Orthogonal Functions Over the Oceans and Applications to the Determination of Orbit Error, Geoid and Sea Surface Topography from Satellite Altimetry

by

Cheinway Hwang

Report No. 414

Department of Geodetic Science and Surveying
The Ohio State University
Columbus, Ohio 43210-1247

December 1991

Abstract

The goal of this study is to look for a set of orthogonal functions over the oceans and then to apply the functions to the expansions of oceanic signals. Ultimately these functions are incorporated in the parameter estimation problem using a model that simultaneously reduces satellite radial orbit errors, improves the geoid and estimates the sea surface topography.

To construct a possible set of orthogonal functions (over the oceans), three methods have been studied. In one method, an attempt was made to solve the spherical Helmholtz equation over the oceans with either the Dirichlet boundary condition or the Newmann Boundary condition or the mixed boundary condition. This method leads to successful developments of the spherical cap harmonics and generalized Fourier-Bessel series for some regularized oceanic domains. The other study employs the Shwarz-Christoffel conformal mapping, as well as an auxiliary mapping, to transform the irregular domain (the oceans) onto the interior of a unit disk where a set of orthogonal functions are easy to find. The set of orthogonal functions over the oceans are then found through the relationship between the two domains implied by the mappings. The third method involves the Gram-Schmidt process for which a new technique for computing the integrals of the products of two associated Legendre functions is developed and a FFT method is used to compute the inner products of spherical harmonics over the oceans. Also, for the entire unit sphere, the generalized 2-D Fourier series and the generalized Fourier-Tschebycheff series are proposed as alternatives for the spherical harmonics.

The expansions of the Levitus SST into the orthonormal functions constructed by the Gram-Schmidt process show that 98.5% of the energy of that signal is contained within degree 10 of the orthonormal functions. Such expansions also render regular spectral behavior of oceanic signal as compared to that from spherical harmonic expansions. A method of detecting band limited oceanic signal is also developed using these orthonormal functions. The applications of these functions to the simultaneous model yield lower correlations between expansion coefficients and show that the cut-off frequency (the highest degree determinable) of the SST from the Geosat data is in the vicinity of degree 15 of the functions. Some other improved simultaneous models have also been tested in an attempt to better separate the geoid and the SST signal. This study concludes that the orthonormal functions are suitable for representing oceanic signal with excellent spectral behavior and can be applied to future altimetric mission such as TOPEX/POSEIDON.

Foreword

This report was prepared by Cheinway Hwang, Graduate Research Associate, Department of Geodetic Science and Surveying at The Ohio State University. The research described in this report was supported under NASA'S TOPEX Altimeter Research in Ocean Circulation Mission and funded through the Jet Propulsion Laboratory under contract 958121, OSURF 720426.

Computer resources were provided by the Ohio State Academic Computing Services, and the Ohio Supercomputer Center through Cray Grant pas 160.

A slightly modified version of this report was submitted to the Graduate School of The Ohio State University in partial fulfillment of the requirements for the Degree Doctor of Philosophy.

The reproduction and distribution of this report was carried out, in part, with funds supplied by the Department of Geodetic Science and Surveying.

Acknowledgements

I am extremely grateful to my advisor, Dr. Richard H. Rapp, without whose guidance and encouragement this work would not have been possible. The knowledge and research experience that I gained from him are invaluable to my future geodetic career.

In addition to Dr. Rapp, thanks are also due to Dr. C. Goad and Dr. B. Schaffrin for their constructive comments on my work as members of the reading committee.

Special thanks go to Dr. Y. Wang and Dr. N. Pavlis for their helpful information on some existing software, to Dr. A. Mainville for providing me software, and to C. Wagner for his constructive suggestion in satellite orbit error work.

The clarifications on some mathematical issues by Dr. P. Nevai in the Dept. of Mathematics of The Ohio State University are greatly appreciated. Conversations with mathematicians T. Koornwinder, A. Magnus and G. Gasper via electronic mail were very helpful in understanding some of the important mathematical problems.

Last, but not least, I am indebted to Ms. Melanie Hennell and to Mrs. Greta Sell for their excellent typing of the draft and their friendly cooperation in making the necessary corrections.

Table of Contents

Abstract ii
Foreword iii
Acknowledgments iv
List of Tables viii
List of Figures x

Chapter Page

I. Introduction 1

II. Expansion in Series of Orthogonal Functions 2

2.1 The Weierstrass Approximation Theorem and Uniform Approximation 3
2.2 The Inner Product Space and Least Squares Approximation 3
2.3 Orthogonal Functions 5
2.4 Some Properties of Orthogonal Function Expansions 7

III. Methods of Constructing Orthonormal Functions Over the Oceans 9

3.1 The Definitions of the Oceans 9
3.2 The Gram-Schmidt Orthonormalizing Process 10
3.2.1 The Principle 10
3.2.2 The Combination Coefficients in the Orthonormalizing Process 12
3.2.3 Relationship with QR Factorization 16
3.3 Method of Eigenvalue-Eigenfunction 17
3.3.1 The Helmholtz Equation and Its Solution 17
3.3.2 Eigenfunctions Orthogonal on a Portion of the Unit Sphere 24
3.3.2.1 Hypergeometric Functions and Legendre's Functions 24
3.3.2.2 Spherical Cap Harmonics 30
3.3.2.3 Generalized Fourier-Bessel Series on the Unit Sphere 38
3.3.3 Frequency Classification of Eigenfunctions 42
3.4 Method of Conformal Mapping for Constructing Complex Orthogonal Functions 46
3.4.1 Transformation Between Two Domains and Finding the Orthonormal Functions 46
3.4.2 Schwarz-Christoffel Transformation Between the Oceans and the Unit Disk 51
3.5 Some Comments on the Proposed Methods 56

IV. Functions Orthogonal on an Entire Sphere 58

4.1 Spherical Harmonics and 2-D Fourier Series 58
4.2 Are the Spherical Harmonics the Unique Set of Orthogonal Functions on an Entire Sphere? 65

V. Gram-Schmidt Orthonormalizing Process Over the Oceans Using Spherical Harmonics 72

5.1 The Inner Products of Spherical Harmonics over the Oceans 73
5.1.1 The Complete Formulae of Inner Products 73
5.1.2 Recursive Formulae for Integrating Products of Two Associated Legendre Functions 78
5.1.3 Product-Sum Formulae for Integrating Products of Two Associated Legendre Functions 80
5.2 Construction of Orthonormal Functions over the Oceans 86
5.3 Choices of Orthonormal Systems and Unitary Transformation 93
5.4 Some Properties of the ON Functions Constructed by the Gram-Schmidt Process 97

VI. Expansions of Sea Surface Topography and Oceanic Geoid in Orthonormal Functions 100

6.1 The SST Data Used in the Expansions and Some Definitions 100
6.2 Expansion Methods 102
6.2.1 Numerical Quadratures Formula 102
6.2.2 Least Squares Fit (ℓsf) 107
6.2.3 Correlation Analysis of Two Signal Components in the ON Function Expansion 116
6.2.4 Harmonic Synthesis 118
6.3 Spherical Harmonic Expansion and ON Function Expansion 120
6.4 Numerical Experiments and Results Using the Levitus SST in the Expansions 125
6.4.1 Some Definitions 125
6.4.2 Results 126
6.4.3 Geostrophic Currents at the Continental Boundaries 132
6.5 Expansions of Oceanic Geoid in ON Functions 134

VII. Applications of Orthonormal Functions to the Simultaneous Radial Orbit Error Reduction and Geoid-SST Estimation 149

7.1 Linearized Lagrange's Equations of Motion and Radial Orbit Error 149
7.2 Mathematical Models for the Simultaneous Radial Orbit Error Reduction and Geoid-SST Estimation 158
7.2.1 Some Existing Problems in the Simultaneous Solution 158
7.2.2 Models for the Simultaneous Solution 161
7.3 Applications of ON Functions in the Simultaneous Solution 168
7.4 The Altimeter Data Used in the Solutions 172
7.5 Numerical Experiments of Simultaneous Solutions and Results 173
7.5.1 Experiments 173
7.5.2 Results 176
7.5.2.1 Sea Surface Topography Models from the Solutions and Geostrophic Currents 176
7.5.2.2 Comparison with Levitus SST and the One-year Geosat Solution 188
7.5.2.3 The Estimability of Geocentric Shift Components ΔX and ΔY 188
7.5.2.4 Accuracies of the SST and Geoid Undulation from the Solutions 192
7.5.2.5 Orbit Accuracies from the Solutions and Some Statistics 201

