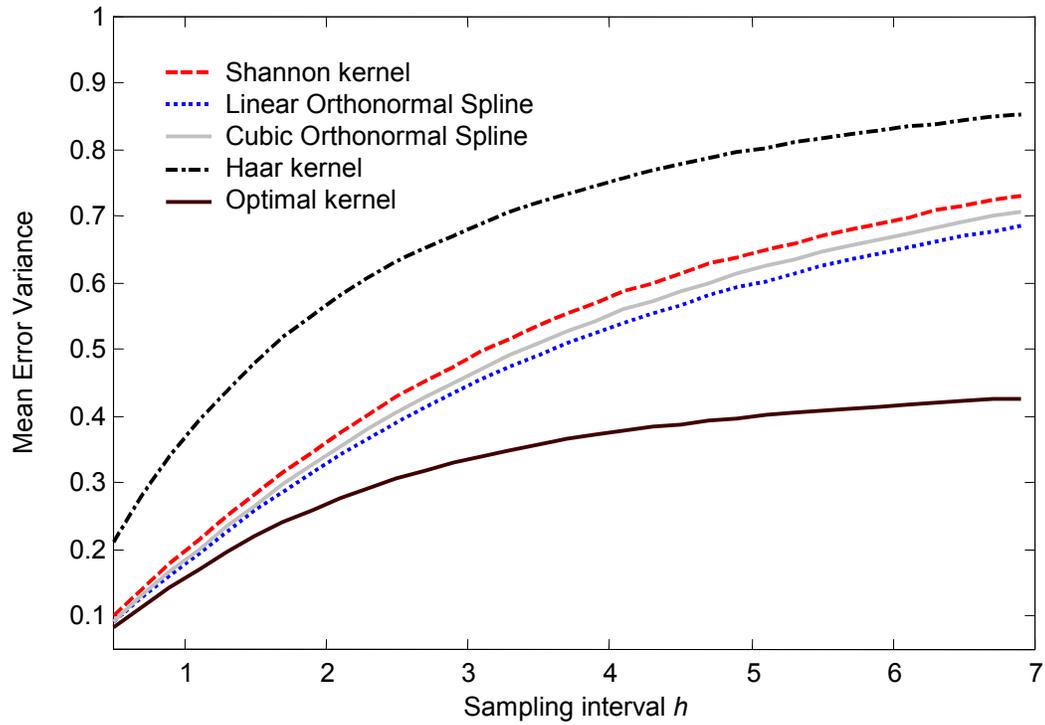


Figure 5.11 Decay rate of the mean error variance using various MR estimation kernels.

The unknown signal follows an $O(\omega^{-2})$ -type power spectrum model.

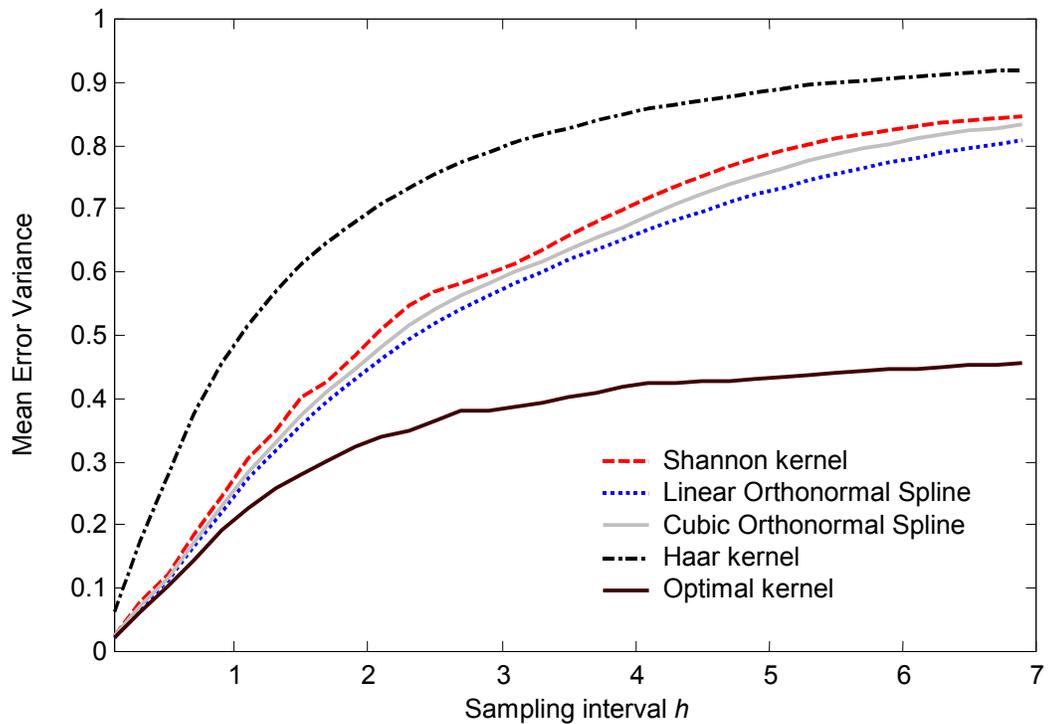


Figure 5.12 Decay rate of the mean error variance using various MR estimation kernels.

The unknown signal follows an 'experimental' power spectrum model.

The Haar scaling function shows the worst performance among all estimation kernels for every signal model, which should be expected due to its discontinuous (step-like) interpolating behaviour. Finally, it is interesting to point out the superiority of the collocation-based optimal kernel, especially for the most realistic case of the experimental signal model $C(\omega)$ (see Figure 5.12), which reveals the potential hidden behind the use of resolution-dependent scaling kernels $\varphi(x, h)$ in linear MR approximation models.

In practice, the ability of the error formula in eq.(5.16) to reliably measure the aliasing effect on the signal approximation $\hat{g}(x)$ depends on how well the function $C(\omega)$ resembles the characteristics of the true signal power spectrum $|G(\omega)|^2$. Many gravity field studies have demonstrated that this kind of uncertainty can significantly affect the accuracy evaluation of optimal linear estimation methods, such as traditional collocation (Moritz, 1980). The problem of properly estimating the power spectrum (or the spatial CV function) of the unknown field is embedded in every error modelling procedure associated with any type of signal approximation method used in geodesy. Although these important empirical estimation problems are not treated in this thesis, it is reasonable to claim that the mean error variance, as given by eq.(5.16), will not be very sensitive with respect to the choice of the signal power spectrum model. This is because $\sigma^2(h)$ is basically the outcome of a smoothing operation applied to the adopted signal model $C(\omega)$, which results from its integration against the auxiliary kernel $K(h\omega)$.

5.1.4 Aliasing Error Propagation in Convolution-Type Integral Formulas

The error formula in eq.(5.16) can be modified for the purpose of evaluating the propagated mean error variance in convolution-type integral formulas, using regularly gridded input data with uniform resolution. This is of special importance in physical geodesy applications where discrete gravity/height data are used as input in many different convolution algorithms for gravity field recovery, including upward and downward continuation, terrain correction computation, indirect effect computation, and of course gravimetric geoid determination based on Stokes' integral. In terms of error analysis, the main interest in such cases is to measure the output signal aliasing error for a given gravity/height grid resolution, as well as to infer the required data grid spacing for a pre-selected mean error level in the estimated output signal.

Such situations can be described mathematically by a general operatorial form

$$g_2(x) = s(x) * g_1(x) \quad (5.21)$$

where $s(x)$ is a given convolution kernel that connects the two continuous fields $g_1(x)$ and $g_2(x)$. Although we will only examine the simple one-dimensional case here, certain generalizations of the following results can easily be made to cover multi-dimensional problems. In practice, the evaluation of eq.(5.21) is based on gridded data values for the input signal, which should be 'referenced' according to the general MR approximation model of eq.(5.1). In this way, we have

$$\hat{g}_2(x) = s(x) * \hat{g}_1(x) = \sum_n g_1(nh) \left[s(x) * \varphi\left(\frac{x}{h} - n\right) \right] \quad (5.22)$$

where $\varphi(x)$ denotes the chosen estimation (reference) kernel, which may not necessarily be the optimal collocation-based interpolation kernel that was determined in Chapter 4.

The situation is illustrated in Figure 5.13.

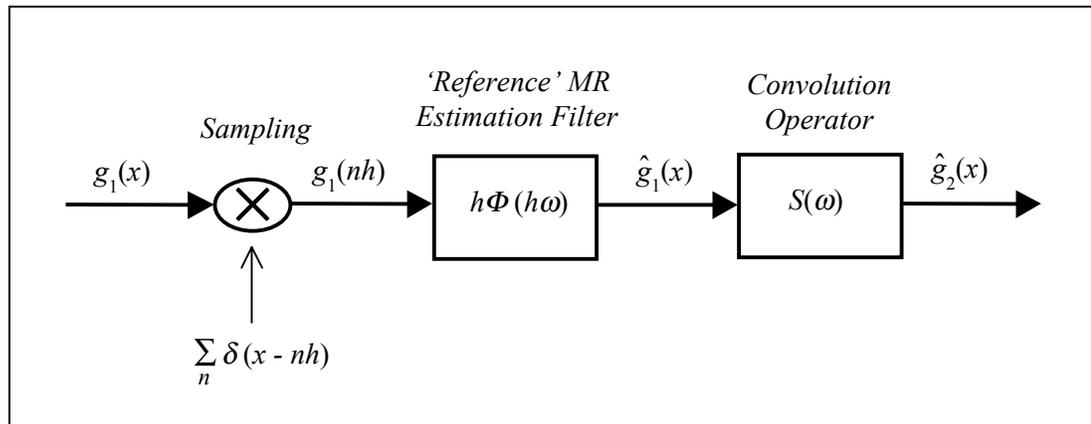


Figure 5.13 Multiresolution filtering configuration of convolution-based integral formulas using discrete input data

The pointwise approximation errors for the referenced input signal and the output signal, at a certain sampling phase and data resolution level, will accordingly be related by the convolution formula

$$e_2(x, x_0, h) = s(x) * e_1(x, x_0, h) \quad (5.23)$$

and the corresponding error power spectra are

$$\left| E_2(\omega, x_o, h) \right|^2 = |S(\omega)|^2 \left| E_1(\omega, x_o, h) \right|^2 \quad (5.24)$$

where $|S(\omega)|^2$ denotes the Fourier power spectrum of the convolution operator under consideration. Obviously, the *mean* error power spectra of the two signals, at the given data resolution level, will be related through the following simple equation:

$$P_{e_2}(\omega, h) = |S(\omega)|^2 P_{e_1}(\omega, h) \quad (5.25)$$

since the convolution kernel $s(x)$ does not depend on the sampling phase x_o of the gridded input data; see eq.(5.9).

The mean error variance of the estimated output signal $\hat{g}_2(x)$ is given by the frequency-domain integral

$$\sigma_2^2(h) = \frac{1}{2\pi} \int P_{e_2}(\omega, h) d\omega \quad (5.26)$$

Based on the same methodology that was used in section 5.1.2 (see also Appendix C), the above error formula can easily be transformed to the following integral expression (for the case of a *symmetric* reference kernel $\varphi(x)$):

$$\sigma_2^2(h) = \frac{1}{2\pi} \int C_1(\omega) K(h\omega) d\omega \quad (5.27)$$

where $C_1(\omega)$ corresponds to the power spectrum of the input field $g_1(x)$. The auxiliary frequency-domain kernel $K(\omega)$ now depends not only on the used estimation filter $\Phi(\omega)$, but also on the connecting convolution kernel, and it has the analytical form

$$K(\omega) = \left| S\left(\frac{\omega}{h}\right) \right|^2 - 2 |\Phi(\omega)| \left| S\left(\frac{\omega}{h}\right) \right|^2 + \sum_k \left| S\left(\frac{\omega}{h} + \frac{2\pi k}{h}\right) \right|^2 |\Phi(\omega + 2\pi k)|^2 \quad (5.28)$$

Note that the behaviour of $K(\omega)$ directly depends on the data resolution level h , in contrast to the simple ‘interpolation’ case that was examined in section 5.1.2, where the corresponding error modelling kernel $K(\omega)$ had a fixed form (independent of the data sampling level); see eq.(5.17). Aliasing error analysis for signal interpolation (or quasi-interpolation) MR models can be treated as a special case of the previous input-output linear system framework, if we simply identify the convolution kernel $s(x)$ with the Dirac delta function $\delta(x)$.

5.1.4.1 Special Case 1: Band-limited Data Referencing Model

Let us now use the previous error propagation methodology for two particularly interesting special cases. The first corresponds to the choice $\varphi(x) = \text{sinc}(x)$ for the data referencing model, which implies a band-limited approximation of the convolution formula in eq.(5.21) over the Nyquist bandwidth $[-\pi/h, \pi/h]$. In order to derive the

simplified expression of the mean error variance for the band-limited output signal $\hat{g}_2(x)$, we should first decompose the power spectrum of the convolution kernel $s(x)$ as follows:

$$|S(\omega)|^2 = |S(\omega)|_{inner}^2 + |S(\omega)|_{outer}^2 \quad (5.29)$$

where the two individual components correspond to the disjoint parts of $|S(\omega)|^2$ within and outside the Nyquist bandwidth, respectively, i.e.

$$|S(\omega)|_{inner}^2 = \begin{cases} |S(\omega)|^2 & , \quad |\omega| \leq \pi / h \\ 0 & , \quad \text{elsewhere} \end{cases} \quad (5.30a)$$

and

$$|S(\omega)|_{outer}^2 = \begin{cases} |S(\omega)|^2 & , \quad |\omega| > \pi / h \\ 0 & , \quad \text{elsewhere} \end{cases} \quad (5.30b)$$

Using the above decomposition and the fact that the reference estimation filter $\Phi(\omega)$ corresponds to the indicator function $\chi_{-\pi, \pi}(\omega)$ (perfect low-pass filter), the general error formula of eq.(5.27) can be now simplified as follows:

$$\begin{aligned}
\sigma_2^2(h) &= \frac{1}{2\pi} \int C_1(\omega) \left[|S(\omega)|^2 - 2|S(\omega)|^2 \chi_{-\pi/h, \pi/h}(\omega) \right. \\
&\quad \left. + \sum_k \left| S(\omega + \frac{2\pi k}{h}) \right|^2 \chi_{-\pi/h, \pi/h}(\omega + 2\pi k) \right] d\omega \\
&= \frac{1}{2\pi} \int C_1(\omega) \left[|S(\omega)|^2 - 2|S(\omega)|_{inner}^2 + \sum_k \left| S(\omega + \frac{2\pi k}{h}) \right|_{inner}^2 \right] d\omega \\
&= \frac{1}{2\pi} \int C_1(\omega) \left[|S(\omega)|^2 - |S(\omega)|_{inner}^2 + \sum_{k \neq 0} \left| S(\omega + \frac{2\pi k}{h}) \right|_{inner}^2 \right] d\omega \\
&= \underbrace{\frac{1}{2\pi} \int C_1(\omega) |S(\omega)|_{outer}^2 d\omega}_{\substack{\text{output signal } g_2(x) \text{ energy} \\ \text{contained outside the Nyquist} \\ \text{bandwidth}}} + \underbrace{\frac{1}{2\pi} \int C_1(\omega) \sum_{k \neq 0} \left| S(\omega + \frac{2\pi k}{h}) \right|_{inner}^2 d\omega}_{\substack{\text{corrective term} \\ \text{depends on the input signal } g_1(x) \text{ energy} \\ \text{contained ONLY outside the Nyquist bandwidth}}} \\
&= \sigma_2^2(h)_a + \sigma_2^2(h)_b
\end{aligned} \tag{5.31}$$

Therefore, we see that in the band-limited approximation case the total error variance of the output field $\hat{g}_2(x)$ consists of two basic components, both of which directly depend on the input signal energy contained only *outside* the Nyquist bandwidth. When the convolution kernel is the Dirac delta function ($S(\omega) = 1$), then the above result is reduced to the mean variance for the band-limited signal interpolation error and the two variance components, $\sigma_2^2(h)_a$ and $\sigma_2^2(h)_b$, become equal.