7.5.2.6 Correlation Analysis 206
7.5.2.7 Conclusions 207

VIII. Conclusions and Recommendations 208

Appendices

A. Proof of Orthonormality of Functions Obtained from the Gram-Schmidt Process 212

B. Green's Second Identity on a Curved Surface in an Orthogonal Curvilinear Coordinate System 213

C. Bessel Functions and Orthogonality Relationship 219

List of References 222

List of Tables

Table Page

2.1 Orthogonal Functions in One-Dimensions and Two-Dimensions 6

3.1 $\partial/\partial\ell(P_\ell^m)$ From Analytical Formula and Numerical Differentiation at $t_0 = \cos 50^\circ$ .. 37

5.1 Total Numbers of Combination Constants Required for Various Maximum Degrees of Spherical Harmonics 85

5.2 Comparison of I_{nmrs} Values from the Product-Sum Formulae and the Recursive Formulae at the nm, rs Pairs Where Maximum Differences Occur 86

5.3 Maximum j Values in System $\{L_j\}$ Before Which the Elements are Independent .. 90

5.4 Maximum Difference Between Combination Coefficients from the Recursive Algorithm and Cholesky Decomposition up to a Maximum Degree 91

5.5 Comparison of Orthonormality of the ON Functions Constructed by the Recursive Algorithm and Cholesky Decomposition 92

6.1 CPU Times Comparison Using FFTSOL and ADJSST on CRAY Y-MP/864 113

6.2 CPU Times on CRAY Y-MP/864 for Forming Normals and U Vectors by FFT 115

6.3 Various Expansions of the Levitus SST and the Comparisons with Original Data 128

6.4 Cumulative Average Powers in Meter**2 and Percentages of Energy for Solutions (1) onlsfto24, $\{X_j\}$ (2) onnqto24, $\{X_j\}$ 130

6.5 Cumulative Average Powers, Percentages of Energy and Figure of Merit for Solution onlsfto24, $\{Z_j\}$ 130

6.6 RMS Differences D_n Between the Unitarily Transformed Coefficients and the Coefficients from the Direct SST Expansion Using System $\{Y_j\}$ 132

6.7 Square Roots of Degree Variances of the ON Functions and the Spherical Harmonics Using 4 Sets of 1° x 1° Mean Oceanic Geoids 136

7.1 Frequency Classification of Satellite Orbit Using Keplerian Elements 155

7.2 Start/Stop Times of the 6 GEM-T2 Arcs Used for Experiments 172

7.3 Square Roots of the SST a priori Degree Variances for Conditioning the Simultaneous Solutions 176

7.4 Categories of Simultaneous Solutions 176

7.5 RMS Differences by Degree Between the SST in Figure 7.4 and the SST in Figure 7.5 Using ON Functions 177

7.6 Correlations by Degree Between the 6-arc SST Solutions in Category I to V and the Levitus SST in Percentage 189

7.7 Correlations Between the Geocentric Shift Components and a_0, a_2 Terms of Orbit Errors 191

7.8 Square Roots of SST Signal/Error and Geoid Error Degree Variances from Geosat 6-arc Solution Using ON Functions, Category III 197

7.9 RMS Values of Corrections to the Residual Sea Surface Heights and Adjusted Crossover Discrepancies in Meters 202

7.10 RMS Values of Orbital Errors (Solution Categories are the Same as Those in Table 7.9) and Geocentric Shift in Meters 203

7.11 Comparison of Orbit Errors of GEM-T1 and GEM-T2 Orbits 206

List of Figures

Figure — Page

3.1 The Domain, Boundary and Outer Normal for the 2-D Helmholtz Equation 17

3.2 An Isosceles for the Helmholtz Equation 23

3.3 Legendre Functions of Non-integer Degree, m = 0 for all Curves, ℓ Plotted Next to the Curves, $\cos 50° \le t \le 1$ 29

3.4 A Domain σ and Its Boundary B and Outer Normal n on the Unit Sphere 32

3.5 A Spherical Cap of Radius θ_0 with Boundary B 33

3.6 Polar Coordinate for a Unit Circle 39

3.7 Normalized Bessel Function in $\cos 10° \le t \le 1$, m = 0, Roots r_{mp} Plotted Next to the Curves 43

3.8 A Region Bounded by Two Parallels and Two Meridians 43

3.9 Zeros of Two Eigenfunctions, $k_2 > k_1$ 44

3.10 Riemann's Mapping Theorem 47

3.11 A Simply Connected Domain R, a Multiply Connected Domain D and a Curve C 48

3.12 Transformation Between Domains B_z and B_w 48

3.13 Mapping the Unit Disk Onto a Polygon 53

4.1 (a) Function $\overline{T}_n(t)$ (b) Function $\overline{U}_n(t)$ 70

5.1 Oceans Implied by the Edited TUG87 1° x 1° Elevation Data (see text) 88

5.2 Oceans as the Area Where the Modified SST of Levitus Exists (Set 3 of Engelis, 1987b), $\overline{H} \le -2250$ Meters 88

6.1 Modified Levitus SST (Set 3 of Engelis, 1987b), CI = 10 cm 101

6.2 Variable ψ in Equation (6.21) 106

6.3 Unit Vectors at the Oceanic Boundaries 133

6.4 Comparisons of the Degree Variances of the ON Function Expansion and the Degree Variances of the Spherical Harmonic Expansion from Various Solutions 137

6.5 Degree Variances from Solutions shnqto24 and shlsfpto24 138

6.6 Degree Variances from Solutions onlsfto24, $\{Y_j\}$ and onlsfto24, $\{Z_j\}$ 139

6.7 SST from Solution onlsfto10, $\{Y_j\}$, CI = 10 cm 140

6.8 Difference Between the Levitus SST and SST from Solution onlsfto10, $\{X_j\}$, CI = 5 cm 140

6.9 SST from Solution onlsfto15, $\{X_j\}$, CI = 10 cm 141

6.10 Difference Between the Levitus SST and SST from Solution onlsfto15, $\{X_j\}$, CI = 5 cm 141

6.11 SST from Solution onlsfto24, $\{X_j\}$, CI = 10 cm 142

6.12 Difference Between the Levitus SST and SST from Solution onlsfto24, $\{X_j\}$, CI = 5 cm 142

6.13 Geostrophic Currents from Solution onlsfto10, $\{X_j\}$ 143

6.14 Geostrophic Currents from Solution onlsfto24 143

6.15 SST from Solution onlsfto10, $\{Z_j\}$, CI = 10 cm 144

6.16 Difference Between the Levitus SST and SST from Solution onlsfto10, $\{Z_j\}$, CI = 5 cm 144

6.17 SST from Solution onlsfto24, $\{Z_j\}$, CI = 10 cm 145

6.18 Difference Between the Levitus SST and SST from Solution onlsfto24, $\{Z_j\}$, CI = 5 cm 145

6.19 Geostrophic Currents from Solution onlsfto10, $\{Z_j\}$ 146

6.20 Geostrophic Currents from Solution onlsfto24, $\{Z_j\}$ 146

6.21 SST from Solution shlsfpto24, CI = 10 cm 147

6.22 Difference Between the Levitus SST and SST from Solution shlsfpto24, CI = 5 cm 147

6.23 SST from Solution shnqto24, CI = 10 cm 148

6.24 Difference Between the Levitus SST and SST from Solution shnqto24, CI = 5 cm 148

7.1 Satellite Orbit Geometry in a Near-circular Orbit 159

7.2 Geometry of Altimetric Observations over the Oceans 162

7.3 Distribution of 20-second Normal Points from Arcs 22, 23 and 24 174

7.4 SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Category I, CI = 10cm 178

7.5 SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Category II, CI = 10cm 179

7.6 Difference Between the SST from ON (24, 24) 6-arc Solution, Category I and ON (24,24) Solution, Category II 179

7.7 SST to ON Degree 15 from ON (24, 24) 6-arc Solution, no Downweighting to Gravity Normal, Category III, CI = 10cm 180

7.8 SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Downweighting Gravity Normal by 1/2, Category III, CI = 10cm 181