5.1.4.2 Special Case 2: 'No' Data Referencing Model

Another interesting special approximation problem occurs when the data referencing model $\varphi(x)$ corresponds to the delta kernel, in which case the MR estimation filter takes the simple constant form $\Phi(\omega) = 1$. Such a situation is closely related to the usual FFT-based discrete numerical computations that are routinely applied in many gravity field convolution algorithms (e.g. geoid determination from gravity grids using Stokes's formula).

In this special case, the operational convolution formula in eq.(5.22) is reduced to the following form:

$$\hat{g}_2(x) = h \sum_n g_1(nh) s(x - nh) \quad (5.32)$$

In practice, the output signal is evaluated only in a finite network of discrete points, usually on the same grid in which the values of the input signal are given, i.e.

$$\hat{g}_2(mh) = h \sum_n g_1(nh) s(mh - nh) \quad (5.33)$$

using efficient FFT algorithms. Note that eqs.(5.32) and (5.33) correspond to a simple discretization process for the numerical (approximate) computation of the original

convolution formula in eq.(5.21). In fact, their use is exactly equivalent to applying the familiar parallelogram rule for numerical integration.

Although no specific continuous reference model $\hat{g}_1(x)$ for the input signal is implied in this case, we can still use the error formula from eq.(5.27) to compute the mean error variance of the continuous output field $\hat{g}_2(x)$, as implied by eq.(5.32). The scaled version of the error modelling kernel $K(\omega)$, according to eq.(5.28), will now have the simple form

$$\begin{aligned} K(h\omega) &= |S(\omega)|^2 - 2 |S(\omega)|^2 + \sum_k \left| S(\omega + \frac{2\pi k}{h}) \right|^2 \\ &= \sum_{k \neq 0} \left| S(\omega + \frac{2\pi k}{h}) \right|^2 \end{aligned} \quad (5.34a)$$

and the corresponding mean error variance for the output signal becomes

$$\begin{aligned} \sigma_2^2(h) &= \frac{1}{2\pi} \int C_1(\omega) \sum_{k \neq 0} \left| S(\omega + \frac{2\pi k}{h}) \right|^2 d\omega \\ &= \underbrace{\frac{1}{2\pi} \int C_1(\omega) \sum_{k \neq 0} \left| S(\omega + \frac{2\pi k}{h}) \right|_{outer}^2 d\omega}_{\text{depends on the input signal } g_1(x) \text{ energy}} + \underbrace{\frac{1}{2\pi} \int C_1(\omega) \sum_{k \neq 0} \left| S(\omega + \frac{2\pi k}{h}) \right|_{inner}^2 d\omega}_{\text{depends on the input signal } g_1(x) \text{ energy}} \\ &\quad \text{contained BOTH inside and outside} \quad \text{contained ONLY outside the Nyquist bandwidth} \end{aligned} \quad (5.34b)$$

The last equation can be compared with the corresponding error formula in eq.(5.31) that was derived for the special case of a band-limited data referencing model. They both have a common error variance component, which depends exclusively on the amount of input signal energy contained outside the Nyquist bandwidth $[-\pi/h, \pi/h]$. However, the first term in eq.(5.34b) is mostly affected by the input signal energy *within* the Nyquist bandwidth (depending on the decay of $|S(\omega)|^2$ and the specific data resolution level h), in contrast to the first error term in eq.(5.31) which still depends only on the input signal energy outside the Nyquist bandwidth. Such a comparison seems to support the initial claim that was made earlier in section 4.1, according to which the use of a proper MR referencing kernel $\varphi(x)$ for the discrete data will generally improve the accuracy of FFT-based numerical computations in linear input-output systems.

An illuminating approach to this important subject can be followed by comparing the mean error power spectra of the output signal, as computed with and without the help of a reference filter for the input data $g_1(nh)$. Let us denote by $P_{e_2}^\varphi(\omega, h)$ the mean error power spectrum for $\hat{g}_2(x)$ when a proper symmetric reference kernel $\varphi(x)$ is used in the operational algorithm of eq.(5.22), and by $P_{e_2}^\delta(\omega, h)$ the mean error power spectrum for $\hat{g}_2(x)$ when no specific data referencing model is employed, i.e. $\varphi(x) = \delta(x)$. In this way, from eqs.(5.10) and (5.25) we have

$$\begin{aligned}
P_{e_2}^{\Phi}(\omega, h) &= |S(\omega)|^2 C_1(\omega) - 2|S(\omega)|^2 C_1(\omega) |\Phi(h\omega)| \\
&\quad + |S(\omega)|^2 |\Phi(h\omega)|^2 \sum_k C_1\left(\omega + \frac{2\pi k}{h}\right)
\end{aligned} \tag{5.35a}$$

and

$$\begin{aligned}
P_{e_2}^{\delta}(\omega, h) &= |S(\omega)|^2 C_1(\omega) - 2|S(\omega)|^2 C_1(\omega) \\
&\quad + |S(\omega)|^2 \sum_k C_1\left(\omega + \frac{2\pi k}{h}\right)
\end{aligned} \tag{5.35b}$$

The reference estimation filter $\Phi(\omega)$ is always of ‘low-pass’ type and it generally satisfies the relation

$$|\Phi(\omega)| \leq 1 \Rightarrow 1 - |\Phi(\omega)| \geq 0 \tag{5.36}$$

Actually, at the origin it takes its maximum value ($|\Phi(0)| = 1$), in accordance with the general convergence condition given in eq.(5.12a). If we now form the difference between the two error power spectra, and taking also into account eq.(5.36), we will have

$$\begin{aligned}
\delta P(\omega, h) &= P_{e_2}^{\delta}(\omega, h) - P_{e_2}^{\Phi}(\omega, h) \\
&= |S(\omega)|^2 \left[\sum_k C_1\left(\omega + \frac{2\pi k}{h}\right) - 2 C_1(\omega) \right] - \\
&\quad - |S(\omega)|^2 \left[|\Phi(h\omega)|^2 \sum_k C_1\left(\omega + \frac{2\pi k}{h}\right) - 2 |\Phi(h\omega)| C_1(\omega) \right] \\
&\geq |S(\omega)|^2 \left[\sum_k C_1\left(\omega + \frac{2\pi k}{h}\right) - 2 C_1(\omega) \right] - \\
&\quad - |S(\omega)|^2 \left[|\Phi(h\omega)| \sum_k C_1\left(\omega + \frac{2\pi k}{h}\right) - 2 |\Phi(h\omega)| C_1(\omega) \right] \\
&= |S(\omega)|^2 [1 - |\Phi(h\omega)|] \underbrace{\left[\sum_k C_1\left(\omega + \frac{2\pi k}{h}\right) - 2 C_1(\omega) \right]}_{A(\omega)} \tag{5.37}
\end{aligned}$$

The above formula can serve as a useful guideline for studying the accuracy gain (or sometimes the accuracy loss) due to the incorporation of a low-pass reference filter $\Phi(\omega)$ in linear input-output (convolution) systems with gridded data. It is seen that the behaviour of the input signal power spectrum $C_1(\omega)$, in conjunction with the data resolution level h , exclusively determine if there is going to be an improvement in the output signal accuracy, since the sign of the difference $\delta P(\omega, h)$ depends *only* on the quantity $A(\omega)$. On the other hand, the actual reference filter $\Phi(\omega)$ and the convolution

operator $S(\omega)$ act basically as scaling factors, which control the significance of the approximation difference between the two methods.

It is quite obvious from eq.(5.37) that a large positive value for the integral $\int \delta P d\omega$ (which favors the use of a low-pass reference filter in the approximation framework) will occur in cases where the data resolution level h leaves a significant amount of input signal energy outside the Nyquist bandwidth, and additionally the power spectrum $|S(\omega)|^2$ of the convolution kernel is not significantly weak outside this bandwidth. On the other hand, as h becomes infinitely small (i.e. increasing data resolution), the difference between the two methods becomes negligible, since

$$\lim_{h \rightarrow 0} [1 - |\Phi(h\omega)|] = 1 - |\Phi(0)| = 0 \quad (5.38)$$

according to the convergence condition for the reference MR estimation filter. Although the preceding theoretical error analysis is very important, actual numerical investigations using simulated input signals and real-life convolution kernels are certainly required to justify the need of incorporating MR estimation reference filters in linear input-output systems with discrete data at varying resolutions.

5.1.5 Comparison with Wiener Filtering

An interesting similarity exists between equation (5.11) for the optimized mean error power spectrum $P_e(\omega, h)$, and the formula giving the PSD of the prediction error in the Wiener filtering theory for stationary random signals. In section 4.3.3, we had identified a similar situation between the actual Wiener filter and the optimal collocation-based MR filter that is implied in eq.(5.11).

According to Wiener's theory, the optimal estimation filter applied to a stationary zero-mean stochastic signal $g(x)$ contaminated with stationary random noise $v(x)$, under the assumption of zero correlation between the signal and the noise, produces a stationary zero-mean prediction error $e(x)$ with an associated PSD given by the formula (Sideris, 1995)

$$P_e(\omega) = P_g(\omega) \left(1 - \frac{P_g(\omega)}{P_g(\omega) + P_v(\omega)} \right) = P_g(\omega) \left(1 - \frac{P_g(\omega)}{P_{g'}(\omega)} \right) \quad (5.39)$$

where $P_g(\omega)$ is the PSD of the actual random signal and $P_v(\omega)$ is the PSD of the stationary input noise. The symbol $P_{g'}(\omega)$ denotes the PSD of the total input signal $g'(x) = g(x) + v(x)$ to the Wiener filter. In the absence of any input noise, the PSD of the prediction error is zero and the Wiener filter becomes a simple identity operator.

The two error formulas in eqs.(5.11) and (5.39) exhibit a strong algorithmic similarity in terms of a certain signal-to-noise ratio (SNR) form. Despite this interesting fact, they are each based on different mathematical principles and assumptions, and their underlying filtering settings correspond to two completely distinct physical situations. In particular, all signals involved in the Wiener filtering scheme are stochastic and continuous, whereas eq.(5.11) is based on a deterministic linear estimation methodology utilizing only discrete ‘noisy’ data at varying resolutions. In the latter case, the noise takes the form of the lost signal information due to the discretization of the original unknown field (see Figure 5.14), and when the data resolution becomes very high (‘zero noise’) then the associated mean error power spectrum in eq.(5.11) converges to zero.

Furthermore, the two estimation schemes employ entirely different concepts to define a signal error measure, in terms of either probabilistic (different experiment repetitions) or spatio-statistical (different sampling phases) averages, which is subsequently used in the optimization of the estimation filter for the input data of each case. The definition of these output error measures, as well as the overall filtering structure of the two estimation methodologies, are illustrated in Figure 5.14.

We should mention once more that no ‘stationarity’ assumption is used in the multiresolution filtering (spatio-statistical collocation) case, in contrast to the Wiener filtering scheme where the stationarity assumption (for both the signal and the observational noise) is essential.

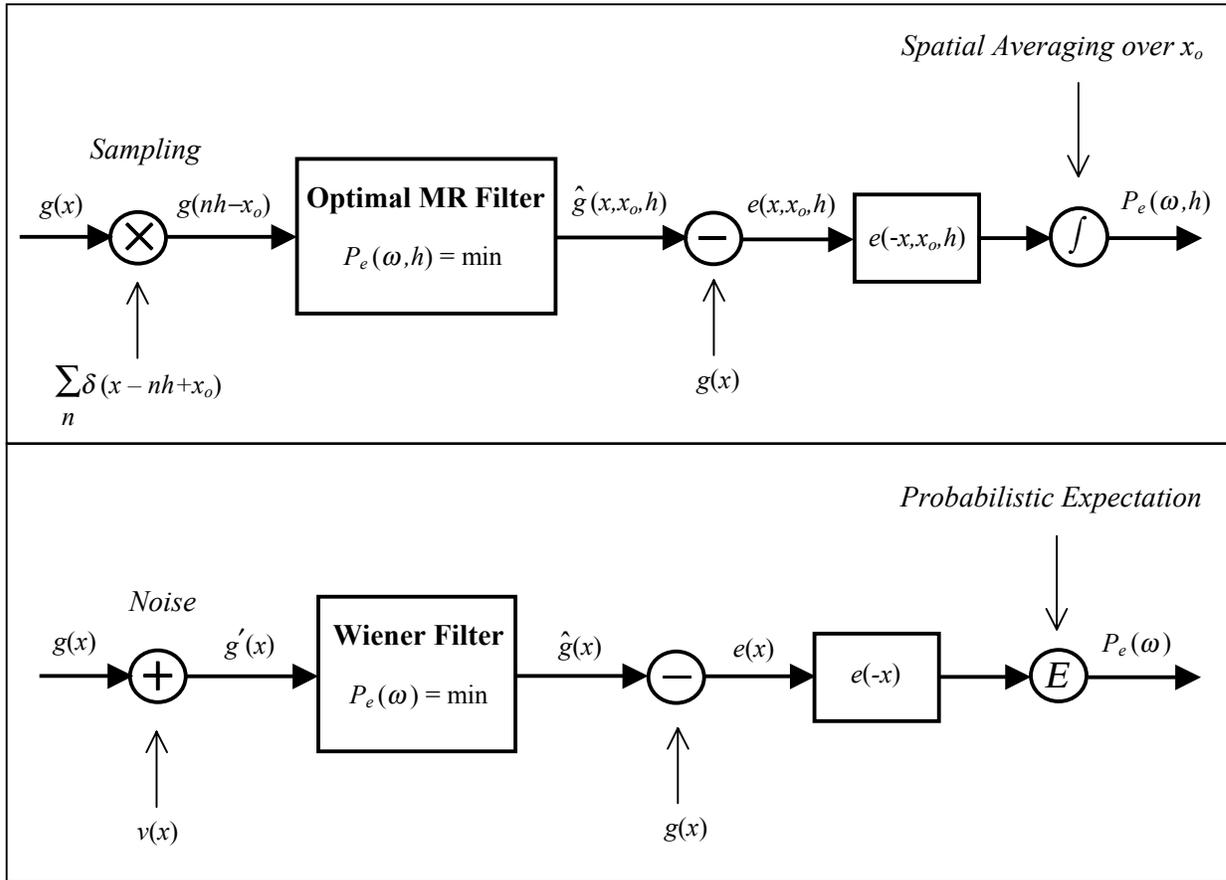


Figure 5.14 Diagrammatic comparison between the classic Wiener estimation filter (bottom system) and optimal translation-invariant linear MR approximation using deterministic gridded data (top system)

However, it is the translation-invariance condition, which was imposed in the deterministic case, that makes the two linear estimation models comparable in terms of convolution-based filters that are applied to the input data of each case. The more general problem where the discrete input data in the MR filtering scheme of eq.(5.1) are influenced by random noise will be examined in the next section.

Concluding this comparison between the spatio-statistical collocation scheme and the Wiener filtering methodology, we should point out one final important difference. Stationary random processes are theoretically defined as infinitely lasting signals with a uniform behaviour across their domain, which obviously do not possess a well defined Fourier transform (i.e. they are not finite-energy signals)*. The stochastic input signal $g(x)$ and the additive random noise $v(x)$ in the Wiener filter case should thus be considered as truncated ensemble realizations of stationary processes with a finite record length T , in order for the linear estimation scheme to be physically realizable. Accordingly, the definition of the output error PSD shown in Figure 5.14 is based on the well known *Wiener-Khinchine* relationship, which takes the expectation of $|E(\omega)|^2$ over different finite data records with length T , multiplies this average error power spectrum by $1/T$, and finally lets T increase beyond bound; see Bendat and Piersol (1993). In this way, the integration of $P_e(\omega)$ will yield the usual (probabilistic) variance σ_e^2 of the stationary prediction error $e(x)$, whose square root is expressed in the same units as the input signal and the input noise.

On the other hand, in the MR filtering scheme of spatio-statistical collocation we do not face the above complications, since the unknown field is a-priori modelled as a finite-energy deterministic signal, with the additional mild requirement that its sampling always

* Note that we do not consider cases of compact signal domains in this thesis. Since most of our developments are restricted in a 1D setting, the domain of the signals should always be understood as the whole real line.

results in a square-summable data sequence. The integration of the mean error power spectrum $P_e(\omega, h)$, at a certain data resolution level h , yields the mean square L^2 norm of the approximation error $e(x, x_o, h)$ averaged over all possible sampling phase values. If we want to convert this deterministic error norm into an RMS-type measure that is compatible with the data signal units, we can simply divide the mean error ‘variance’ $\sigma^2(h)$ by the finite extent of the spatial field $g(x)$, and then take the square root of the resulting value. Note that such an operation only gives an *estimate* of the RMS signal error, since the support of the error signal $e(x, x_o, h)$ may be larger than the support of the unknown field, depending on the localization properties of the approximation kernel $\varphi(x)$ and the actual data resolution level. Furthermore, it should always be kept in mind that the value of $\sigma^2(h)$ corresponds to an average performance at a specific resolution level h , which may deviate from the actual L^2 error norm produced by a given data set with a certain sampling phase x_o . For this type of error variability problems, see Blu and Unser (1999a).

5.2 Noise Filtering

The signal error analysis in the previous sections was based entirely on a *noiseless* data setting, taking into account only the finite resolution of the available observations. In practice, however, the discrete data can hardly be considered noise-free. Hence, it is very

important to also study the case of having additive random noise in the samples of an unknown deterministic field when using the linear MR estimation model of eq.(5.1). In particular, we should consider the problem of modifying the original scaling kernel $\varphi(x)$, that we would normally use with noiseless data, in order to reduce the effect of the propagated noise in the final approximated field as much as possible.

5.2.1 Continuous Versus Discrete Noise

Let us assume that we have available noisy samples $d(nh)$ of an unknown field $g(x)$, taken at a spatial resolution level h . Then, we can write the following equation:

$$d(nh) = g(nh) + v(nh) \quad (5.40)$$

where $g(nh)$ are the true values of the unknown deterministic signal, and $v(nh)$ represents a discrete (in general non-stationary) random noise sequence. The associated stochastic model which is used to describe the behaviour of the measurement noise is expressed by the formulae

$$E\{v(nh)\} = 0$$

$$E\{v^2(nh)\} = \sigma_v^2(nh) = \sigma_v(nh, nh) \quad (5.41)$$

$$E\{v(nh)v(mh)\} = \sigma_v(nh, mh)$$

where $\sigma_v^2(nh)$ is the noise variance at each data point, and $\sigma_v(nh, mh)$ is the noise spatial covariance between the data points $x = nh$ and $x = mh$. Note that we use a purely discrete model for the data noise, in contrast to other formulations of the optimal estimation problem in gravity field signals where a spatially continuous (and stationary) model is usually employed with the help of a stationary noise CV function (Sideris, 1995; Li, 1996c; Li and Sideris, 1997; Tziavos et al., 1996). The use of continuous noise models is a rather questionable issue within an estimation framework utilizing only discrete spatial data. The measurement noise does not generally exist in a physical sense as a continuous spatial signal (i.e. we do not ‘sample the noise’), but it originates only because we performed an observation with an imperfect instrument under certain external influences at a specific point in space.

On the other hand, there exist cases of signal approximation problems with discrete spatial data where it does make sense to consider continuous noise models. For example, a data acquisition device may change its noise characteristics, as it moves from one spatial point to another, according to a given continuous (time-dependent) stochastic model. However, since we always collect (and process) observations at a finite network of data points, the input noise will still be a discrete signal (in a spatial sense) with an associated discrete stochastic model. The latter is determined in such cases, at the points of interest, from a continuous time-dependent error model and the spatio-temporal ‘path’ of the measuring device. One may also consider extreme cases in signal estimation

problems with discrete spatial data where the observational noise (or some part of it) does indeed exist in a spatially continuous sense (e.g. atmospheric effects in various types of measurements). Such cases are not considered here and they can usually be treated by applying a-priori corrections to the discrete data for these spatially continuous noise effects, before the optimal estimation takes place.

In any case, the important aspect to be emphasized here is that we do not really need continuous noise models for our finite-resolution discrete data in signal approximation schemes as in eq.(5.1). As it will be demonstrated in the next sections, all that is **algorithmically** necessary is the discrete model given in eq.(5.41), even if the underlying data noise is generated by continuous (time/spatial) stochastic phenomena. For the treatment of noise in continuous fields of measurements, see Sanso and Sona (1995).

5.2.2 General Formulation

As in the case of the noiseless approximation problem (see section 4.3), the optimal signal estimate $\hat{g}(x)$ will employ the discrete noisy data $d(nh)$ in a linear and translation-invariant fashion. The ‘need’ to obtain a translation-invariant signal estimate (i.e. independent of the spatial reference system) is **not** affected by the presence of (stationary or non-stationary) noise in the observations. Based on these two assumptions, the estimation formula should have the familiar convolution-type form

$$\hat{g}(x) = \sum_n d(nh) \xi_h(x - nh) \quad (5.42)$$

where $\xi_h(x)$ is an unknown kernel, which generally depends on the data resolution level h . The above equation can be illustrated through the linear filtering procedure shown in Figure 5.15.

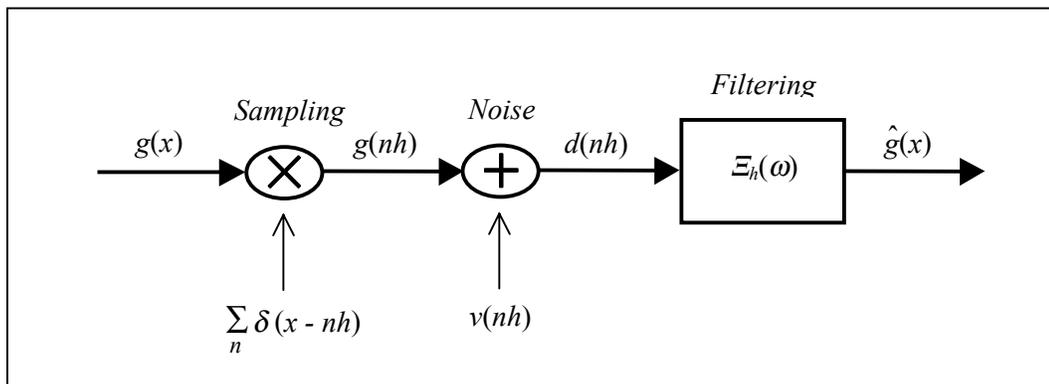


Figure 5.15 Linear translation-invariant signal estimation using discrete noisy data

In order to determine an optimal form for the kernel $\xi_h(x)$, we have to introduce an associated optimality principle. The signal error produced by the filtering equation (5.42) can be decomposed into two distinct components

$$e(x) = e_h(x) + e_v(x) = g(x) - \hat{g}(x) \quad (5.43)$$

where $e_h(x)$ is the part of the total estimation error caused from the use of discrete data with finite sampling resolution (aliasing error), and $e_v(x)$ is the additional part due to the

noise presence in the data samples. Note that the fundamental difference between these two error components is that $e_h(x)$ is a purely deterministic signal (whose behaviour has been modelled and studied in section 5.1), whereas $e_v(x)$ is a stochastic signal originating exclusively from the presence of random noise in the discrete data.

In the absence of any noise from the input data, the best we can do is to obtain just an *interpolated model* $m_g(x)$ for the unknown field, that will depend on the true signal samples $g(nh)$ at the given spatial resolution. We will assume that such a signal model is given in terms of a linear translation-invariant form, as follows:

$$m_g(x) = \sum_n g(nh) \varphi_h(x-nh) \quad (5.44)$$

The noise-dependent error component $e_v(x)$ should be measured relative to such a **noiseless ‘reference model’** for the unknown field. In this way, we will have

$$\begin{aligned} e_v(x) &= m_g(x) - \hat{g}(x) \\ &= \sum_n g(nh) \varphi_h(x-nh) - \sum_n d(nh) \xi_h(x-nh) \end{aligned} \quad (5.45)$$

whereas the resolution-dependent deterministic error term $e_h(x)$ is

$$\begin{aligned}
e_h(x) &= g(x) - m_g(x) \\
&= g(x) - \sum_n g(nh) \varphi_h(x - nh)
\end{aligned}
\tag{5.46}$$

The signal model $m_g(x)$ is controlled by a certain MR scaling kernel $\varphi_h(x) = \varphi(x/h)$, which should obey all the admissibility conditions addressed in section 5.1. Note that the above error decomposition provides great flexibility for the design of the noise filtering kernel $\xi_h(x)$, since the choice of $\varphi(x)$ may be arbitrarily based on noise-independent modelling criteria for the behaviour of the unknown field at the given resolution level h . As a matter of fact, the problem of linear noise filtering for signals that belong in an arbitrary MRA subspace can be embedded into the previous formulation, if we identify the reference kernel $\varphi(x)$ with the scaling sampling function of the MRA under consideration. Regardless of the behaviour of the unknown signal, it is convenient to choose an orthonormal reference scaling kernel because of the simplification that is achieved in the study of the resolution-dependent error component $e_h(x)$ (see section 5.1.2).

The optimal determination of the data noise filter will be based on the minimization of a suitable functional of the stochastic error component $e_v(x)$ given in eq.(5.45). Such an optimization procedure is analytically described in the next section.

5.2.3 Optimization of the Noise Filter

The optimization of the noise filtering kernel $\xi_h(x)$ will be carried out in this section, following a frequency-domain methodology. In particular, the optimal estimation criterion that we will use has the familiar mean-square-error (MSE) expression

$$E \left\{ \left| E_v(\omega) \right|^2 \right\} = \min \quad (5.47)$$

where $E_v(\omega)$ denotes the Fourier transform of the noise-dependent error component $e_v(x)$. Note that the term ‘mean’ corresponds to its usual probabilistic interpretation (i.e. E is the classic expectation operator), in contrast to the optimization scheme that was followed in the previous chapter where the MSE was defined in a spatio-statistical deterministic sense. The reference scaling kernel $\varphi_h(x) = \varphi(x/h)$ was then optimized using the spatio-statistical power spectrum of the resolution-dependent error $e_h(x)$, whereas the noise filtering kernel $\xi_h(x)$ will now be optimized using the mean power spectrum of the noise-dependent error $e_v(x)$.

From eq.(5.45), the Fourier transform of $e_v(x)$ will take the following form:

$$\begin{aligned} E_v(\omega) &= M_g(\omega) - \hat{G}(\omega) \\ &= \Phi_h(\omega) \sum_n g(nh) e^{-i\omega nh} - \Xi_h(\omega) \sum_n d(nh) e^{-i\omega nh} = \end{aligned}$$

$$\begin{aligned}
&= \Phi_h(\omega) \frac{1}{h} \sum_k G\left(\omega + \frac{2\pi k}{h}\right) - \Xi_h(\omega) \sum_n (g(nh) + v(nh)) e^{-i\omega nh} \\
&= \Phi_h(\omega) \bar{G}_h(\omega) - \Xi_h(\omega) [\bar{G}_h(\omega) + \bar{V}_h(\omega)] \\
&= \Phi_h(\omega) \bar{G}_h(\omega) - \Xi_h(\omega) \bar{G}_h(\omega) - \Xi_h(\omega) \bar{V}_h(\omega)
\end{aligned} \tag{5.48}$$

where $\Phi_h(\omega)$ and $\Xi_h(\omega)$ denote the Fourier transforms of the reference kernel $\varphi_h(x)$ and the noise filtering kernel $\xi_h(x)$, respectively. The term $\bar{G}_h(\omega)$ corresponds to the periodic Fourier transform of the sequence formed by the true signal values $g(nh)$, whereas $\bar{V}_h(\omega)$ is the periodic Fourier transform of the sequence formed by the actual noise values at every data point, i.e.

$$\bar{V}_h(\omega) = \sum_n v(nh) e^{-i\omega nh} \tag{5.49}$$

In order to ensure the existence (convergence) of the two periodic Fourier transforms $\bar{G}_h(\omega)$ and $\bar{V}_h(\omega)$, both the signal and the noise values should be measurable in the sense of eq.(4.7). One might think that such a condition is contradictory with the concept of *stationary noise*, which is generally understood as an infinitely lasting process with uniform statistical behaviour across the whole real line. However, in physical applications the unknown spatial fields cover only finite regions and, as a result, the condition in eq.(4.7) for both the signal and the noise (stationary or not) is always met. As Ronald

Bracewell noted in his classic book (Bracewell, 1986; p. 9): ‘*the question of the existence of Fourier transforms may safely be ignored when the signal to be transformed is an accurately specified description of a physical quantity. Physical possibility is a valid sufficient condition for the existence of the Fourier transform*’.

Using eq.(5.48), we can derive the noise-dependent error power spectrum as follows:

$$\begin{aligned}
|E_v(\omega)|^2 &= E_v(\omega) E_v^*(\omega) \\
&= \Phi_h(\omega)\Phi_h^*(\omega)\bar{G}_h(\omega)\bar{G}_h^*(\omega) - \Phi_h(\omega)\Xi_h^*(\omega)\bar{G}_h(\omega)\bar{G}_h^*(\omega) - \Phi_h(\omega)\Xi_h^*(\omega)\bar{G}_h(\omega)\bar{V}_h^*(\omega) \\
&\quad - \Phi_h^*(\omega)\Xi_h(\omega)\bar{G}_h(\omega)\bar{G}_h^*(\omega) + \Xi_h(\omega)\Xi_h^*(\omega)\bar{G}_h(\omega)\bar{G}_h^*(\omega) + \Xi_h(\omega)\Xi_h^*(\omega)\bar{G}_h(\omega)\bar{V}_h^*(\omega) \\
&\quad - \Phi_h^*(\omega)\Xi_h(\omega)\bar{G}_h^*(\omega)\bar{V}_h(\omega) + \Xi_h(\omega)\Xi_h^*(\omega)\bar{G}_h^*(\omega)\bar{V}_h(\omega) + \Xi_h(\omega)\Xi_h^*(\omega)\bar{V}_h(\omega)\bar{V}_h^*(\omega)
\end{aligned} \tag{5.50}$$

and by applying the expectation operator to the above expression, we obtain the mean error power spectrum

$$\begin{aligned}
E \left\{ |E_v(\omega)|^2 \right\} &= \Phi_h(\omega)\Phi_h^*(\omega)\bar{G}_h(\omega)\bar{G}_h^*(\omega) - \Phi_h(\omega)\Xi_h^*(\omega)\bar{G}_h(\omega)\bar{G}_h^*(\omega) \\
&\quad - \Phi_h^*(\omega)\Xi_h(\omega)\bar{G}_h(\omega)\bar{G}_h^*(\omega) + \Xi_h(\omega)\Xi_h^*(\omega)\bar{G}_h(\omega)\bar{G}_h^*(\omega) \\
&\quad + \Xi_h(\omega)\Xi_h^*(\omega)\bar{P}_v(\omega)
\end{aligned} \tag{5.51}$$

where the term $\bar{P}_v(\omega)$ is used to denote the following quantity:

$$\bar{P}_v(\omega) = E \left\{ \bar{V}_h(\omega)\bar{V}_h^*(\omega) \right\} = E \left\{ |\bar{V}_h(\omega)|^2 \right\} \tag{5.52}$$

The determination of the optimal noise filter $\Xi_h(\omega)$ can now be easily made using eqs.(5.47) and (5.51). The underlying procedure is straightforward (see, e.g. Bendat and Piersol, 1986; pp. 182-183) and it gives the final Wiener-like result

$$\begin{aligned}
 \Xi_h(\omega) &= \frac{\overline{G}_h(\omega) \overline{G}_h^*(\omega)}{\overline{G}_h(\omega) \overline{G}_h^*(\omega) + \overline{P}_v(\omega)} \Phi_h(\omega) \\
 &= \frac{|\overline{G}_h(\omega)|^2}{|\overline{G}_h(\omega)|^2 + \overline{P}_v(\omega)} \Phi_h(\omega) \\
 &= \overline{W}_h(\omega) \Phi_h(\omega)
 \end{aligned} \tag{5.53}$$

In the following section, the two-step (separable) filtering structure of the above optimal result is analytically discussed.