7.9 SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Downweighting Levitus's Normal by 1/4, Category IV, CI = 10cm 183

7.10 SST from Spherical Harmonic (15, 15) 6-arc Solution, Category V, CI = 10cm 184

7.11 Difference Between the SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Category III and the Spherical Harmonic (15, 15) Solution, Category V, CI = 5 cm 185

7.12 Geostrophic Currents Implied by the SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Category I 185

7.13 Geostrophic Currents Implied by the SST to ON Degree 15 from ON (24, 24) 6-arc Solution, No Downweighting to Gravity Normal, Category III 186

7.14 Geostrophic Currents Implied by the SST from Spherical Harmonic (15, 15) 6-arc Solution, Category V 187

7.15 SST to ON Degree 10 Using ON System $\{Y_j\}$ from Unitary Transformation of ON (24, 24) 6-arc Solution in Category III, CI = 10cm 190

7.16 Geoid Undulation Errors Corresponding to ON (24, 24) 6-arc Solution, Category III, CI = 5cm 194

7.17 SST Errors Corresponding to the SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Category III, CI = 5cm 194

7.18 SST Errors Corresponding to the Spherical Harmonic (15, 15) Solution, Category V, CI = 5cm 195

7.19 Degree Variances and Error Degree Variance from Geosat 6-Arc Solution Using Spherical Harmonic Functions, Category V 196

7.20 Degree Variance and Error Degree Variance from Geosat 6-Arc Solution Using Orthonormal Function, Category I 198

7.21 Error Degree Variances of Geoid Undulations from Geosat 6-arc Solutions Using Spherical Harmonics (SH) and Orthonormal Functions (ON), Category I and III 199

7.22 Degree Variance and Error Degree Variance from Geosat 6-arc Solution Using Orthonormal Functions, Category IV 200

7.23 Geoid Undulation Corrections from ON (24, 24) 6-arc Solution, Category III, CI = 1m 205

1. Introduction

A satellite borne altimeter has the capability to measure the range between the center of the mass of the satellite and the instantaneous sea surface. Two important signals that the geodesists and oceanographers desire to extract from the range information are the geoid and the sea surface topography (SST). However, in order to obtain meaningful signals, numerous corrections to the observations must be applied. The standard corrections, provided with the altimeter data, are the ionospheric and tropospheric refraction corrections, ocean tide and solid tides corrections, instrument and timing bias corrections, and sea state bias corrections. The additional and necessary corrections for the precise geoid and the SST determination are the reductions of satellite radial orbit errors.

One of the promising techniques to recover the geoid and the SST is within a model that simultaneously determines orbital correction parameters, the geoid and the SST. In such a model, the SST is usually represented by the surface spherical harmonics

$$\zeta(\theta, \lambda) = \sum_{n=0}^{N_1} \sum_{m=0}^{n} (\bar{a}_{nm}\cos m\lambda + \bar{b}_{nm}\sin m\lambda)\bar{P}_n^m(\cos\theta) \tag{1.1}$$

and the geoid undulations posesses a similar form:

$$N(r, \theta, \lambda) = \frac{GM}{r\gamma} \sum_{n=2}^{N_2} \left(\frac{a_e}{r}\right)^n \sum_{m=0}^{n} (\bar{C}_{nm}\cos m\lambda + \bar{S}_{nm}\sin m\lambda)\bar{P}_n^m(\cos\theta) \tag{1.2}$$

In these equations, the SST and the geoid have been expanded up to maximum degrees N_1 and N_2, respectively. There are several problems with the spherical harmonic representation of the SST in (1.1). Engelis and Knudsen (1989) found in their Seasat work that high correlations exist between degree 1 terms of the SST coefficients and the parameters of orbital corrections of 1 cyc/rev terms, despite some conditioning of parameters applied in the adjustment process. In addition, Engelis (1987a), Wagner (1986), Wagner (1989) and others also presented the similar results on the high correlation problem in the joint geoid-SST solution using the spherical harmonic representation scheme.

Denker and Rapp (1990) addressed the problem of using spherical harmonics even more comprehensively. First of all, they pointed out that the spherical harmonic representation is not appropriate for data defined only over the oceans. Since they are not orthogonal with respect to a scalar product defined as integral over the oceans, spectral analyses of oceanic signals using such a representation may lead to misleading results and implications. Also, using such a global representation of N and ζ will create difficulty in interpreting the degree variances since the accuracy distribution is extremely non-uniform over the earth. The spherical harmonic representation also creates problems in representing the current flow near the continental boundaries since the current vectors with this representation will not be parallel to the continental boundaries.

As a further example, we show a problem found in Engelis' (1987b) work. Analyzing the Levitus SST (Levitus, 1982) using spherical harmonic expansions, Engelis (ibid.) has to sacrifice the approximation accuracy and restricted his expansions to only degree 10 due to the high correlations between the SST harmonic coefficients and the

excessively large magnitudes of the coefficients. In view of such a dilemma between the requirement of approximation accuracy and explainable spectral contents of oceanic signals, the non-orthogonal basis functions over the oceans such as the spherical harmonic functions indeed have serious defects.

With these problems as background, a different set of basis functions other than the spherical harmonics is thought to be necessary for representing oceanic signals such as the SST and the oceanic geoid. The ideal basis functions will be a set of orthogonal functions over the oceans with which spectral analyses such as decomposition of signals, energy distribution of oceanic signals, signal-to-noise ratio studies can be conducted over the oceans and produce correct conclusions and implications. Thus in the present study effort will be made to construct a possible set of orthogonal functions to fulfill such a need.

The outline of this report is as follows. Chapter 2 provides some definitions and properties of approximation theory related to expansions of a function in orthogonal functions. It also serves as an introductory chapter that will describe the advantage of using orthogonal functions and will provide some necessary literature about some existing orthogonal functions for some specific domains. Chapter 3 is a relatively long chapter which is devoted to three methods of constructing orthogonal functions. All the necessary equations and software for the proposed orthogonal functions will be given in that chapter. In Chapter 4, the orthogonal functions over the entire unit sphere are investigated and subsequently two sets of orthogonal functions as alternatives for the spherical harmonics are presented.

Upon choosing the Gram-Schmidt process as the method for further detailed study of the construction of the orthonormal functions over the oceans, all necessary techniques and software for that method are exploited in Chapter 5. Especially the new techniques of integrating products of two associated Legendre functions and computing the inner products of spherical harmonics over the oceans will be stressed. Finally some properties of the orthonormal functions constructed by the Gram-Schmidt process will be given in that chapter.

Chapter 6 is then devoted to the expansions of the Levitus SST and the oceanic geoid into the orthonormal functions constructed in Chapter 5. Analyses on the spectral behavior and the fit to the expanded data in such expansions will be made. In this chapter some highly improved methods which deal with the "old" problems will be also found.

In Chapter 7, some improved models for the simultaneous estimation of satellite radial orbit errors, the geoid and the SST are developed. The orthonormal functions will then be incorporated into these models as the basis functions for the SST for the purpose of parameter estimation. A method is developed to assess the accuracy of the geoid from the simultaneous solutions using the orthonormal functions. Finally in Chapter 8 conclusions and summary will be made and future studies will be proposed.

2. Expansions in Series of Orthogonal Functions

In this chapter, several definitions of operations and quantities will be provided. These definitions will help the understanding of the important issues in the chapters that follow. The discussion is primarily related to the general properties of approximation, especially those using orthogonal series.

2.1 The Weierstrass Approximation Theorem and Uniform Approximation

We begin our discussion with a well-known approximation theorem introduced by Weierstrass. The Weierstrass approximation theorem states that (Davis, 1975, p. 107):

If $f(x)$ is continuous on the interval $[a, b]$ and $\varepsilon > 0$ is arbitrary, we can find a polynomial $P_n(x)$ (of sufficiently high degree) for which

$$|f(x) - Pn(x)| \leq \varepsilon, a \leq x \leq b \tag{2.1}$$

The implication of Weierstrass theorem is that any continuous function (for the moment, it is also one-dimensional) can be approximated by a polynomial as close as possible. This theorem can be proved in many ways, among them the use of Bernstein's polynomials (Rivlin, 1981, p. 12) is the most elegant.