5.2.4 The Cascade Structure of the Optimal Noise Filter

There are several important remarks that should be made at this point regarding the final result of the previous section. First, it is seen that the optimal estimation procedure is basically decomposed into two individual steps (filters), which are connected in a linear cascading manner. The first step, expressed by the periodic filter component $\overline{W}_h(\omega)$, has the role of ‘denoising’ the discrete input data $d(nh)$ using certain information about the average behaviour of the input noise and the underlying unknown field. In contrast, the

second filter component $\Phi_h(\omega)$ is solely used to obtain a continuous representation for the final estimated field $\hat{g}(x)$, based on a properly selected reference scaling kernel $\varphi_h(x) = \varphi(x/h)$. In this way, the result of the optimization procedure in Chapter 4 (for the resolution-dependent signal error $e_h(x)$) can easily be incorporated into the current optimal result for the noise-dependent signal error $e_v(x)$, if we set the quantity $\Phi_h(\omega)$ equal to the collocation-based MR filter according to eq.(4.17). The two basic steps of the optimal noise filtering procedure are illustrated in Figure 5.16.

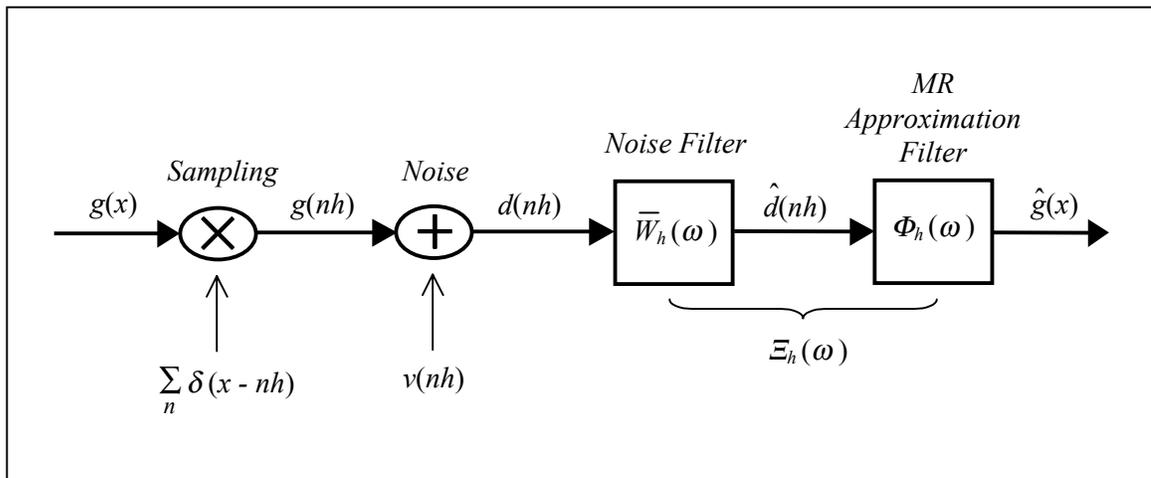


Figure 5.16 Two-step optimal translation-invariant filtering of discrete noisy data

From the previous figure, we can see that it is not necessary to modify the scaling kernel $\varphi(x)$ of the linear MR approximation model in eq.(5.1) when dealing with noisy input data. The optimization of the noise-dependent output error, according to eq.(5.45), adds just an intermediate filtering step that is applied to the original discrete data $d(nh)$ and it

produces a new estimated sequence $\hat{d}(nh)$ in which the effect of the observational noise has been minimized in a certain translation-invariant linear fashion. We can then use this ‘synthetic’ data sequence as input to the basic estimation model of eq.(5.1) in order to get a continuous approximation of the underlying unknown field at the given resolution level h .

The structure of the noise filter $\bar{W}_h(\omega)$ is very similar to the classic Wiener estimation filter used in random field theory. There exist, however, significant conceptual differences between the two optimal estimation schemes as well, since: (i) the underlying unknown signal $g(x)$ has not been modelled as a stationary stochastic process, and (ii) the additive input noise has not been assumed to be stationary. In order to better understand the Wiener-like behaviour of the filter $\bar{W}_h(\omega)$, let us analyze its two main building components in more detail. We can write the following:

$$\begin{aligned} \bar{W}_h(\omega) &= \frac{|\bar{G}_h(\omega)|^2}{|\bar{G}_h(\omega)|^2 + \bar{P}_v(\omega)} = \frac{\frac{h}{X} |\bar{G}_h(\omega)|^2}{\frac{h}{X} |\bar{G}_h(\omega)|^2 + \frac{h}{X} \bar{P}_v(\omega)} \\ &= \frac{\bar{C}_1(\omega)}{\bar{C}_1(\omega) + \bar{C}_2(\omega)} \end{aligned} \tag{5.54}$$

where X denotes the finite extent of the unknown spatial field $g(x)$. The two auxiliary functions, $\bar{C}_1(\omega)$ and $\bar{C}_2(\omega)$, correspond to the periodic Fourier transforms of two associated space-domain sequences, $c_1(nh)$ and $c_2(nh)$, which have the form

$$c_1(nh) = \frac{h}{X} \sum_m g(mh) g(mh + nh) \quad (5.55a)$$

and

$$c_2(nh) = \frac{h}{X} \sum_m \sigma_v(mh, mh + nh) \quad (5.55b)$$

We shall demonstrate the above fact for the case of the noise-dependent Fourier pair $c_2(nh) \leftrightarrow \bar{C}_2(\omega)$; a similar methodology can also be applied for the signal-dependent Fourier pair $c_1(nh) \leftrightarrow \bar{C}_1(\omega)$. We have the following:

$$\begin{aligned} \bar{C}_2(\omega) &= \frac{h}{X} \bar{P}_v(\omega) = \frac{h}{X} E \left\{ \left| \bar{V}_h(\omega) \right|^2 \right\} = \frac{h}{X} E \left\{ \bar{V}_h(\omega) \bar{V}_h^*(\omega) \right\} \\ &= \frac{h}{X} E \left\{ \sum_n \sum_m v(nh) v^*(mh) e^{-i\omega nh} e^{i\omega mh} \right\} \\ &= \frac{h}{X} E \left\{ \sum_n \sum_m v(nh) v(mh) e^{-i\omega(n-m)h} \right\} \\ &= \frac{h}{X} \sum_n \sum_m E \{ v(nh) v(mh) \} e^{-i\omega(n-m)h} = \end{aligned}$$

$$\begin{aligned}
&= \frac{h}{X} \sum_n \sum_m \sigma_v(nh, mh) e^{-i\omega(n-m)h} \\
&= \frac{h}{X} \sum_k \sum_m \sigma_v(mh + kh, mh) e^{-i\omega kh} \\
&= \frac{h}{X} \sum_n \sum_m \sigma_v(mh, mh + nh) e^{-i\omega nh} \\
&= \sum_n c_2(nh) e^{-i\omega nh}
\end{aligned} \tag{5.56}$$

The sequence $c_1(nh)$ in eq.(5.55a) can be identified as the *discrete* spatial CV function of the unknown signal at a given resolution level h . This function contains less information than the *continuous* signal CV function that was used in the previous chapter for optimizing the MR reference kernel $\varphi_h(x)$, since it takes into account only the discrete values of the unknown field at a certain resolution and at a certain sampling phase (the sampling phase x_o of the input data has been assumed zero in this section, see eq.(5.40)). Note that the discrete spatial CV function does not generally correspond to a sampled version of the continuous signal CV function given in eq.(4.15) or, equivalently, the quantity $|\overline{G}_h(\omega)|^2$ is not generally obtained by a simple periodization of the true signal power spectrum $|G(\omega)|^2$, since

$$|\overline{G}_h(\omega)|^2 = \left| \frac{1}{h} \sum_k G(\omega + \frac{2\pi k}{h}) \right|^2 \neq \frac{1}{h^2} \sum_k \left| G(\omega + \frac{2\pi k}{h}) \right|^2 \tag{5.57}$$

The second sequence $c_2(nh)$ in eq.(5.55b), on the other hand, cannot be interpreted as the discrete noise CV function and, as a result, the quantity $\bar{P}_v(\omega)$ should not generally be viewed as the data noise PSD. Such an interpretation is possible only in the special situation where the additive noise $v(nh)$ is stationary, in which case the sequence $c_2(nh)$ is reduced to the probabilistic CV function of the discrete input noise.

The similarities of our noise filtering framework with the Wiener filtering formalism stem from our initial modelling choice in eq.(5.42) that the signal estimation procedure should *always* be linear and translation-invariant. This results in a convolution-based algorithmic scheme in terms of an SNR-type optimal filter. The differences associated with the two methodologies are due to the fact that the data noise is not restricted to being stationary in our case, in contrast to Wiener filtering where the stationarity assumption, for both the signal and the noise, is crucial. The unknown field $g(x)$ has been modelled as a deterministic finite-energy signal in our case, with the only additional restriction being that it should have a compact spatial support X (such an assumption is needed in order for $c_2(nh)$ and $\bar{C}_2(\omega)$ to be well defined in the stationary noise case). Note that if the input noise is non-stationary then the signal estimation algorithm, according to Wiener's linear theory, can no longer be reduced to a simple filtering operation but, instead, it takes the form of a complicated integral equation (Wiener-Hopf formula) whose solution determines the best linear (*but not translation-invariant in this case*) signal estimate in a MSE sense; see Sanso and Sideris (1997). In our approximation framework, however, the

a-priori imposed condition of translation-invariance for the optimal signal estimate allows us to treat both stationary and non-stationary noise cases in a unified linear filtering (convolution-type) manner, which can be efficiently implemented via FFT techniques.

5.2.5 Additional Remarks

Concluding this chapter, a few additional remarks will be given about the noise filtering framework that was presented in the preceding paragraphs. One important aspect for practical applications is the **realization** of the periodic noise filter $\overline{W}_h(\omega)$ given in eq.(5.54). Although the input noise ‘PSD’ $\overline{P}_v(\omega)$ can always be determined from the noise variances and covariances using the discrete Fourier transform of the sequence $c_2(nh)$, the signal ‘PSD’ $|\overline{G}_h(\omega)|^2$ is generally unknown since it depends on the true signal values $g(nh)$ at the specific resolution. The same problem also exists in the classic Wiener filter theory, which requires the knowledge of the noiseless (stochastic) signal PSD function. In our case, one possibility to overcome this filter realization problem is to introduce a certain ‘external’ model for the discrete signal CV function $c_1(nh)$ at the given resolution level, and then simply compute its discrete Fourier transform needed in eq.(5.54).

An alternative methodology can also be followed using the noisy data set $d(nh)$ in order to empirically estimate the deterministic signal ‘PSD’. If we denote by $\bar{D}_h(\omega)$ the discrete Fourier transform of the input data sequence, we obviously have

$$\bar{D}_h(\omega) = \sum_n d(nh) e^{-i\omega nh} \quad (5.58a)$$

and

$$\bar{D}_h(\omega) = \bar{G}_h(\omega) + \bar{V}_h(\omega) \quad (5.58b)$$

where the discrete Fourier transform $\bar{V}_h(\omega)$ of the noise is given by eq.(5.49). Using the last equation, we can obtain the following relationship:

$$\left| \bar{D}_h(\omega) \right|^2 = \left| \bar{G}_h(\omega) \right|^2 + \bar{G}_h(\omega) \bar{V}_h^*(\omega) + \bar{G}_h^*(\omega) \bar{V}_h(\omega) + \left| \bar{V}_h(\omega) \right|^2 \quad (5.59a)$$

and by applying the expectation operator

$$\begin{aligned} E \left\{ \left| \bar{D}_h(\omega) \right|^2 \right\} &= \left| \bar{G}_h(\omega) \right|^2 + E \left\{ \left| \bar{V}_h(\omega) \right|^2 \right\} \\ &= \left| \bar{G}_h(\omega) \right|^2 + \bar{P}_v(\omega) \end{aligned} \quad (5.59b)$$

The signal ‘PSD’ $|\bar{G}_h(\omega)|^2$ can now be determined empirically through the last formula, using the available realization $|\bar{D}_h(\omega)|^2$ of the data power spectrum as an estimate of its expected value $E\left\{|\bar{D}_h(\omega)|^2\right\}$.

The use of the optimal (separable) estimation filter $\Xi_h(\omega)$, according to eq.(5.53), gives rise to the following expression for the mean power spectrum of the noise-dependent error component $e_v(x)$:

$$\begin{aligned}
 E\left\{|E_v(\omega)|^2\right\} &= \frac{|\bar{G}_h(\omega)|^2}{|\bar{G}_h(\omega)|^2 + \bar{P}_v(\omega)} \bar{P}_v(\omega) |\Phi_h(\omega)|^2 \\
 &= \bar{W}_h(\omega) \bar{P}_v(\omega) |\Phi_h(\omega)|^2 \\
 &= h^2 \bar{W}_h(\omega) \bar{P}_v(\omega) |\Phi(h\omega)|^2
 \end{aligned} \tag{5.60}$$

since the reference kernel of the noiseless signal model is always assumed to have the scaling form $\varphi_h(x) = \varphi(x/h)$. An interesting point in the last formula is that the data resolution level h acts as a scaling factor, which makes the noise-dependent signal error decrease as the data sampling density increases! This result should not be surprising to signal analysts, since it is well known that linear signal approximation schemes based on oversampling can lead to significant noise reduction (Benedetto, 1998). Note also that the

inverse Fourier transform of the above equation will not necessarily correspond to an ‘error CV function’, since the noise-dependent error component $e_v(x)$ is not necessarily a stationary random signal.

A key point in the development of our translation-invariant estimation procedure was the decomposition of the total signal error into a resolution-dependent part $e_h(x)$ and a noise-dependent part $e_v(x)$; see eqs.(5.45) and (5.46). Such a modelling choice offers the possibility of studying/optimizing individually the effects of the finite data resolution and the additive input noise to the final signal estimate $\hat{g}(x)$, using appropriate error measures and criteria for each case. Note that the second error component $e_v(x)$ is not entirely independent from the actual data resolution, but it actually depends on h according to eq.(5.60). It is also interesting that the optimization of the noise-dependent signal error led to a separable filter solution $\Xi_h(\omega)$, which simplifies the linear estimation algorithm into two basic and distinct steps (discrete data denoising + signal ‘interpolation’); see Figure 5.16. An additional advantage due to the separability of the optimal noise filter is that it allows us to treat the final estimated field $\hat{g}(x)$ as a member of a Hilbert subspace $V_h \subset L^2(\mathfrak{R})$, where V_h is the linear space spanned by the translation-invariant Riesz basis $\varphi(\frac{x}{h} - n)_{n \in \mathbb{Z}}$ generated from the chosen reference scaling kernel. In this way, we may incorporate the various MRA tools (e.g. wavelet

spectral analysis) into our optimal signal approximation framework not only for noiseless data schemes (see Chapter 4), but also for cases with noisy input data.

The partition of the total signal error $e(x)$ in terms of resolution-dependent and noise-dependent disjoint parts is actually more than an arbitrary modelling choice. If, instead of the optimal criterion given in eq.(5.47), we had used the following ‘total’ MSE estimation principle:

$$E \left\{ |E(\omega)|^2 \right\} = \min \quad (5.61)$$

then the optimized filtering kernel $\xi_h(x)$ would have the frequency-domain form

$$\bar{\Xi}_h(\omega) = \frac{G(\omega) \bar{G}_h^*(\omega)}{\bar{G}_h(\omega) \bar{G}_h^*(\omega) + \bar{P}_v(\omega)} = \frac{\bar{G}_h^*(\omega)}{|\bar{G}_h(\omega)|^2 + \bar{P}_v(\omega)} G(\omega) \quad (5.62)$$

The above result is practically useless, since it requires the knowledge of the total unknown field $G(\omega)$ beforehand. The signal error decomposition, according to eqs.(5.45) and (5.46), is thus necessary in order to obtain a reasonable (i.e. realizable) optimal solution for the noise filtering kernel $\xi_h(x)$ in terms of signal/noise ‘PSD’ information.

This decomposition requires the introduction of an intermediate translation-invariant noiseless model $m_g(x)$ for the unknown signal, which can be either arbitrary or it may be determined through a separate optimization procedure according to Chapter 4. Such a

procedure leads to the substitution of the true signal Fourier transform $G(\omega)$ in eq.(5.62) with a noiseless model $M_g(\omega) = \overline{G}_h(\omega) \Phi_h(\omega)$ at the given resolution level; see eq.(5.44).

The theoretical analysis presented in section 5.2 provides an original approach to the linear approximation problem in deterministic spatial fields using discrete noisy samples at a given resolution level. The classic methodology that has traditionally been applied for this type of approximation problems in geodesy is *Tikhonov regularization*, which employs a hybrid optimal estimation principle in terms of a deterministic signal norm and a quadratic (Euclidean) observation error norm; see Moritz (1980, pp. 238-249). The simultaneous minimization of these two quantities yields a unique optimal solution, which corresponds to the ‘smoothest’ field with the smallest possible deviations from its observed noisy values. Such an approximation procedure is identical to using the deterministic collocation method within a certain Hilbert space, when the available linear functionals are influenced by zero-mean random errors; for more details, see Moritz (1980, sect. 28-30) and Dermanis and Sanso (1997, ch. 8). The main problems associated with the above estimation methodology encompass all the issues that were previously discussed in Chapter 3 of this thesis (i.e. choice of the norm/reproducing kernel, stability with respect to increasing data resolution, compatibility between data and model resolution). In order to overcome these types of problems, we have essentially replaced the arbitrary signal norm (field smoothing condition) that appears in Tikhonov’s

regularization principle with the notion of a noiseless signal model at the given data resolution level, according to the linear multiresolution expansion of eq.(5.44).

5.3 A Physical Geodesy Example

A simple theoretical example will now be presented in order to demonstrate the applicability of the signal estimation framework that was developed in the last two chapters. The example is taken from physical geodesy and it corresponds to the problem of local gravimetric geoid determination using Stokes's integral formula under a planar approximation. We will assume that the input data are discrete noisy gravity anomalies, given on a 2D orthogonal planar grid with uniform resolution level h in both directions. The true gravity anomaly signal will be denoted by $\Delta g(\bar{\mathbf{x}})$ and its gridded noisy samples by $\Delta g^v(h \cdot \bar{\mathbf{k}})$, where $\bar{\mathbf{x}}$ and $\bar{\mathbf{k}}$ are (2×1) vectors in the Euclidean spaces \mathfrak{R}^2 and Z^2 , respectively. We will also use the symbol $v(h \cdot \bar{\mathbf{k}})$ to denote a 2D (non-stationary in general) random noise sequence, such that

$$\Delta g^v(h \cdot \bar{\mathbf{k}}) = \Delta g(h \cdot \bar{\mathbf{k}}) + v(h \cdot \bar{\mathbf{k}}) \quad (5.63)$$

where $\Delta g(h \cdot \bar{\mathbf{k}})$ are the true (noiseless) values of the gravity anomaly signal at the data points. As usual, the sampling procedure is assumed to cover the entire (finite) support of the local gravity anomaly field $\Delta g(\bar{\mathbf{x}}) \in L^2(\mathfrak{R}^2)$. The optimal estimation of the geoid

undulation signal $N(\vec{x})$ from the gravity anomaly data will then follow the procedure shown in the linear system of Figure 5.17.

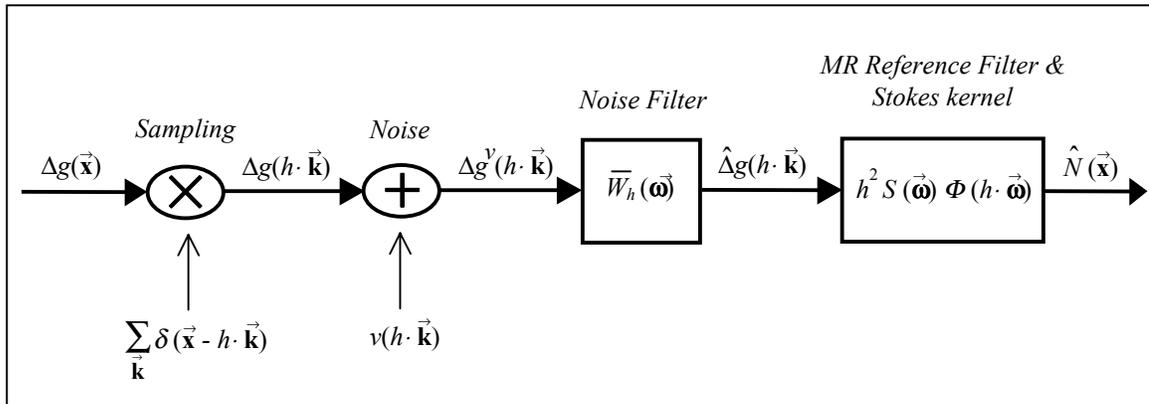


Figure 5.17 Optimal geoid estimation from gridded noisy gravity anomalies using a certain multiresolution reference filter

As seen from the above figure, the geoid estimation algorithm consists of two basic separable parts. The first step is a straightforward extension of the one-dimensional noise filtering methodology that was described in section 5.2. The optimal noise filter that is applied to the gravity anomaly data $\Delta g^y(h \cdot \vec{k})$ will have the two-dimensional Wiener-like form

$$\bar{W}_h(\vec{\omega}) = \frac{|\overline{\Delta G}_h(\vec{\omega})|^2}{|\overline{\Delta G}_h(\vec{\omega})|^2 + \bar{P}_v(\vec{\omega})} \quad (5.64)$$

where $\left| \overline{\Delta G}_h(\vec{\omega}) \right|^2$ is the periodic ‘PSD’ of the deterministic (true) gravity anomaly values (not of the total gravity anomaly signal) at the given resolution level h , $\overline{P}_v(\vec{\omega})$ is the periodic noise ‘PSD’, and $\vec{\omega}$ denotes a (2×1) real frequency vector. The quantity $\overline{\Delta G}_h(\vec{\omega})$ corresponds to the following 2D discrete Fourier transform:

$$\overline{\Delta G}_h(\vec{\omega}) = \sum_{\vec{k}} \Delta g(h \cdot \vec{k}) e^{-ih\vec{k}^T \vec{\omega}} = \frac{1}{h^2} \sum_{\vec{k}} \Delta G(\vec{\omega} + \frac{2\pi}{h} \cdot \vec{k}) \quad (5.65)$$

with $\Delta G(\vec{\omega})$ being the 2D continuous Fourier transform of the true gravity anomaly signal $\Delta g(\vec{x})$. Alternatively, we can also express the gravity anomaly ‘PSD’ function in the form

$$\left| \overline{\Delta G}_h(\vec{\omega}) \right|^2 = \sum_{\vec{k}} c_{\Delta g}(h \cdot \vec{k}) e^{-ih\vec{k}^T \vec{\omega}} \quad (5.66)$$

where $c_{\Delta g}(h \cdot \vec{k})$ is the 2D discrete spatial CV function of the gravity anomaly field $\Delta g(\vec{x})$ at a uniform resolution level h , i.e.

$$c_{\Delta g}(h \cdot \vec{k}) = \sum_{\vec{m}} \Delta g(h \cdot \vec{m}) \Delta g(h \cdot \vec{m} + h \cdot \vec{k}) \quad (5.67)$$

For the practical determination of the quantity $\left| \overline{\Delta G}_h(\vec{\omega}) \right|^2$, see the comments given in the beginning of section 5.2.5. Similarly, the data noise ‘PSD’ will be given by the 2D discrete Fourier transform

$$\overline{P}_v(\vec{\omega}) = \sum_{\vec{k}} c_v(h \cdot \vec{k}) e^{-ih\vec{k}^T \vec{\omega}} \quad (5.68)$$

where the two-dimensional discrete sequence $c_v(h \cdot \vec{k})$ is defined through the summation formula (see section 5.2.4)

$$c_v(h \cdot \vec{k}) = \sum_{\vec{m}} \sigma_v(h \cdot \vec{m}, h \cdot \vec{m} + h \cdot \vec{k}) \quad (5.69)$$

In the last equation, the symbol $\sigma_v(h \cdot \vec{m}, h \cdot \vec{m} + h \cdot \vec{k})$ denotes the covariance of the gravity anomaly noise between the 2D data points $h \cdot \vec{m}$ and $h \cdot \vec{m} + h \cdot \vec{k}$. The optimal noise filter of eq.(5.64) is multiplied by the 2D discrete Fourier transform of the data sequence $\Delta g^v(h \cdot \vec{k})$, and the result is a new ‘synthetic’ set of estimated gravity anomaly values $\hat{\Delta g}(h \cdot \vec{k})$ from which the effect of the observational noise has been minimized in a certain translation-invariant MSE sense (see section 5.2.3). The underlying linear filtering procedure can be described by the equation

$$\hat{\Delta}g(h \cdot \bar{\mathbf{k}}) = \mathfrak{S}^{-1} \left\{ \overline{W}_h(\bar{\boldsymbol{\omega}}) \overline{\Delta G}^v(\bar{\boldsymbol{\omega}}) \right\} \quad (5.70)$$

where the term $\overline{\Delta G}^v(\bar{\boldsymbol{\omega}})$ corresponds to the 2D discrete Fourier transform of the noisy data sequence, i.e.

$$\overline{\Delta G}^v(\bar{\boldsymbol{\omega}}) = \sum_{\bar{\mathbf{k}}} \Delta g^v(h \cdot \bar{\mathbf{k}}) e^{-ih\bar{\mathbf{k}}^T \bar{\boldsymbol{\omega}}} \quad (5.71)$$

and \mathfrak{S}^{-1} denotes the 2D inverse discrete Fourier transform operator.

The second step in the geoid estimation procedure utilizes the filtered gravity anomaly data $\hat{\Delta}g(h \cdot \bar{\mathbf{k}})$ to obtain a certain linear approximation $\hat{N}(\bar{\mathbf{x}})$ for the geoid signal (see Figure 5.17). The underlying translation-invariant algorithm can be expressed by the general convolution formula

$$\hat{N}(\bar{\mathbf{x}}) = s(\bar{\mathbf{x}}) * \hat{\Delta}g(\bar{\mathbf{x}}) \quad (5.72a)$$

where the estimated (referenced) input signal $\hat{\Delta}g(\bar{\mathbf{x}})$ is given in terms of a 2D multiresolution approximation model, similar to the 1D case of eq.(5.1), i.e.

$$\hat{\Delta}g(\bar{\mathbf{x}}) = \sum_{\bar{\mathbf{k}}} \hat{\Delta}g(h \cdot \bar{\mathbf{k}}) \varphi\left(\frac{1}{h} \cdot \bar{\mathbf{x}} - \bar{\mathbf{k}}\right) \quad (5.72b)$$

The function $s(\bar{\mathbf{x}})$ in eq.(5.72a) corresponds to the 2D planar Stokes kernel (Schwarz et al., 1990), whereas $\varphi(\bar{\mathbf{x}}) \in L^2(\mathfrak{R}^2)$ represents some 2D scaling interpolating kernel that is chosen to model the behaviour of the local gravity anomaly field at the given resolution level. For an optimal (in a spatio-statistical sense) determination of the reference kernel $\varphi(\bar{\mathbf{x}})$, a two-dimensional generalization of the optimization procedure given in section 4.3 should be followed. Eqs.(5.72a) and (5.72b) can also be combined into a single estimation step, as follows:

$$\hat{N}(\bar{\mathbf{x}}) = \sum_{\bar{\mathbf{k}}} \hat{\Delta}g(h \cdot \bar{\mathbf{k}}) \left[s(\bar{\mathbf{x}}) * \varphi\left(\frac{1}{h} \cdot \bar{\mathbf{x}} - \bar{\mathbf{k}}\right) \right] \quad (5.73a)$$

or, equivalently, in the frequency domain

$$\hat{N}(\bar{\omega}) = h^2 S(\bar{\omega}) \Phi(h \cdot \bar{\omega}) \sum_{\bar{\mathbf{k}}} \hat{\Delta}g(h \cdot \bar{\mathbf{k}}) e^{-ih\bar{\mathbf{k}}^T \bar{\omega}} \quad (5.73b)$$

In practice, the evaluation of the output geoid signal $\hat{N}(\bar{\mathbf{x}})$ is performed only at the points of the input gravity data grid. For this purpose, we can apply efficient FFT techniques in the numerical computation of the formulas (5.73a) and (5.73b), which obviously requires the knowledge of the 2D Fourier transform of the convolution product $s(\bar{\mathbf{x}}) * \varphi\left(\frac{1}{h} \cdot \bar{\mathbf{x}}\right)$ between the Stokes kernel and the MR approximation (referencing) kernel.

Chapter 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary – Conclusions

The discussion in this section will focus on the main conclusions that can be drawn from this thesis work. However, it should be noted that in the preceding investigations there were many interesting and key theoretical findings that have been individually identified, summarized and discussed in detail in the various sections of each chapter and, therefore, will not be repeated here.

The central scope of this research work was to address and study various problems that occur in linear signal estimation from discrete gridded data. Such problems included: solution stability, solution convergence, optimal adaptation of the estimation kernel to the data resolution level, resolution-dependent analysis of the signal interpolation error, aliasing error propagation in convolution integral formulas with discrete input data, and noise filtering. All these issues are very important in modern operational geodesy and

they affect both theoretical aspects (e.g. discrete BVPs) and practical applications (e.g. local geoid determination) in gravity field modelling and approximation.

A general linear estimation framework for dealing with the above problems has been developed, based on the use of concepts from multiresolution analysis theory. The thesis aimed at showing that the multiresolution signal analysis methods, according to Mallat's pioneering work, are not just an additional mathematical tool with a competitive role against traditional approximation techniques used in geodesy. From the findings presented in the previous chapters, it can be concluded that the MRA concept actually has a complementary (regularization) role within the signal estimation framework of deterministic collocation in Hilbert spaces (chapter 3). Furthermore, it was found to be complemented (or rather extended) by the spatio-statistical collocation principle (chapter 4).

A major portion of the thesis was devoted to discussing the classic collocation method for linear approximation in Hilbert spaces, and identifying its drawbacks for deterministic signal modelling and estimation. The use of frame theory was essential for studying the basic aspects and limitations of this method. In fact, it has expanded its applicability for more general cases than the ones usually treated in the geodetic literature (i.e. cases of linearly dependent observational representers).