Extension of the Weierstrass approximation theorem to a higher dimension case is given by Davis (ibid., p. 122). For this study, our concern about approximation is mainly for a two-dimensional case (e.g., on the surface of a unit sphere); the extension thus ensures the successful construction of polynomials in two variables with which the quantities of interest can be approximated to the degree specified.

The Weierstrass approximation theorem in fact describes a uniform approximation. Since, upon choosing an $\varepsilon > 0$ value, the deviation of $P_n(x)$ in (2.1) from $f(x)$ will be less than ε for any x within $[a, b]$. Another type of approximation, the least-squares approximation, is more commonly used and will be addressed in the following section.

2.2 The Inner Product Space and Least-Squares Approximation

Least-squares approximation, in a general sense, is a way of approximating functions by minimizing some scalar product. The concept of scalar product will be discussed in connection with the inner product.

An inner product space is a linear space equipped with a scalar product (x, y), x and y being its two members, which possesses the following properties (Davis, 1975 and Dettman, 1988):

(i) Symmetry:
$$(x, y) = (y, x)^* \tag{2.2}$$

(ii) Linearity:
$$(x, y + z) = (x, y) + (x, z) \tag{2.3}$$

(iii) Homogeneity:
$$(x, \alpha y) = \alpha^*(x, y) \tag{2.4}$$

(iv) Posrive-definiteness:
$$(x, x) \geq 0, (x, x) = 0 \text{ if only if } x = 0 \tag{2.5}$$

where * is the conjugate operator.

For this study, we give the following two definitions of scalar products which will be used in later developments:

• Definition 1:

f, g are two square-integrable functions defined on a domain D, or f, g ∈ $L^2[D]$ and w is a nonnegative weight function defined on D, then

$$(f, g) = \frac{1}{\dim(D)} \int_D fg^* w d(D) \tag{2.6}$$

where dim(D) is the weighted dimension of domain D, being length for a one-dimensional case and area for a two-dimensional case. In (2.6), f, g are two complex functions. For real functions f, g, we have

$$(f, g) = \frac{1}{\dim(D)} \int_D fgwd(D) \tag{2.7}$$

In the classical definition of inner products, dim(D) is normally omitted, i.e.

$$(f, g) = \int_D fgwd(D) \tag{2.8}$$

However, to be consistent with the definition of the fully normalized (surface) spherical harmonics for which the area of the unit sphere, 4π, is used, the definition in (2.7) will be adopted as the standard definition of inner product unless otherwise specified.

• Definition 2:

Let f_i, g_i be functions f, g evaluated at point i, w_i be the weight function w evaluated at point i, i = 1, ... n, then

$$(f, g) = \sum_{i=1}^{n} f_i g_i^* w_i \tag{2.9}$$

defines a discrete version of the inner product according to (2.8).

In many occasions, observations are available only at some random points of D; the operations (such as approximations, spectral analysis) are thus very difficult to be carried out over the entire domain D. A compromise is to do the operations only at the points where observations exist. By doing this one essentially regards a function as a vector of finite elements. In section 3.3.4, the comparison between an orthogonal series of functions and a series obtained by the QR factorization (Stewart, 1973) will be based on this idea.

To make this report self-contained, we also introduce the following definition of inner product related to the Lebesque integral:

• Definition 3:

Let f, g be two square integrable (in the Lebesque sense) functions with no more than countably many distinct values on respective domains, namely

$$y_1, y_2, \ldots, y_n, \ldots \text{ for } f$$

and

$$z_1, z_2, \ldots, z_n, \ldots \text{ for } g$$

then the inner product in the sense of (2.8) is

$$(f, g) = \int_A f(x)g(x)d\mu = \sum_n y_n z_n \mu(A_n) \tag{2.10}$$

where

$$A_n = \{x: \ x \in A, f(x) = y_n, g(x) = z_n\}$$

μ is the measure on set A. This definition has been based on Taylor (1965) and Kolmogorov and Fomin (1970). An infinite set is called countable if its elements can be put in one-to-one correspondence with the elements of a set containing all postive integers (Kolmogorov and Fomin, 1970, p. 10)

Having defined the inner products, we may now define a norm as

$$\| x \| = (x, x)^{1/2} \tag{2.11}$$

Thus if x is an error function obtained by differencing the approximated function and the approximating function, then a least-squares approximation can be achieved by minimizing the error norm $\| x \|$. This is the least-squares (of error) principle.

2.3 Orthogonal Functions

A complete normed inner-product space is called a Hilbert space. It turns out that the set of all square integrable functions in domain D constitute a Hilbert space, $L^2(D)$ (Sansone, 1959). Let the infinite set of functions in $L^2(D)$ be f_i, $i = 1, \ldots \infty$, then f_i's are mutually orthogonal if

$$(f_i, f_j) = \begin{cases} \text{constant} , & i = j \\ 0 , & i \neq j \end{cases} \tag{2.12}$$

f_i's are mutually orthonormal if

$$(f_i, f_j) = \begin{cases} 1 , & i \neq j \\ 0 , & i \neq j \end{cases} \tag{2.13}$$

A set of functions $\{f_i\}$ fulfilling condition (2.12) or (2.13) is called an orthogonal set (or system) or an orthonormal set (or system) respectively. In terms of spectral analysis, $\{f_i\}$ can be used for generalized Fourier series expansions.

For a domain of regular geometry (such as a real line, a square, a unit circle, the surface of a sphere), numerous orthogonal functions can be found in the literature. Examples are the sine-cosine functions for $[0, 2\pi]$; the Fourier-Bessel functions for a unit circle; the surface spherical harmonics for the surface of a unit sphere. In the current study, the primary domain (D) of concern is the oceans which appear to be quite irregular. For such a domain, an orthogonal system cannot be readily found, thus a study is needed to construct a possible one.

The above discussions automatically include the real orthogonal functions and the complex orthogonal functions. Table 2.1 summarizes the orthogonal functions, domains and types (real or complex) that frequently appear in literature. Standard treatment of orthogonal functions may be found in Davis (1975, chapters 5-7) and Sanone (1959). The physical applications of orthogonal functions, especially those arising from eigenvalue-eigenfunction problems (to be discused in section 3.4), may be found in Courant and Hilbert (1953, chapters 5-7) and Morse and Feshbach (1953, chapter 6). A good discussion on complex orthogonal functions can be found in Gaier (1987).

Table 2.1 Orthogonal Functions in One-Dimensions and Two-Dimensions

Name	Notation	Domain	Type*	Reference
Fourier	cosnx, sinnx	$0 \le x \le 2\pi$	R	Tolstov, 1976, chapter 1
Ass. Legendre	$P_n^m(x)$	$-1 \le x \le 1$	R	Hobson, 1965, p. 91
Tschebysheff, 1st	$T_n(x)$	$-1 \le x \le 1$	R	Davis, 1975, p. 365
Jacobi	$P_n^{(\alpha,\beta)}(x)$	$-1 \le x \le 1$	R	Davis, 1975, p. 366
Generalized Laguerre	$L_n^{(\alpha)}(x)$	$0 \le x \le \infty$	R	Sansone, 1959, p. 295
Hermite	$H_n(x)$	$-\infty \le x \le \infty$	R	Sansone, 1959, p. 303
Bessel, int. order	$J_n(x_{nm}\frac{x}{a})$	$0 \le x \le a$	R	Lebedev, 1972, p. 128
Mathieu function	$S_{em}J_{em}$, $S_{om}J_{om}$	ellipse	R	Morse et al. 1953, p. 562
Fourier-Bessel‡	$J_n(x_{nm}r)\pm e^{in\lambda}$	circle	R	Kaplan, 1981, p. 449
Fourier-Bessel‡	$[J_n(K_{nm}r)+cN_m(K_{nm}r)]e^{\pm in\lambda}$	circle ring	R	Sommerfeld, 1949, p. 167
Fourier-Bessel‡	$J_\mu(K_{\mu m})\sin\mu\lambda$	circular sector	R	Sommerfeld, 1949, p. 168
Polynomial in z	z^n	circle	C	Davis, 1975, p. 240
Tschebyscheff$^\Delta$	$T_n(z)$	ellipse	C	Davis, 1975, p. 240
Spherical harmonics	$R_{nm}(\theta, \lambda)$, $S_{nm}(\theta, \lambda)$	unit sphere	R	H/M, 1967, p. 29

* R: real; C: complex.

‡ The use of $e^{\pm in\lambda}$ is just for convenient notation, the functions are real in nature.