The main problem that was identified in the deterministic collocation method is the use of a fixed estimation kernel, which does not take into account the actual data resolution. This affects the numerical stability of its underlying signal interpolation algorithm, as well as the overall quality of the minimum-norm signal estimate when the given data resolution is not ‘compatible’ with the implied model resolution. The multiresolution analysis theory in infinite-resolution Hilbert spaces (such as $L^2(\mathfrak{R})$), on the other hand, provides an ideal and effective tool for dealing with stability, convergence and modelling issues in linear approximation problems with gridded data of increasing resolution. It has been shown that we can actually arrive at the MRA concept through the deterministic collocation formalism by requiring unperturbed stability and convergence in the optimal solution algorithm for increasing data density, and additionally assuming a certain scale-invariance condition for the underlying signal norm.

Although the use of the multiresolution analysis concept was initially suggested from a stability and convergence point of view, in Chapter 4 we re-established its link with the linear estimation problem from a completely different viewpoint. In particular, the spatio-statistical collocation principle (with gridded data) was proven to give rise to an MRA-type interpolation algorithm, with its basic kernel being completely adapted to the data resolution level in a certain optimal fashion. This result provided a powerful extension of Mallat’s classic approximation framework, since it allows not only for the spread of the estimation kernel to be tuned to the data grid density, but also for the functional form of

the kernel itself. The accuracy improvement that is obtained by using such resolution-dependent interpolating kernels in the linear estimation framework, over the usual fixed scaling kernels of MRA theory, was also demonstrated through some numerical examples in Chapter 5, which revealed the practical significance of our theoretical developments for actual signal approximation applications in geodesy.

Another key result in this thesis was the development of a rigorous, yet simple, algorithmic procedure that can be used to measure the decay rate of the mean square interpolation error in linear multiresolution estimation models, as a function of the data resolution level and the scaling approximation kernel. The same procedure has also been used for propagating the effect of the aliasing error in convolution integral formulas with gridded input data at uniform resolution. Such resolution-dependent error modelling techniques are extremely important in various geodetic applications (e.g. gravimetric geoid determination), especially for identifying data resolution requirements for a mean accuracy level in the estimated signal. They can also be used for comparing the performance of different estimation kernels for varying data grid density, as well as for evaluating the need to incorporate an additional multiresolution reference filter for the discrete input data in the classic FFT-based geodetic computations.

Finally, the last part of the thesis presented a theoretical treatment of various issues related to the noise filtering problem within a linear multiresolution estimation model. Our developments provided a similar optimal solution with the one given in Wiener

filtering theory, without employing the usual stationarity assumption for the input data noise, but by using only a translation-invariance restriction for the signal estimation algorithm. Such a result demonstrated that the FFT-based computational techniques can be a very efficient tool in signal approximation problems even with non-stationary noisy data, if we are willing to keep the translation-invariance property for our linear signal estimate. Furthermore, our modelling choice of decomposing the total signal error in terms of resolution-dependent and noise-dependent disjoint parts resulted in the possibility of studying/optimizing individually the effects of the finite data resolution and the additive input noise to the final signal estimate, using appropriate error measures and criteria for each case. This new approach overcomes some of the traditional limitations that we face when we use other classic geodetic approximation procedures, such as Tikhonov's regularization.

6.2 Recommendations for Further Research

Considerable extensions of the research work presented herein are certainly required in order to further explore the potential of multiresolution (and wavelet-based) analysis methods for geodetic estimation problems. The present thesis has served the purpose of laying out various aspects of these methods that make their further study, in both theoretical and practical levels, justifiable and, hopefully, more interesting.

The rigorous formulation and extension for all of our theoretical developments, which were presented in Chapters 4 and 5 of this thesis, in higher dimensional settings should obviously be made. This provides a challenging problem for cases of signals with compact spherical domains that are needed in global geodetic applications. However, simpler two-dimensional planar generalizations do not introduce major complications and they can easily be employed for many local applications in gravity field modelling with real (or synthetic) data. Of particular interest is the incorporation of multiresolution reference kernels or filters, for the discrete input data, in geodetic convolution algorithms. As was suggested in more detail in previous sections of the thesis, such an estimation approach should be compared with the classical FFT-based numerical methods that are routinely used in many geodetic applications, where no specific reference model for the discrete input data is employed.

An interesting, but more mathematically oriented and considerably more difficult, topic is also to investigate the possibility of building wavelet-type bases within the generalized MRA structure that was constructed from the resolution-dependent form of the collocation scaling kernel. This would provide a very powerful ‘non-stationary’ spectral system that can be directly linked with the optimal signal estimation methodology developed in this thesis. Additional studies should also be made on the determination of closed expressions for the resolution-dependent optimal collocation filter, based on the use of different analytical models for the signal power spectrum.

The present research has only treated the special case where the input data in the linear estimation framework are gridded values of the unknown field itself. Single input-output convolution systems have been additionally considered, but only from an aliasing error propagation point of view. The existence of a generalized MRA-type behaviour in the estimated output signal of such linear systems should also be investigated, as was suggested in more detail at the end of Chapter 4. Furthermore, the stability problem of the linear estimation algorithm has been considered only from a data resolution point of view. The additional effect due to the ill-conditioned structure of various inverse convolution operators that appear in many geodetic formulas must be further regularized within the multiresolution estimation framework. Finally, the problem of optimally (and efficiently) combining heterogeneous noisy data given at different scale levels, within an integrated MRA-type estimation framework, provides the ultimate challenge in terms of geomathematical applications.

Given the continuously increasing interest of the geodetic community in multiresolution and wavelet methods and the significant achievements that have been obtained up to date, it is evident that further research on this area is warranted.

REFERENCES

- Albert, A. (1972): *Regression and the Moore-Penrose PseudoInverse*. Academic Press.
- Aldroubi, A. and Unser, M. (1992): Families of wavelet transforms in connection with Shannon's sampling theory and the Gabor transform. In "*Wavelets: A tutorial in Theory and Applications*", Chui, C.K. (ed.), Academic Press.
- Aldroubi, A. and Unser, M. (1993): Families of Multiresolution and Wavelet Spaces with Optimal Properties. *Numerical Functional Analysis and Optimization*, 14 (5,6), pp. 417-446.
- Aldroubi, A. and Unser, M. (1994): Sampling Procedures in Function Spaces and Asymptotic Equivalence with Shannon's Sampling Theory. *Numerical Functional Analysis and Optimization*, 15 (1,2), pp. 1-21.
- Alpert, B.K. (1993): A class of bases in L^2 for the sparse representation of integral operators. *SIAM Journal of Mathematical Analysis*, vol. 24, no. 1, pp. 246-262.
- Aronszajn, N. (1950): Theory of Reproducing Kernels. *Transactions of the American Mathematical Society*, vol. 68, pp. 337-404.
- Ballani, L. (1995): Solving the Inverse Gravimetric Problem: On the benefit of Wavelets. *IAG Symposia Proceedings*, vol. 114, pp. 151-161. Springer-Verlag.
- Barthelmes, F., Ballani, L. and Klees, R. (1995): On the application of Wavelets in Geodesy. *IAG Symposia Proceedings*, vol. 114, pp. 394-403. Springer-Verlag.
- Barzaghi, R. and Sanso, F. (1986): New Results on Convergence Problem in Collocation Theory. *Proceedings of the I Hotine-Marussi Symposium on Mathematical Geodesy*, Rome, Italy, June 3-6, 1985, pp. 417-457.
- Batha, L., Benciolini, B. and Zatelli, P. (1995): Geodetic applications of Wavelets: Proposals and simple numerical examples. *IAG Symposia Proceedings*, vol. 114, pp. 404-412. Springer-Verlag.
- Benciolini, B. and Zatelli, P. (1997): Wavelets applications in local geoid computations. Presented at the *IAG Scientific Assembly*, Rio de Janeiro, Brazil, Sept. 3-9, 1997.
- Bendat, J.S. and Piersol, A.G. (1986): *Random Data: Analysis and Measurement Procedures*. John Wiley & Sons, Inc.
- Bendat, J.S. and Piersol, A.G. (1993): *Engineering Applications of Correlation and Spectral Analysis*. John Wiley & Sons, Inc.

- Benedetto, J.J. (1994): Frame decompositions, sampling, and uncertainty principle inequalities. In “*Wavelets: Mathematics and Applications*”, Benedetto, J.J. and Frazier, M.W. (eds.), CRC Press, Inc.
- Benedetto, J.J. (1998): Noise reduction in Terms of the Theory of Frames. In “*Signal and Image Representation in Combined Spaces*”, Zeevi, Y. and Coifman, R. (eds.), Academic Press.
- Benedetto, J.J. and Walnut, D.F. (1994): Gabor Frames for L^2 and related spaces. In “*Wavelets: Mathematics and Applications*”, Benedetto, J.J. and Frazier, M.W. (eds.), CRC Press, Inc.
- Beylkin, G. (1992): On the representation of operators in bases of compactly supported wavelets. *SIAM Journal of Numerical Analysis*, vol. 6, no. 6, pp. 1716-1740.
- Beylkin, G., Coifman, R. and Rokhlin, V. (1991): Fast Wavelet Transforms and Numerical Algorithms I. *Communications in Pure and Applied Mathematics*, vol. 44, pp. 141-183.
- Bjerhammar, A. (1964): A New Theory of Gravimetric Geodesy. *Transactions of the Royal Institute of Technology*, no. 243, Stockholm.
- Bjerhammar, A. (1973): On the Discrete Boundary Value Problem in Physical Geodesy. *Reports of the Royal Institute of Technology, Geodetic Division*, Stockholm.
- Bjerhammar, A. (1975): Discrete approaches to the solution of the boundary value problem of physical geodesy. *Bollettino di Geodesia e Scienze Affini*, vol. 34, no. 2, pp. 185-240.
- Bjerhammar, A. (1982): On the Foundation of Collocation in Physical Geodesy. *Bulletin Geodesique*, vol. 56, pp. 312-328.
- Bjerhammar, A. (1983): Inversion-free predictors. *Reports of the Royal Institute of Technology, Geodetic Division*, Stockholm.
- Bjerhammar, A. (1987): Discrete Physical Geodesy. *Report no.380*, Dept. of Geodetic Science, Ohio State University, Columbus, Ohio.
- Blu, T. and Unser, M. (1997): Quantitative L_2 error analysis for interpolation methods and wavelet expansions. *Proceedings of the IEEE International Conference on Image Processing*, Santa Barbara, California, October 26-29, 1997, vol. I, pp. 663-666.
- Blu, T. and Unser, M. (1999a): Quantitative Fourier Analysis of Approximation Techniques: Part I – Interpolators and Projectors. *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2783-2795.

- Blu, T. and Unser, M. (1999b): Quantitative Fourier Analysis of Approximation Techniques: Part II – Wavelets. *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2796-2806.
- Blu, T. and Unser, M. (1999c): Approximation error for quasi-interpolators and (multi-) wavelet expansions. *Applied and Computational Harmonic Analysis*, vol. 6, no. 2, pp. 219-251.
- Bottoni, G.P. and Barzaghi, R. (1993): Fast collocation. *Bullettin Geodesique*, vol. 67, pp. 119-126.
- Bracewell, R.N. (1986): *The Fourier Transform and its Applications*. McGraw-Hill, Inc.
- Butzer, P.L., Splettstoßer, W. and Stens, R.L. (1988): The Sampling Theorem and Linear Prediction in Signal Analysis. *Jber. D. dt. Math. Verein.*, vol. 70, pp. 1-70.
- Christakos, G. (1992): *Random Field Models in Earth Sciences*. Academic Press.
- Cohen, L. (1989): Time-Frequency Distributions – A Review. *Proceedings of the IEEE*, vol. 77, pp. 941-981.
- Cohen, L. (1993): The Scale Representation. *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3275-3292.
- Cohen, L. (1995): *Time-Frequency Analysis*. Prentice Hall.
- Cvetkovic, Z. and Vetterli, M. (1998): Overcomplete Expansions and Robustness. In “*Signal and Image Representation in Combined Spaces*”, Zeevi, Y. and Coifman, R. (eds.), Academic Press.
- Daubechies, I. (1990): The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961-1005.
- Daubechies, I. (1992): *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA.
- Daubechies, I. (1996): Where Do Wavelets Come From ? – A Personal Point of View. *Proceedings of the IEEE*, vol. 84, no. 4, pp. 510-513.
- Daubechies, I., Grossmann, A. and Meyer, Y. (1986): Painless nonorthogonal expansions. *Journal of Mathematical Physics*, vol. 27, no. 5, pp. 1271-1283.
- Davis, P.J. (1975): *Interpolation and Approximation*. Dover Publications, Inc.
- De Boor, C., De Vore, R.A. and Ron, A. (1994): Approximation from shift-invariant subspaces of $L_2(\mathbb{R}^d)$. *Transactions of the American Mathematical Society*, vol. 341, no. 2, pp. 787-806.
- De Vore, R.A. and Lucier, B.J. (1992): Wavelets. *Acta Numerica*, vol. 1, pp. 1-56.

- Debnath, L. and Mikusinski, P. (1999): *An Introduction to Hilbert Spaces with Applications*. Academic Press.
- Dermanis, A. (1976): Probabilistic and Deterministic Aspects of Linear Estimation in Geodesy. *Report no.244*, Dept. of Geodetic Science, Ohio State University, Columbus, Ohio.
- Dermanis, A. (1977): Geodetic Linear Estimation Techniques and the Norm Choice Problem. *Manuscripta Geodaetica*, vol. 2, pp. 15-97.
- Dermanis, A. and Sanso, F. (1997): *Statistical Foundations of Geomatics*. Lecture Notes, Dept. of Geomatics Engineering, University of Calgary, Calgary, Alberta.
- Djokovic, I. and Vaidyanathan, P.P. (1997): Generalized sampling theorems in multiresolution subspaces. *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 583-599.
- Donoho, D.L. (1992): Interpolating Wavelet Transforms. *Technical Report*, Department of Statistics, Stanford University, Stanford.
- Duffin, R.J. and Schaeffer, A.C. (1952): A class of nonharmonic Fourier series. *Transactions of the American Mathematical Society*, vol. 72, pp. 341-366.
- Eeg, J. and Krarup, T. (1973): Integrated Geodesy. *Internal Report no.7*, Danish Geodetic Institute, Copenhagen.
- Eren, K. (1980): Spectral Analysis of GEOS-3 Altimeter Data and Frequency-Domain Collocation. *Report no.297*, Dept. of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio.
- Feichtinger, H.G. and Grochenig, K. (1994): Theory and practice of irregular sampling. In *“Wavelets: Mathematics and Applications”*, Benedetto, J.J. and Frazier, M.W. (eds.), CRC Press, Inc.
- Forsberg, R. (1986): Spectral Properties of the Gravity Field in the Nordic Countries. *Bollettino di Geodesia e Scienze Affini*, vol. 45, no. 4, pp. 361-383.
- Freeden, W. (1983): Interpolation and Best Approximation by Harmonic Spline Functions - Theoretical and Computational Aspects. *Proceedings of the 8th Symposium on Mathematical Geodesy (5th Hotine Symposium)*, Como, Italy, Sept. 7-9, 1981, pp. 105-121.
- Freeden, W. (1999): *Multiscale Modelling of Spaceborne Geodata*. B.G. Teubner, Stuttgart-Leipzig.
- Freeden, W. and Schneider, F. (1998): An Integrated Wavelet Concept of Physical Geodesy. *Journal of Geodesy*, vol. 72, pp. 259-281.