$^\Delta$ Now Tschebyscheff polynomials have a complex variable.

2.4 Some Properties of Orthogonal Function Expansions

For a convenient discussion of the properties of orthogonal function expansions, we restrict the case to real functions and a set of orthonormal functions, $\phi_1, \phi_2, \ldots \phi_M$ (M may be infinite). An arbitrary function $f \in L^2(D)$ can be expanded into ϕ_i as follows:

$$\hat{f} = \sum_{i=1}^{M} c_i \phi_i \tag{2.14}$$

where $\hat{f}$ is the approximation of f and M is the truncation degree. The minimum property states that the error, $\| f - \hat{f} \|^2$, measured in a norm according to (2.7), is minimum if c_i's are obtained by

$$c_i = (f, \phi_i) \tag{2.15}$$

This is easily seen as follows. The error φ is:

$$\varphi = \|f - \hat{f}\|^2 = \|f - \sum c_i \phi_i\|^2 = \frac{1}{\dim(D)} \int_D \left(f - \sum_{i=1}^{M} c_i \phi_i \right)^2 d(D) \tag{2.16}$$

In order to have a minimum φ, we take

$$\frac{\partial \varphi}{\partial c_i} = 0 \quad , \quad i = 1, 2, \cdots, M \tag{2.17}$$

Thus we have

$$\int_D \left(f - \sum_{j=1}^{M} c_j \phi_j \right) \phi_i d(D) = 0 \tag{2.18}$$

Using the orthogonality relationship $(\phi_i, \phi_j) = \delta_{ij}$, we finally arrive at (2.15).

Since $\varphi = \|f - \hat{f}\|^2 > 0$, we have, according to (2.16),

$$\int \left(f - \sum_{i=1}^{M} c_i \phi_i \right)^2 d(D) = \int_D f^2 d(D) - \sum_{i=1}^{M} c_i^2 \int_D \phi_i^2 d(D) > 0$$

Since $(\phi_i, \phi_i) = 1$, we have

$$\|f\|^2 \geq \sum_{i=1}^{M} c_i^2 \tag{2.19}$$

This is the Bessel inequality. When M approaches ∞, we have

$$\lim_{M\to\infty} \left\| f - \sum_{i=1}^{M} c_i\phi_i \right\| = 0 \tag{2.20}$$

The set $\{\phi_i\}$ satisfying (2.20) for any f defined and square-integrable on D is complete in $L^2(D)$. When (2.20) is true, we get

$$\|f\|^2 = \sum_{i=1}^{\infty} c_i^2 \tag{2.21}$$

which is Parseval's theorem.

The final discussion will concern the statistical properties of orthogonal function expansions. We assume that we expand function f into ϕ_i as in (2.14) and obtain the covariance matrix for c_i from some parameter estimation process. We will be able to get an error estimate of function f at point s using the error propagation

$$\sigma_s^2 = \sum_{i=1}^{M} \sigma_i^2\phi_i^2(s) + \sum_{\substack{i=1 \\ i\neq j}}^{M}\sum_{j=1}^{M} \sigma_{ij}\phi_i(s)\phi_j(s) \tag{2.22}$$

where σ_s^2 is the (error) variance of f at point s, σ_i^2 is the variance of c_i and σ_{ij} is the covariance between c_i and c_j. The average cumulative error variance of f, σ_f^2, is

$$\sigma_f^2 = \frac{1}{\dim(D)} \int_D \sigma_s^2 d(D)$$

$$= \frac{1}{\dim(D)} \left(\sum_{i=1}^{M} \sigma_i^2 \int_D \phi_i^2 d(D) + \sum_{\substack{i=1 \\ i\neq j}}^{M}\sum_{j=1}^{M} \sigma_{ij} \int_D \phi_i\phi_j d(D) \right)$$

$$= \sum_{i=1}^{M} \sigma_i^2 \tag{2.23}$$

Eq. (2.23) shows that the covariance σ_{ij} does not play a role in calculating σ_f^2. Such a property will provide a good estimate of geoid error over the oceans and SST error in the

process of simultaneous estimation using orthonormal functions that will be described in Chapter 7.

Defining the i^{th} signal component as $f_i = c_i\phi_i$, we can measure the correlation between components i and j using the average product (Heiskanen and Moritz, 1967, p. 256)

$$\overline{M}(f_i f_j) = \frac{1}{\dim(D)} \int_D c_i c_j \phi_i \phi_j d(D)$$

$$= c_i^2 \delta_{ij} \qquad (2.24)$$

where $\overline{M}$ is the averaging operator. Note that (2.24) requires the underlying process to be ergodic. Eq. (2.24) reveals the fact that two distinct signal components can possibly be considered statistically independent in orthogonal function expansions.

3.0 Methods of Constructing Orthonormal Functions Over the Oceans

Three methods for constructing orthonormal functions (or, in general, orthogonal functions) over a specified domain, especially over the oceans, will be investigated here. One final choice will be made for the numerical experiments of the simultaneous geoid-SST estimation in altimetry.

3.1 The Definition of the Oceans

Why do we need to define the oceans? In the geodetic literature, the term "the domain of the oceans" is often mentioned, but not clearly defined. We illustrate two kinds of geodetic problems that involve the domain of the oceans:

— The altimetry-gravimetry problem (Sansò and Stock, 1985). The task is to solve for the disturbing potential T given data on land and on the oceans.

— The overdetermined boundary value problem (Sacerdote and Sansò, 1985). Here T is solved given the redundant boundary values on land and on the oceans.

In the first problem, a single type of data is given at a portion of the boundary while another type of data is given on the rest of the boundary. In the second problem, a portion of the boundary could have two or more kinds of data; thus redundancy is created. In these two problems, it is likely that "the oceans" is defined as the domain where geoid undulations (or gravity disturbances) are given, since the problems can still be attacked without knowing whether the domain is really "ocean" or not (in the oceanographic sense). For one thing; we can conduct gravimetry on the oceans and can obtain geoid undulations by differencing precise ellipsoidal height (from, e.g., GPS) and orthometric height (from levelling plus gravimetry) on land. Thus the clear demarcaration between land and the oceans is not very much required.

It is not so in the simultaneous geoid-SST estimation in altimetry (see Chapter 7 for the definition). SST is defined only over the area where oceanic waters exist (we exclude the in-land seas such as the Caspian Sea from our discussion). However, our world is

surrounded by complexities and details, therefore we have to "generalize" the phenomena or quantities to the extent that research of a global scale can be carried out. But the generalization will never intend to mix the true oceans with land.

In summary, the following options are proposed for defining a domain of the oceans suitable for constructing a set of orthornormal functions (over the oceans) and for the analysis of the results from the simultaneous geoid-SST estimation in altimetry:

Option 1: Define the boundary of the oceans using a set of polygons of shoreline data. A point interior to the polygons is regarded as land, otherwise it is ocean. Additional editings can be made to remove the undesired "oceans", e.g., the Caspian Sea.

Option 2: Define the oceans as the domain consisting of a finite set of equiangular blocks (in latitude and longtitude). Given a global (complete) DTM data, such as the 1° x 1° mean elevations from TUG87 (Wieser, 1987), a block is ocean if $\bar{H} < 0$ otherwise it is land. $\bar{H}$ is the mean elevation. Again, editing can be made to remove the in-land seas, basins, etc. which obviously do not belong to the oceans.

Option 3: Define the oceans as the domain where observations exist. For example, set 3 of the SST data analyzed by Engelis (1987b) implies a domain constituting a set of 1° x 1° blocks where $\bar{H} < -2500$ m (except at the Mediterranean Sea and the Black Sea, where $\bar{H} > -2500$ m).

None of the options could provide a "true" boundary of the oceans, due to the factors such as shoreline resolution, the accuracy and resolution of the DTM, etc. In the numerical experiments conducted in Chapters 6 and 7, the definitions of the oceans according to Option 2 or Option 3 will be adopted. However, the analytical derivation of orthonormal functions for some special cases will imply the use of Option 1.

In contrast to the orthonormal functions such as the fully normalized spherical harmonics over the unit sphere, the resulting orthonormal functions according to any of the above mentioned options will be "data-dependent". Here the "data" means the (smoothed) shoreline or the DTM used in defining the domain. In short, the resulting orthonormal functions are not universal to all the users.