- Gelfand, I.M. and Shilov, G.E. (1964): *Generalized Functions*, vols. 1 and 2, Academic Press.
- Gerstl, M. and Rummel, R. (1981): Stability Investigations of Various Representations of the Gravity Field. *Reviews of Geophysics and Space Physics*, vol. 19, no. 3, pp. 415-420.
- Giacaglia, G.E.O. and Lundquist, C.A. (1972): Sampling Functions for Geophysics. *Special Report no.344, Smithsonian Astrophysics Observatory*, Washington, D.C.
- Hadamard, J. (1923): *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. Yale University Press, New Haven.
- Halmos, P.R. (1991): *Measure Theory*. Graduate Texts in Mathematics Series, vol. 18, Springer-Verlag.
- Heil, C.E. and Walnut, D.F. (1994): Continuous and Discrete Wavelet Transforms. *Siam Review*, vol. 31, no. 4, pp. 628-666.
- Heiskanen, W.A. and Moritz, H. (1967): *Physical Geodesy*. W.H. Freeman, San Francisco.
- Hernandez, E. and Weiss, G. (1996): *A First Course on Wavelets*. CRC Press, Inc.
- Higgins, J.R. (1985): Five Short Stories about the Cardinal Series. *Bulletin of the American Mathematical Society*, vol. 12, no. 1, pp. 45-89.
- Hlawatsch, F. and Boudreaux-Bartels, G.F. (1992): Linear and Quadratic Time-Frequency Signal Representations. *IEEE SP Magazine*, April 1992, pp. 21-67.
- Holschneider, M. (1995): *Wavelets, An Analysis Tool*. Oxford University Press, Inc.
- Hou, H.S. and Andrews, H.C. (1978): Cubic splines for image interpolation and digital filtering. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-26, pp. 508-517.
- Janssen, A.J.E.M. (1993): The Zak Transform and Sampling Theorems for Wavelet Subspaces. *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3360-3364.
- Jawerth, B. and Sweldens, W. (1994): An Overview of Wavelet Based Multiresolution Analyses. *Siam Review*, vol. 36, no. 3, pp. 377-412.
- Jerri, A.J. (1977): The Shannon Sampling Theorem – Its Various Extensions and Applications: A Tutorial Review. *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1565-1596.
- Kailath, T. (1974): A View of Three Decades of Linear Filtering Theory. *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 146-181.

- Kaula, W.M. (1959): Statistical and harmonic analysis of gravity. *Journal of Geophysical Research*, vol. 64, pp. 2401-2421.
- Keller, W. (1995): Harmonic Downward Continuation Using a Haar Wavelet Frame. *Proceedings of the IAG Symposium on Airborne Gravity Field Determination*, IUGG General Assembly, Boulder, Colorado, July 2-14, pp. 81-86.
- Keller, W. (1997): A wavelet-vaguelette analysis of geodetic integral formulae. *IAG Symposia Proceedings*, vol. 117, pp. 557-564. Springer-Verlag.
- Keys, R.G. (1981): Cubic Convolution Interpolation for Digital Image Processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, no. 6, pp. 1153-1160.
- Kirsch, A. (1996): *An Introduction to the Mathematical Theory of Inverse Problems*. Springer Verlag.
- Kotsakis, C. and Sideris, M.G. (1998): Study of the Gravity Field Spectrum in Canada in View of cm-Geoid Determination. Presented at the 2nd Joint Meeting of the International Gravity Commission and the International Geoid Commission, Trieste, Italy, Sept. 7-12, 1998.
- Kotsakis, C. and Sideris, M.G. (1999): The High-Frequency Structure of the Gravity Field in Canada. Presented at the 25th Annual Meeting of Canadian Geophysical Union, Banff, Canada, May 9-13, 1999.
- Krarp, T. (1969): A Contribution to the Mathematical Foundations of Physical Geodesy. *Report of the Danish Geodetic Institute*, no. 44, Copenhagen.
- Krarp, T. (1978): Some Remarks about Collocation. In "Approximation Methods in Geodesy", Moritz, H. and Sunkel, H. (eds.), Herbert Wichmann Verlag, Karlsruhe.
- Leglemann, D. (1979): Analytical Collocation with Kernel Functions. *Bullettin Geodesique*, vol. 53, pp. 273-289.
- Li, Z. (1996a): Multiresolution approximation of the gravity field. *Journal of Geodesy*, vol. 70, pp. 731-739.
- Li, Z. (1996b): Multiresolution Approximation in Gravity Field Modelling. *UCGE Report no. 20103*, Dept. of Geomatics Engineering, University of Calgary, Calgary, Alberta.
- Li, J. (1996c): Detailed Marine Gravity Field Determination by Combination of Heterogeneous Data. *UCGE Report no. 20102*, Dept. of Geomatics Engineering, University of Calgary, Calgary, Alberta.
- Li, J. and Sideris, M.G. (1997): Marine gravity and geoid determination by optimal combination of satellite altimetry and shipborne gravimetry data. *Journal of Geodesy*, vol. 71, pp. 209-216.

- Mallat, S.G. (1989a): A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693.
- Mallat, S.G. (1989b): Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\mathbb{R})$. *Transactions of the American Mathematical Society*, vol. 315, no. 1, pp. 69-87.
- Mallat, S.G. (1998a): Applied Mathematics Meets Signal Processing. *Documenta Mathematica*, Extra Volume ICM (Proceedings of the International Congress of Mathematicians, Berlin, 1998), pp. 1-18.
- Mallat, S.G. (1998b): *A Wavelet Tour of Signal Processing*. Academic Press.
- Mallat, S. and Hwang, W.L. (1992): Singularity Detection and Processing with Wavelets. *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617-643.
- Meissl, P. (1976): Hilbert Spaces and their Applications to Geodetic Least-Squares Problems. *Bolletino di Geodesia e Scienze Affini*, vol. 35, no. 1, pp. 181-210.
- Mertins, A. (1996): *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications*. Wiley, New York.
- Moritz, H. (1962): Interpolation and Prediction of Gravity and their Accuracy. *Report no.24*, Institute of Geodesy, Photogrammetry and Cartography, Ohio State University, Columbus, Ohio.
- Moritz, H. (1970): Least-Squares Estimation in Physical Geodesy. *Report no.130*, Dept. of Geodetic Science, Ohio State University, Columbus, Ohio.
- Moritz, H. (1976a): Integral Formulas and Collocation. *Manuscripta Geodaetica*, vol. 1, pp. 1-40.
- Moritz, H. (1976b): Least-Squares Collocation as a Gravitational Inverse Problem. *Report no.249*, Dept. of Geodetic Science, Ohio State University, Columbus, Ohio.
- Moritz, H. (1978a): Statistical Foundations of Collocation. *Report no.272*, Dept. of Geodetic Science, Ohio State University, Columbus, Ohio.
- Moritz, H. (1978b): Introduction to Interpolation and Approximation. In “*Approximation Methods in Geodesy*”, Moritz, H. and Sunkel, H. (eds.), Herbert Wichmann Verlag, Karlsruhe.
- Moritz, H. (1980): *Advanced Physical Geodesy*. Herbert Wichmann Verlag, Karlsruhe.
- Moritz, H. and Sanso, F. (1980): A Dialogue on Collocation. *Bolletino di Geodesia e Scienze Affini*, vol. 39, no. 1, pp. 49-51.

- Nash, R.A. and Jordan, S.K. (1978): Statistical Geodesy: An Engineering Perspective. *Proceedings of the IEEE*, vol. 66, no. 5, pp. 532-550.
- Nashed, M.Z. (1976): Aspects of Generalized Inverses in Analysis and Regularization. In "Generalized Inverses and Applications", Nashed, M.A. (ed.), New York.
- Nashed, M.Z. and Walter, G.G. (1991): General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces. *Mathematics of Control, Signals and Systems*, 4, pp. 363-390.
- Naylor, A.W. and Sell, G.R. (1982): *Linear Operator Theory in Engineering and Science*. Springer Verlag.
- Oppenheim, A.V. and Schafer R.W. (1989): *Discrete-Time Signal Processing*. Prentice Hall.
- Papoulis, A. (1977): Generalized sampling expansions. *IEEE Transactions on Circuits and Systems*, vol. 24, no. 11, pp. 652-654.
- Papoulis, A. (1991): *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc.
- Park, S.K. and Showengerdt, R.A. (1983): Image reconstruction by parametric convolution. *Computer Vision, Graphics and Image Processing*, vol. 20, no. 3, pp. 258-272.
- Parker, J.A., Kenyon, R.V. and Troxel, D.E. (1983): Comparison of interpolating methods for image resampling. *IEEE Transactions on Medical Imaging*, vol. MI-2, no. 1, pp. 31-39.
- Parzen, E. (1967): *Time Series Analysis Papers*. Holden-Day.
- Phillips, G.M. and Taylor, P.J. (1996): *Theory and Applications of Numerical Analysis*. Academic Press.
- Pratt, W.K. (1978): *Digital Image Processing*. Wiley, New York.
- Priestley, M.B. (1981): *Spectral Analysis and Time Series*. Vol. 1&2, Academic Press.
- Rao, C.R. and Mitra, S.K. (1971): *Generalized Inverses of Matrices and their Applications*. Wiley, New York.
- Rapp, R.H. (1978): Results of the Application of Least-Squares Collocation to Selected Geodetic Problems. In "Approximation Methods in Geodesy", Moritz, H. and Sunkel, H. (eds.), Herbert Wichmann Verlag, Karlsruhe.
- Rioul, O. and Vetterli, M. (1991): Wavelets and Signal Processing. *IEEE SP Magazine*, October 1991, pp. 14-38.

- Rummel, R., Schwarz, K.P. and Gerstl, M. (1979): Least-Squares Collocation and Regularization. *Bulletin Geodesique*, vol. 53, pp. 343-361.
- Saito, N. and Beylkin, G. (1993): Multiresolution Representations Using the Auto-Correlation Functions of Compactly Supported Wavelets. *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3584-3590.
- Sanso, F. (1978): The Minimum Mean Square Estimation Error Principle in Physical Geodesy (Stochastic and Non-Stochastic Interpretation). Presented at the 7th *Symposium on Mathematical Geodesy (4th Hotine Symposium)*, Assisi, Italy, June 8-10. (also in *Bollettino di Geodesia e Scienze Affini*, vol. 39, no. 2, pp. 112-129, 1980)
- Sanso, F. (1986): Statistical Methods in Physical Geodesy. In “*Mathematical and Numerical Techniques in Physical Geodesy*”, Sunkel, H. (ed.), Springer Verlag.
- Sanso, F. (1987): Talk on the Theoretical Foundations of Physical Geodesy. *IAG Section IV Report “Contributions to Geodetic Theory and Methodology”*, IUGG General Assembly, Vancouver B.C., Canada, August 9-22, 1987, pp. 5-27.
- Sanso, F. and Tscherning, C.C. (1980): Notes on Convergence in Collocation Theory. *Bollettino di Geodesia e Scienze Affini*, vol. 39, no. 3, pp. 123-134.
- Sanso, F. and Schuh, W.D. (1987): Finite Covariance Functions. *Bullettin Geodesique*, 61, pp. 331-347.
- Sanso, F. and Sideris, M.G. (1997): On the similarities and differences between systems theory and least-squares collocation in physical geodesy. *Bollettino di Geodesia e Scienze Affini*, vol. 54, no. 2, pp. 173-206.
- Sanso, F. and Sona, G. (1995): The Theory of Optimal Linear Estimation for Continuous Fields of Measurements. *Manuscripta Geodaetica*, vol. 20, pp. 204-230.
- Schmidt, H.F. (1981): Sampling Function and Finite Element Method Representation of the Gravity Field. *Reviews of Geophysics*, vol. 19, no. 3, pp. 421-436.
- Schwarz, K.P. (1979): Geodetic improperly-posed problems and their regularization. *Bollettino di Geodesia e Scienze Affini*, vol. 38, no. 3, pp. 389-416.
- Schwarz, K.P. (1984): Data types and their spectral properties. *Proceedings of the International Summer School on Local Gravity Field Approximation*, Beijing, China, Aug. 21-Sept. 4, 1984, pp. 1-67.
- Schwarz, K.P., Sideris, M.G. and Forsberg, R. (1990): The use of FFT techniques in Physical Geodesy. *Geophysical Journal International*, vol. 100, pp. 485-514.
- Sideris, M.G. (1995): On the use of heterogeneous noisy data in spectral gravity field modeling methods. *Journal of Geodesy*, vol. 70, pp. 470-479.

- Sjoberg, L. (1978): A Comparison of Bjerhammar's Methods and Collocation in Physical Geodesy. *Report no.273*, Dept. of Geodetic Science, Ohio State University, Columbus, Ohio.
- Sobolev, S.L. (1964): *Partial Differential Equations of Mathematical Physics*. Pergamon Press, London.
- Strang, G. (1988): *Linear Algebra and its Applications*. Saunders College Publishing.
- Strang, G. (1989): Wavelets and Dilation Equations: A Brief Introduction. *Siam Review*, vol. 31, no. 4, pp. 614-627.
- Strang, G. and Fix, G. (1971): A Fourier analysis of the finite element variational method. In "*Constructive Aspect of Functional Analysis*", Edizioni Cremonese, Rome, Italy, pp. 796-830.
- Strohmer, T. (1995): *Irregular Sampling, Frames and Pseudoinverse*. Ph.D. Dissertation, Dept. of Mathematics, University of Vienna.
- Sunkel, H. (1981): Cardinal Interpolation. *Report no.312*, Dept. of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio.
- Sunkel, H. (1984): Splines: their equivalence to collocation. *Report no.353*, Dept. of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio.
- Svensson, S.L. (1983): A New Geodesy Based upon Inversion-free Bjerhammar Predictors. *Proceedings of the 8th Symposium on Mathematical Geodesy (5th Hotine Symposium)*, Como, Italy, Sept. 7-9, 1981, pp. 449-468.
- Sweldens, W. (1996): Wavelets: What Next? *Proceedings of the IEEE*, vol. 84, no. 4, pp. 665-670.
- Thomas, S.W. and Heller, W.G. (1976): Efficient Estimation Techniques for Integrated Gravity Data Processing. *Technical Report AFGL-TR-76-0232*. The Analytical Sciences Corporation. Reading, Massachusetts.
- Tscherning, C.C. (1977): A note on the choice of the norm when using collocation for the computations of approximations to the anomalous potential. *Bulletin Geodesique*, vol. 51, pp. 137-147.
- Tscherning, C.C. (1978a): On the Convergence of Least Squares Collocation. *Bolletino di Geodesia e Scienze Affini*, vol. 37, no. 2-3, pp. 507-517.
- Tscherning, C.C. (1978b): Introduction to Functional Analysis with a View to Its Applications in Approximation Theory. In "*Approximation Methods in Geodesy*", Moritz, H. and Sunkel, H. (eds.), Herbert Wichmann Verlag, Karlsruhe.