3.2 The Gram-Schmidt Orthonormalizing Process

3.2.1 The Principle

We now introduce a purely "mechanical" approach of constructing orthornomal functions, the Gram-Schmidt orthonormalizing process (Davis, 1975, p. 165). Assuming that we are given a finite or countably infinite set of linearly independent functions, f_1, f_2, ... , and $f_i \in L^2(D)$, as shown in Appendix A, the set of functions $\bar{f}_i$ obtained by the following process are mutually orthonormal:

$$h_1 = f_1 \quad , \quad \bar{f}_1 = h_1/\|h_1\|$$

$$h_2 = f_2 - (f_2, \bar{f}_1)\bar{f}_1 \quad , \quad \bar{f}_2 = h_2/\|h_2\|$$

$$\vdots$$

$$h_n = f_n - \sum_{k=1}^{n-1} (f_n, \bar{f}_k)\bar{f}_k, \; \bar{f}_n = h_n/\|h_n\| \qquad (3.1)$$

etc.

In general, we can write (3.1) as

$$\bar{f}_i = \sum_{j=1}^{i} c_{ij} f_j \qquad , \; i = 1, \cdots \qquad (3.2)$$

where c_{ij} are the combination coefficients in the orthonormalizing process.

The functions f_i are independent if $\sum_{i=1}^{n} a_i f_i = 0$ implies that $a_i = 0$, for $i = 1, \cdots, n$; if at least one a_i could be chosen non-zero, then f_i are not linearly independent or some of them are dependent. To check the mutual dependence among the set of functions f_i in D, one can also use the Gram matrix G, defined as

$$G = G(f_1, f_2, \cdots, f_n) = \begin{bmatrix} (f_1, f_1)(f_1, f_2) \cdots (f_1, f_n) \\ (f_2, f_1)(f_2, f_2) \cdots (f_2, f_i) \\ \vdots \\ (f_n, f_1) \cdots\cdots\cdots (f_n, f_n) \end{bmatrix} \qquad (3.3)$$

It can be shown that the determinant of G, |G|, fulfills the condition (Davis, 1975, p. 178)

$$0 \le |G| \le \|f_1\|^2 \|f_2\|^2 \cdots \|f_n\|^2 \qquad (3.4)$$

The necessary and sufficient condition for f_i to be linearly dependent is that (Sansone, 1959)

$$|G| = 0 \qquad (3.5)$$

Eq. (3.3), together with (3.5), will provide an important tool for checking the dependence of surface spherical harmonics over some domains in later developments (n in (3.3) must be finite in such a checking).

Davis (1975, p. 168) has shown that some classical orthogonal polynomials, such as $P_n(x)$, $T_n(x)$ in Table 1.1 can be derived through this process. However, if n in (3.1) is very large or the evaluation of the inner product of two functions, (f_i, f_j), is complicated, an efficient computational algorithm should be developed to find the combination coefficients c_{ij}. Two algorithms will be described in the following section.

It should be noted that the orthonormalizing process according to (3.2) is the triangular scheme, since

$$\begin{bmatrix} \bar{f}_1 \\ \bar{f}_2 \\ \vdots \\ \bar{f}_n \end{bmatrix} = \begin{bmatrix} c_{11} & & & \\ c_{21} & c_{22} & & 0 \\ \vdots & & & \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \tag{3.6}$$

Written in a matrix form, (3.6) is equivalent to

$$y = Cx \tag{3.7}$$

where $y = (\bar{f}_1, \cdots, \bar{f}_n)^T$, $x = (f_1, \cdots, f_n)^T$. The combination matrix C is a lower triangular matrix. Another combination scheme, called the general scheme, has a full combination matrix. That is, instead of adding the functions f_i up to i in (3.2), an orthonormal function $\bar{f}_i'$ is obtained by adding all f_j's as follows:

$$\bar{f}_i' = \sum_{j=1}^{n} c_{ij}' f_j \quad , \quad i = 1, \cdots, n \tag{3.8}$$

As pointed out by Davis (ibid.), given a certain order of the functions f_j, the combination coefficients are uniquely determined in the triangular scheme, but not in the general scheme. For this study, the triangular scheme is preferred. One reason is related to the decay of the power of the signal analyzed in this study: Let f_i be the power series in t, i.e. $1, t^1, t^2, \cdots, t^n$. The resulting orthonormal functions $\bar{f}_i$ will be a set of polynomials $P_i(t)$, $i = 1, \cdots, n$ if the triangular scheme is used. However, (3.8) will yield a set of polynomials of the same degree, i.e., $P_n(t)$. In most of the signals encountered in geophysics, the amplitude of the signal component will decay as the frequency increases. An appropriate frequency decomposition can be achieved from the use of polynomials of variable degree, such as $P_i(x)$, $i = 1, \cdots, n$, but not a set of polynomials of the same degree.

3.2.2 The Combination Coefficients in the Orthonormalizing Process

To find the combination coefficients c_{ij} in (3.2), the first step is to evaluate the inner products of the given functions, f_i. Given n functions, a total of $n(n + 1)/2$ inner products need to be evaluated. These functionals correspond to the elements in the lower triangular part of the Gram matrix in (3.3). Since $(f_i, f_j) = (f_j, f_i)^*$, the upper triangular part of G can also be formed by these inner products. If f_i's are real, it can be shown that G is a real, symmetric positive definite matrix assuming all f_i's are independent. Once the inner products are formed, two algorithms can be employed to find the coefficients c_{ij}. This is shown in the following discussion.

The first algorithm is based on Mainville (1987, Chapter V). For a convenient discussion, we assume f_i's are real. We observe from (3.7) that functions f_i can also be expressed as a linear combination of $\bar{f}_i$, since

$$x = C^{-1}y \tag{3.9}$$

or

$$f_i = \sum_{j=1}^{i} \bar{c}_{ij}\bar{f}_j \tag{3.10}$$

Thus

$$(\bar{f}_i, f_j) = \begin{cases} \bar{c}_{ii} & , \text{ for all } j \text{ with } j \geq i \\ 0 & , \text{ for all } j \text{ with } j < i \end{cases} \tag{3.11}$$

In (3.9), C^{-1} exists since all c_{ii}'s are positive (so the determinant of C is not zero!). Using the property of a triangular matrix, one gets

$$\bar{c}_{ii} = \frac{1}{c_{ii}} \tag{3.12}$$

Based on (3.10), we can write (3.2) as

$$\bar{f}_i = c_{ii}\left(\sum_{j=1}^{i-1} b_{ij}\bar{f}_j + f_i\right) \tag{3.13}$$

For $p < i$, one has

$$(\bar{f}_i, \bar{f}_p) = 0 = c_{ii}\left[\sum_{j=1}^{i-1} b_{ij}(\bar{f}_j, \bar{f}_p) + (f_i, \bar{f}_p)\right]$$

Thus

$$b_{ij} = -(f_i, \bar{f}_j) = -\sum_{p=1}^{i-1} c_{jp}(f_i, f_p) \tag{3.14}$$

Since $(\bar{f}_i, \bar{f}_i) = 1$, using (3.13) and (3.14) one gets

$$\frac{1}{c_{ii}^2} = \| \sum_{j=1}^{i-1} b_{ij}\bar{f}_j + f_i\|^2$$

$$= \|f_i\|^2 - \sum_{j=1}^{i-1} b_{ij}^2 \tag{3.15}$$

Substituting $\bar{f}_j = \sum_{p=1}^{j} c_{jp}f_p$ into (3.13), after some index manipulation, we have

$$\bar{f}_i = \sum_{j=1}^{i-1} \left(c_{ii} \sum_{p=j}^{i-1} b_{ip} c_{pj} \right) f_j + c_{ii} f_i \tag{3.16}$$

Comparing (3.16) with (3.2), we find

$$c_{ij} = c_{ii} \sum_{p=j}^{i-1} b_{ip} c_{pj} \quad , \quad j < i \tag{3.17}$$

Summarizing the above development, the computational formulae (for $i > 2$) are

$$b_{ij} = -\sum_{p=1}^{j} c_{jp}(f_i, f_p), \; j = 1, \ldots, i-1$$

$$c_{ii} = \left(\|f_i\|^2 - \sum_{j=1}^{i-1} b_{ij}^2 \right)^{-1/2} \tag{3.18}$$

$$c_{ij} = c_{ii} \sum_{p=j}^{i-1} b_{ip} c_{pj} \quad , \; j = 1, \ldots, i-1$$

The only required starting value is $c_{11} = 1/\|f_1\|$. This is a recursive algorithm. A similar recursive formula is given by Gaier (1987, p. 8). Gaier shows that $\bar{f}_i$ can be derived from

$$h_k = \det \begin{pmatrix} (f_1, f_1) \ldots (f_1, f_{k-1}) & f_1 \\ \cdots\cdots\cdots\cdots\cdots & \\ (f_k, f_1) \ldots (f_k, f_{k-1}) & f_k \end{pmatrix} \tag{3.19}$$

and

$$\bar{f}_k = h_k / \|h_k\| \tag{3.20}$$

Rearranging (3.19), we get

$$h_k = (-1)^{k-1} \det \begin{pmatrix} f_1 \ldots\ldots\ldots f_{k-1} & f_k \\ (f_1, f_1) \ldots\ldots (f_1, f_{k-1}) & (f_1, f_k) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ (f_{k-1}, f_1) \ldots\ldots (f_{k-1}, f_{k-1}) & (f_{k-1}, f_k) \end{pmatrix}$$

$$= (-1)^{k-1} \sum_{i=1}^{k} (-1)^{i-1} M_i f_i = \sum_{i=1}^{k-1} (-1)^{k+i-2} M_i f_i + A_{k-1} f_k \tag{3.21}$$

where M_i is the minor at row 1 and column i. In (3.21), the minor at row 1 and column k, i.e., M_k, has been denoted as A_{k-1} which is precisely the Gram determinant |G| for the first k-1 functions. Now, the norm of h_k can be expressed as (Gaier, 1987)

$$(h_k, h_k) = \det \begin{pmatrix} (f_1, f_1) \ldots (f_1, f_{k-1}) & (f_1, h_k) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ (f_k, f_1) \ldots (f_k, f_{k-1}) & (f_k, h_k) \end{pmatrix} \tag{3.22}$$

From (3.11), we see that all the the entries in the last column of (3.22) will vanish except (f_k, h_k). By the definition of A_{k-1} mentioned above and by (3.19), it is not difficult to see that $A_k = (f_k, h_k)$. Therefore, from (3.22) we have

$$(h_k, h_k) = A_k A_{k-1} \tag{3.23}$$

Using (3.21) and (3.24), we finally arrive at

$$\tilde{f}_i = \sum_{j=1}^{i-1} c_{ij} f_j + \sqrt{A_{k-1}/A_k}\, f_i \tag{3.24}$$

which shows that $c_{ii} = \sqrt{A_{k-1}/A_k}$. The rest of c_{ij} can be found using (3.19), (3.20) and (3.23).

The above development has been based on Gaier (1987). Mainville's algorithm and Gaier's algorithm are all based on recursive formulae. However, usually the computation of determinants (such as minors) is rather expensive, thus Mainville's algorithm should be chosen over Gaier's if one is to use a recursive algorithm.

A more efficient algorithm for computing c_{ij} is the Cholesky decomposition of matrix $G(f_1, \ldots, f_n)$ or the Gram matrix.

Since $y = Cx$, the relationship between the Gram matrices of y and x is (Davis, 1975, p. 177).

$$G(\tilde{f}_1, \ldots, \tilde{f}_n) = CG(f_1, \ldots, f_n)C^T \tag{3.25}$$

Now $G(\tilde{f}_1, \ldots, \tilde{f}_n)$ is an identity matrix of order n due to the orthonormality of $\tilde{f}_i$. Denoting $B = G(f_1, \ldots, f_n)$, we write (3.25) as

$$B = C^{-1}(C^{-1})^T \tag{3.26}$$

which shows that B is a product of a lower triangular matrix and its transpose whose diagonal elements are all positive. Since B is a positive definite matrix and its Cholesky decomposition is unique, we conclude that

$$C^{-1} = R^T \text{ or } C = (R^{-1})^T \tag{3.27}$$

where R^T is the lower triangular matrix from the Cholesky decomposition of $G(f_1, \ldots, f_n)$. Eq. (3.27) states that, to find the combination coefficients c_{ij}, all we need to do is to decompose $G(f_1, \ldots, f_n)$ (in the Cholesky sense) and find the inverse of the lower triangular matrix, R^T.

In the numerical experiments described in Chapter 6, both the recursive algorithm and the Cholesky decomposition will be used to compute c_{ij}. The comparisons on the computational efficiency and accuracy will be made in that particular chapter.

3.2.3 Relationship with QR Factorization

The aim of the development in Sections 3.2.1 and 3.2.2 is to derive a set of orthonormal functions over a continuous domain D in a Hilbert Space. It turns out that , for practical application, the domain D itself does not need to be continuously connected, instead it can contain only a finite number of points. Or, in the Lebesque integral, the concept of domain has been totally replaced by the concept of set.

Assume now the domain D consists of a finite number of points and the functions f_i's are defined only at those particular points. Denoting the number of points as m, we see that under this condition a function now can equivalently be represented as a vector of length m and we can write $f_i \in R^m$, i = 1, ..., n. (Again, for the moment, we deal with real functions only). A well-known matrix operation is the so-called QR factorization (Stewart, 1973) in which a matrix is factorized into a product of two matricies as follows:

$$A = QR \tag{3.28}$$

where $A \in R^{m \times n}$, $Q \in R^{m \times n}$, $R \in R^{n \times n}$, and $m \geq n$. Matrix A has n linearly independent columns, Q has n orthonormal columns and R is an upper triangular matrix whose diagonal elements are all positive. Let f_i's (now vectors) be the column vectors of A and the inner products take the definition according to (2.9), we have

$$Q = AR^{-1} \text{ or } Q^T = (R^{-1})^T A^T \tag{3.29}$$

It is easy to see that (3.29) is a discrete version of (3.7) (the reader should note the difference in notations and bear in mind that every element in vectors x and y is a function). Explicitly, we have

$$\begin{bmatrix} \bar{f}_1^T \\ \vdots \\ \bar{f}_n^T \end{bmatrix} = \begin{bmatrix} c_{11} & & \\ \vdots & \ddots & 0 \\ c_{n1} & \cdots & c_{nn} \end{bmatrix} \begin{bmatrix} f_1^T \\ \vdots \\ f_n^T \end{bmatrix} \tag{3.30}$$

Therefore, the orthonormalized (column) vectors are $\bar{f}_i$'s and the elements in $(R^{-1})^T$ are the combination coefficients. Nothing we have said in Sections 3.2.1 and 3.2.2 needs to be changed except the definitions of domain and inner product. However, the orthonormalizing process according to the QR factorization (or, the discrete version) will not be used here since

- Its impossible to derive a set of analytically defined orthonormal functions such as the Legendre polynomials $P_n(t)$ and the Tschebyscheff polynomial, $T_n(t)$ in [-1, 1] using such a discrete scheme.

- The domain is defined by a set of observation points and varies from case to case; therefore the resulting functions are not unique and lack physical meaning . In fact, the

concept of "points" is not necessarily required in geodetic observations so the interpretation becomes even more difficult.

Although the discrete orthormalizing process (i.e., QR factorization) is not to be used in this study, some of the concepts involved in the process can be very useful. For example, Goad and Mueller (1988) have employed the Gram determinant of the observation vectors to efficiently remove the correlation between GPS observables.

3.3 Method of Eigenvalue - Eigenfunction

3.3.1 The Helmholtz Equation and Its Solution

In Section 3.3, a method of obtaining orthogonal functions, subject to a specified boundary condition, will be discussed. First of all, we introduce the well-known Helmholtz equation (Page, 1955, p. 50):

$$\Delta u + ku = 0 \tag{3.31}$$

where Δ is the Laplace operator in E^1, E^2 or E^3, and k is a positive number (we restrict our study to a real u and a real k) associated with the "frequency" of the solution function u. The solutions of the Helmholtz equations for various domains and various boundary conditions is an important topic in physics, especially in the area of wave propagation. Standard treatment of this equation can be found in Morse and Feshbach (1953), Courant and Hilbert (1953), etc. The book by Miller (1977) even devoted two chapters to the solutions of this equation by the use of Lie algebra.