- Tscherning, C.C. (1986): Functional Methods for Gravity Field Approximation. In “*Mathematical and Numerical Techniques in Physical Geodesy*”, Sunkel, H. (ed.), Springer Verlag.
- Tziavos, I.N., Li, J. and Sideris, M.G. (1996): Marine gravity field modelling using non-isotropic a-priori information. *IAG Symposia Proceedings*, vol. 117, pp. 400-408. Springer-Verlag.
- Unser, M. (1997): Ten Good Reasons for Using Spline Wavelets. *Proceedings of SPIE, Wavelet Applications in Signal and Image Processing V*, vol. 3169, pp. 422-431.
- Unser, M. (1999): Splines: A perfect fit for signal/image processing. *IEEE SP Magazine*. vol. 16, no. 6, pp. 22-38.
- Unser, M. (2000): Sampling – 50 years after Shannon. *Proceedings of the IEEE*. (to appear)
- Unser, M., Aldroubi, A. and Eden, M. (1991): Fast B-Spline transforms for continuous image representation and interpolation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 277-285.
- Unser, M. and Aldroubi, A. (1992): Polynomial Splines and Wavelets – a signal processing perspective. In “*Wavelets: A tutorial in Theory and Applications*”, Chui, C.K. (ed.), pp. 543-601, Academic Press.
- Unser, M., Aldroubi, A. and Eden, M. (1992): Polynomial Spline Signal Approximations: Filter Design and Asymptotic Equivalence with Shannon’s Sampling Theorem. *IEEE Transactions on Information Theory*, vol. 38, no. 1, pp. 95-103.
- Unser, M. and Daubechies, I. (1997): On the Approximation Power of Convolution-Based Least Squares versus Interpolation. *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 1697-1711.
- Unser, M. and Zerubia, J. (1997): Generalized sampling: Stability and performance analysis. *IEEE Transactions on Signal Processing*, vol. 45, no. 12, pp. 2941-2950.
- Unser, M. and Zerubia, J. (1998): A generalized sampling theory without bandlimiting constraints. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, vol. 45, no. 8, pp. 959-969.
- Walter, G.G. (1992): A Sampling Theorem for Wavelet Subspaces. *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 881-884.
- Walter, G.G. (1993): Approximation of the Delta Function by Wavelets. *Journal of Approximation Theory*, vol. 71, pp. 329-343.

- Walter, G.G. (1994): *Wavelets and Other Orthogonal Systems with Applications*. CRC Press, Inc.
- Wojtaszczyk, P. (1997): *A Mathematical Introduction to Wavelets*. Cambridge University Press.
- Xia, X-G. and Zhang, Z. (1993): On Sampling Theorem, Wavelets, and Wavelet Transforms. *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3524-3535.
- Yao, K. (1967): Applications of reproducing kernel Hilbert spaces – bandlimited signal models. *Information and Control*, vol. 11, pp. 429-444.
- Young, R.M. (1980): *An Introduction to Non-Harmonic Fourier Series*. Academic Press.
- Zayed, A.I. (1993): *Advances in Shannon's Sampling Theory*. CRC Press, Inc.

APPENDIX A

In this appendix we will prove the following equation (see also eq.(4.14), page 137):

$$\int_{-h/2}^{h/2} |E(\omega, x_o)|^2 dx_o = hC(\omega) - \Phi_h^*(\omega)C(\omega) - \Phi_h(\omega)C(\omega) + \Phi_h(\omega)\Phi_h^*(\omega)\bar{C}_h(\omega) \quad (\text{A.1})$$

where the asterisk * denotes the complex conjugate, $\Phi_h(\omega)$ is the Fourier transform of the approximation kernel $\varphi_h(x)$ at data resolution level h , and $C(\omega)$ is the Fourier transform of the spatial CV function $c(x)$ of the unknown signal, i.e.

$$c(x) = \int g(y) g(y+x) dy \quad (\text{A.2})$$

The function $\bar{C}_h(\omega)$ has the following periodic form:

$$\bar{C}_h(\omega) = \frac{1}{h} \sum_k C(\omega + \frac{2k\pi}{h}) \quad (\text{A.3})$$

Taking into account eqs.(4.11) and (4.12) from Chapter 4, the error power spectrum at an arbitrary value of the sampling phase x_o has the following form:

$$\begin{aligned}
|E(\omega, x_o)|^2 &= G(\omega)G^*(\omega) - \Phi_h^*(\omega)G(\omega)A^*(\omega, x_o) - \\
&\quad - \Phi_h(\omega)G^*(\omega)A(\omega, x_o) + \Phi_h(\omega)\Phi_h^*(\omega)A(\omega, x_o)A^*(\omega, x_o)
\end{aligned} \tag{A.4}$$

where the auxiliary function $A(\omega, x_o)$ is given by the formula

$$A(\omega, x_o) = \frac{1}{h} \sum_k G(\omega + \frac{2\pi k}{h}) e^{-i\frac{2\pi k}{h}x_o} \tag{A.5}$$

Integrating equation (A.4) over x_o , we get analytically for every term

$$\int_{-h/2}^{h/2} G(\omega)G^*(\omega) dx_o = h |G(\omega)|^2 = h C(\omega) \tag{A.6}$$

$$\begin{aligned}
&\int_{-h/2}^{h/2} \Phi_h^*(\omega) G(\omega) A^*(\omega, x_o) dx_o = \\
&= \Phi_h^*(\omega) G(\omega) \int_{-h/2}^{h/2} \frac{1}{h} \sum_k G^*(\omega + \frac{2\pi k}{h}) e^{i\frac{2\pi k}{h}x_o} dx_o \\
&= \frac{1}{h} \Phi_h^*(\omega) G(\omega) \sum_k G^*(\omega + \frac{2\pi k}{h}) \int_{-h/2}^{h/2} e^{i\frac{2\pi k}{h}x_o} dx_o \\
&= \frac{1}{h} \Phi_h^*(\omega) G(\omega) \sum_k G^*(\omega + \frac{2\pi k}{h}) \int_{-\pi}^{\pi} \frac{h}{2\pi} e^{ik\xi} d\xi \\
&= \Phi_h^*(\omega) G(\omega) \sum_k G^*(\omega + \frac{2\pi k}{h}) \frac{\sin k\pi}{k\pi} \\
&= \Phi_h^*(\omega) G(\omega) G^*(\omega) \\
&= \Phi_h^*(\omega) C(\omega)
\end{aligned} \tag{A.7}$$

Following similar derivations as in eq.(A.7), we also have that

$$\int_{-h/2}^{h/2} \Phi_h(\omega) G^*(\omega) A(\omega, x_o) dx_o = \Phi_h(\omega) C(\omega) \quad (\text{A.8})$$

Finally, the integration of the last term in equation (A.4) yields

$$\begin{aligned} & \int_{-h/2}^{h/2} \Phi_h(\omega) \Phi_h^*(\omega) A(\omega, x_o) A^*(\omega, x_o) dx_o = \\ & = \Phi_h(\omega) \Phi_h^*(\omega) \int_{-h/2}^{h/2} \frac{1}{h^2} \sum_n \sum_m G(\omega + \frac{2\pi n}{h}) G^*(\omega + \frac{2\pi m}{h}) e^{i \frac{2\pi(m-n)}{h} x_o} dx_o \\ & = \frac{1}{h^2} \Phi_h(\omega) \Phi_h^*(\omega) \sum_n \sum_m G(\omega + \frac{2\pi n}{h}) G^*(\omega + \frac{2\pi m}{h}) \int_{-h/2}^{h/2} e^{i \frac{2\pi(m-n)}{h} x_o} dx_o \\ & = \frac{1}{h^2} \Phi_h(\omega) \Phi_h^*(\omega) \sum_n \sum_m G(\omega + \frac{2\pi n}{h}) G^*(\omega + \frac{2\pi m}{h}) \int_{-\pi}^{\pi} \frac{h}{2\pi} e^{i(m-n)\xi} d\xi \\ & = \frac{1}{h} \Phi_h(\omega) \Phi_h^*(\omega) \sum_n \sum_m G(\omega + \frac{2\pi n}{h}) G^*(\omega + \frac{2\pi m}{h}) \frac{\sin \pi(m-n)}{\pi(m-n)} \\ & = \frac{1}{h} \Phi_h(\omega) \Phi_h^*(\omega) \sum_k G(\omega + \frac{2\pi k}{h}) G^*(\omega + \frac{2\pi k}{h}) \\ & = \frac{1}{h} \Phi_h(\omega) \Phi_h^*(\omega) \sum_k C(\omega + \frac{2\pi k}{h}) \\ & = \Phi_h(\omega) \Phi_h^*(\omega) \bar{C}_h(\omega) \end{aligned} \quad (\text{A.9})$$

Combining together the results from equations (A.6) - (A.9), we finally get the initially claimed statement of equation (A.1).

APPENDIX B

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

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

Each element $V_j \subset L^2(\mathfrak{R})$ of this subspace sequence is defined as the closed linear span of the set $\varphi(\frac{x}{h_j} - n, h_j)_{n \in \mathbb{Z}}$, where the kernel $\varphi(x, h_j)$ is defined by eq.(4.23a) and the scaling parameter h_j associated with each subspace V_j is assumed to satisfy the two general conditions given in eqs.(4.33) and (4.35). Furthermore, the power spectrum and the CV function of the underlying unknown field are assumed to satisfy all the conditions given for them in section 4.6.1. Every signal $f_j(x) \in V_j$ will have the following general form:

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

where $\{b_n\}_{n \in \mathbb{Z}}$ is a certain square-summable sequence of coefficients. Taking into account eq.(4.23b), the last equation can be equivalently expressed in the frequency domain as follows:

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

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

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

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

$$\begin{aligned}
F_j(\omega) &= h_j \frac{C(\omega)}{\sum_k C(\omega + \frac{2\pi k}{h_j})} \frac{h_{j+1} \sum_k C(\omega + \frac{2\pi k}{h_{j+1}})}{h_{j+1} \sum_k C(\omega + \frac{2\pi k}{h_{j+1}})} \bar{B}_{2\pi/h_j}(\omega) \\
&= h_{j+1} \frac{C(\omega)}{\sum_k C(\omega + \frac{2\pi k}{h_{j+1}})} \frac{h_j \sum_k C(\omega + \frac{2\pi k}{h_{j+1}})}{h_{j+1} \sum_k C(\omega + \frac{2\pi k}{h_j})} \bar{B}_{2\pi/h_j}(\omega) \\
&= h_{j+1} \frac{C(\omega)}{\sum_k C(\omega + \frac{2\pi k}{h_{j+1}})} \bar{\Lambda}_{2\pi/h_{j+1}}(\omega) \bar{B}_{2\pi/h_j}(\omega) \\
&= h_{j+1} \frac{C(\omega)}{\sum_k C(\omega + \frac{2\pi k}{h_{j+1}})} \bar{N}_{2\pi/h_{j+1}}(\omega) , \quad \forall f_j(x) \in V_j
\end{aligned} \tag{B.5}$$

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

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

Obviously, the above function is $(2\pi/h_{j+1})$ -periodic, since the two scaling parameters (h_j and h_{j+1}) are assumed to be related through a positive *integer* number, according to condition VI in eq.(4.35). For the same reason, the product of the two periodic functions

$\bar{A}_{2\pi/h_{j+1}}(\omega)$ and $\bar{B}_{2\pi/h_j}(\omega)$, which is denoted by $\bar{N}_{2\pi/h_{j+1}}(\omega)$ in eq.(B.5), will also be a $(2\pi/h_{j+1})$ -periodic function.

Furthermore, under conditions I and II [see eqs.(4.28) and (4.30)], 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 $\bar{A}_{2\pi/h_{j+1}}(\omega)$ will have a finite $L^2(0, 2\pi/h_{j+1})$ norm. As a result, the auxiliary periodic function $\bar{N}_{2\pi/h_{j+1}}(\omega)$ in eq.(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 into the higher resolution subspace V_{j+1} , according to the general formula given in eq.(B.4), and the nesting property therefore has been established.

APPENDIX C

In this appendix we will prove the fundamental integral formula that gives the mean error ‘variance’ $\sigma^2(h)$ in linear MR signal approximation models of the form of eq.(5.1), as a function of the data resolution level h , i.e.

$$\sigma^2(h) = \frac{1}{2\pi} \int C(\omega) K(h\omega) d\omega \quad (\text{C.1})$$

where $C(\omega)$ denotes the power spectrum of the unknown signal, and $K(\omega)$ corresponds to the auxiliary frequency-domain kernel

$$K(\omega) = 1 - 2|\Phi(\omega)| + \sum_k |\Phi(\omega + 2\pi k)|^2 \quad (\text{C.2})$$

with $\Phi(\omega)$ being the Fourier transform of the approximation scaling kernel $\varphi(x)$. The latter will be assumed to correspond to a *symmetric* function.

According to eq.(5.15) in Chapter 5, the mean error variance is defined as the integral of the mean (spatio-statistical) error power spectrum over the whole real line, i.e.

$$\sigma^2(h) = \frac{1}{2\pi} \int P_e(\omega, h) d\omega \quad (\text{C.3})$$

where the mean error power spectrum is given by the general formula (see section 5.1.1)

$$P_e(\omega, h) = C(\omega) - C(\omega)\Phi(h\omega) - C(\omega)\Phi^*(h\omega) + h\bar{C}_h(\omega) |\Phi(h\omega)|^2 \quad (\text{C.4})$$

In order to simplify eq.(C.4), we should express the basic MR approximation filter in the following polar complex form:

$$\Phi(\omega) = |\Phi(\omega)| e^{i\angle\Phi(\omega)} \quad (\text{C.5})$$

where $\angle\Phi(\omega)$ denotes the phase of the Fourier transform $\Phi(\omega)$. According to Euler's formula, we have that

$$\begin{aligned} \Phi(\omega) + \Phi^*(\omega) &= |\Phi(\omega)| \left(e^{i\angle\Phi(\omega)} + e^{-i\angle\Phi(\omega)} \right) \\ &= 2 |\Phi(\omega)| \cos(\angle\Phi(\omega)) \end{aligned} \quad (\text{C.6})$$

In this way, the mean error power spectrum in eq.(C.4) takes the form

$$P_e(\omega, h) = C(\omega) \left[1 - 2 |\Phi(h\omega)| \cos(\angle\Phi(h\omega)) \right] + h\bar{C}_h(\omega) |\Phi(h\omega)|^2 \quad (\text{C.7})$$

and under the assumption that the approximation filter $\Phi(\omega)$ corresponds to a symmetric space-domain kernel $\varphi(x)$, we finally have

$$P_e(\omega, h) = C(\omega) [1 - 2|\Phi(h\omega)|] + h \bar{C}_h(\omega) |\Phi(h\omega)|^2 \quad (\text{C.8})$$

Hence, the integral formula for the mean error variance in eq.(C.3) takes the analytical form

$$\sigma^2(h) = \frac{1}{2\pi} \int C(\omega) [1 - 2|\Phi(h\omega)|] d\omega + \frac{1}{2\pi} \int h \bar{C}_h(\omega) |\Phi(h\omega)|^2 d\omega \quad (\text{C.9})$$

The second integral at the right-hand side of the last equation can be further modified, taking into account eq.(4.16) from Chapter 4, as follows:

$$\begin{aligned} \int h \bar{C}_h(\omega) |\Phi(h\omega)|^2 d\omega &= \int h \frac{1}{h} \sum_k C(\omega + \frac{2k\pi}{h}) |\Phi(h\omega)|^2 d\omega \\ &= \int_{-\pi/h}^{\pi/h} \sum_n |\Phi(h\omega + 2n\pi)|^2 \sum_k C(\omega + \frac{2k\pi}{h}) d\omega \quad (\text{C.10}) \\ &= \int C(\omega) \sum_n |\Phi(h\omega + 2n\pi)|^2 d\omega \end{aligned}$$

If we substitute eq.(C.10) into eq.(C.9), we can finally obtain the initially claimed integral formula for the mean error variance, i.e.

$$\begin{aligned}
\sigma^2(h) &= \frac{1}{2\pi} \int C(\omega) [1 - 2|\Phi(h\omega)|] d\omega + \frac{1}{2\pi} \int C(\omega) \sum_k |\Phi(h\omega + 2k\pi)|^2 d\omega \\
&= \frac{1}{2\pi} \int C(\omega) \left[1 - 2|\Phi(h\omega)| + \sum_k |\Phi(h\omega + 2\pi k)|^2 \right] d\omega \quad (\text{C.11}) \\
&= \frac{1}{2\pi} \int C(\omega) K(h\omega) d\omega
\end{aligned}$$

where the auxiliary frequency-domain kernel $K(\omega)$ is given by eq.(C.2).