A value k_i, in (3.31) is called an eigenvalue associated with a solution u_i which is designated an eigenfunction. A solution of the Helmholtz equation, u_i, together with k_i, normally satisfies a boundary condition. Let σ be the domain where (3.31) holds, B the boundary of σ, and n the outer normal of B, normally three types of boundary condition (B.C.) are possible:

(1) $u = 0$ on B. The Dirichlet B.C.

(2) $\frac{\partial u}{\partial n} = 0$ on B. The Neumann B.C.

(3) $u = c \frac{\partial u}{\partial n}$ on B. The mixed B.C., c depends on the position on B only.

The relationship between σ, B and n is shown in Figure 3.1 (a 2-D planar case).

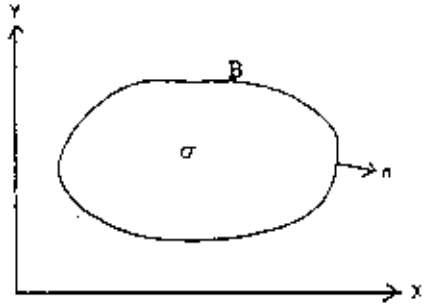


Figure 3.1 The Domain, Boundary and Outer Normal for a 2-D Helmholtz Equation

To illustrate the orthogonality of the eigenfunctions found from the solution of the Helmholtz equation, we consider a two-dimensional Laplacian (Δ) in a X-Y rectangular coordinate system:

$$\Delta u_i + k_i u_i = 0 \tag{3.32}$$

$$\Delta u_j + k_j u_j = 0 \tag{3.33}$$

Where $\Delta = \partial^2/\partial X^2 + \partial^2/\partial Y^2$, u_i and u_j are two solutions (eigenfunctions) associated with two eigenvalues, k_i and k_j. Multiplying (3.32) by u_j and (3.33) by u_i, then subtracting one from the other, we get

$$(k_i - k_j)u_i u_j = u_i \Delta u_j - u_j \Delta u_i \tag{3.34}$$

Taking the integrals on both sides of (3.34) and applying Green's second identity (see Appendix B) on a plane, we obtain

$$(k_i - k_j)\iint_\sigma u_i u_j d\sigma = \iint_\sigma (u_i \Delta u_j - u_j \Delta u_i) d\sigma$$

$$= \oint_B \left(u_i \frac{\partial u_j}{\partial n} - u_j \frac{\partial u_i}{\partial n} \right) ds \tag{3.35}$$

where ds is the differential arc length along B. Substituting any of the three B.C. into eq. (3.35), it is easy to see that

$$(k_i - k_j)\iint u_i u_j d\sigma = 0 \tag{3.36}$$

If $k_j \neq k_j$, then the integral part must vanish in order to make (3.36) true. Thus we show that two distinct eigenfunctions are mutually orthogonal. The eigenvalue-eigenfunction becomes degenerate if one eigenvalue corresponds to several eigenfunctions (Sommerfield, 1949). In such a degenerate case, any linear combination of these eigenfunctions is still a solution of the Helmholtz equation. Using the Gram-Schmidt process, we then can orthogonalize these eigenfunctions (of the same eigenvalue) and still maintain the orthogonality for all the eigenfunctions.

When k_j approaches k_i in (3.35), we can obtain the normalizing factor

$\iint_\sigma u_i^2 d\sigma$ using L'Hopital's rule as follows:

$$\iint_\sigma u_i^2 d\sigma = \lim_{k_j \to k_i} \iint_\sigma u_i u_j d\sigma$$

$$= \lim_{k_j \to k_i} \frac{1}{k_i - k_j} \oint_B \left(u_i \frac{\partial u_j}{\partial n} - u_j \frac{\partial u_i}{\partial n} \right) ds$$

$$= - \oint_B \left[u_i \frac{\partial}{\partial k} \left(\frac{\partial u}{\partial n} \right)_{k=k_i} - \frac{\partial u_i}{\partial n} \frac{\partial}{\partial k} (u)_{k=k_i} \right] ds \qquad (3.37)$$

In (3.37), we have considered u and $\partial u/\partial n$ as functions of the eigenvalue k. This is an important technique for deriving the normalizing factors of the generalized Fourier-Bessel functions and the spherical cap harmonics that will be described shortly.

It is worth mentioning that the 2-D Helmholtz equation which we have just discussed is a special case of the 2-D Sturm-Liouville problem (Courant and Hilbert, 1953):

$$L(u) + ku = \frac{\partial}{\partial x}\left(p \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y}\left(p \frac{\partial u}{\partial y} \right) - qu + ku = 0 \quad , \quad p(x, y) \geq 0 \qquad (3.38)$$

where L is an operator whose definition is clear in the above equation. The operator L in (3.38) is a self-adjoint operator which satisfies $(L(u), v) = (u, L(v))$, where u and v are two functions fulfilling any B.C. and are the solutions of (3.38). The orthogonality of the eigenfunctions is easier to derive by the properties of a self-adjoint operator, since

$$(L(u_i), u_j) = (u_i, L(u_j)) \Leftrightarrow$$

$$(-k_j u_i, u_j) = (u_i, -k_j u_j) \Leftrightarrow$$

$$(k_i - k_j)(u_i, u_j) = 0 \qquad (3.39)$$

Clearly (3.39) is equivalent to (3.36). Putting $p = 1$, $q = 0$ in (3.37), one immediately gets (3.32) or (3.33).

The set of eigenfunctions from the solution of the Helmholtz equation (or in general, the eigenvalue-eigenfunction problem) are complete in the sense of (2.20) (Courant and Hilbert, 1953). Thus any square integrable functions over σ may be expanded into eigenfunctions. The proof of the completeness of the eigenfunctions and the derivations of expansion theorem can be found in Courat and Hilbert (1953, pp. 424-427) or Toslov (1976, pp. 255-258).

So far we have only discussed the properties of eigenfunctions. We shall now proceed to solve for the Helmholtz equation and derive the analytic forms of the

eigenfunctions. The solution of (3.32) depends on the shape of the domain and the type of B.C. Upon knowing the shape and the B.C., normally the technique of separation of variables is employed to solve the partial differential equation. The solution u is factorized into a product of functions according to the chosen variables. When a successful separation of variables is achieved, the Helmholtz equation is then split into a set of second order ordinary differential equations which are then solved individually. To give an example, we consider the 3-D Helmholtz equation in spherical coordinates (r, θ, λ). For the boundary condition, we require that the solution u should vanish at the boundary, i.e., on the surface of the unit sphere $r = 1$ (so now the boundary B is defined by $r = 1$). It is well-known that (see also Appendix B) the Laplacian Δ in an curvilinear coordinate system is

$$\Delta = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \left(\sqrt{g}\, g^{ij} \frac{\partial}{\partial x^j} \right) \tag{3.40}$$

where the definitions of g, g^{ij} can be found in Appendix B. x^i are the curvilinear coordinates which also include the ordinary rectangular coordinates. The contravariant tensor g^{ij} is the inverse of the covariant tensor g_{ij}. In the Euclidean Space, g_{ij} is obtained from

$$[g_{ij}] = J^T J \tag{3.41}$$

where J is Jacobian from the curvilinear coordinate system, x^i, to the rectangular coordinate system, $\bar{x}^i$. Specifically,

$$J_{3x3} = \left[\frac{\partial \bar{x}^i}{\partial x^j} \right] \tag{3.42}$$

The reader may find (3.40), (3.41) and (3.42) in an ordinary textbook of tensor calculus (e.g., Kay, 1988; Synge and Schild, 1978). If the system x^i is orthogonal then both $[g_{ij}]$ and $[g^{ij}]$ become diagonal matrices. For an orthogonal system x^i, the arc length element is usually denoted as

$$(ds)^2 = h_1^2 (dx^1)^2 + h_2^2 (dx^2)^2 + h_3^2 (dx^3)^2 \tag{3.43}$$

In such a system, the following relationships hold:

$$g_{ij} = g^{ij} = 0, \qquad i \neq j \tag{3.44}$$

$$\begin{aligned} g_{11} &= \frac{1}{g^{11}} = h_1^2 \\ g_{22} &= \frac{1}{g^{22}} = h_2^2 \\ g_{33} &= \frac{1}{g^{33}} = h_3^2 \end{aligned} \tag{3.45}$$

Now $(x^1, x^2, x^3) = (r, \theta, \lambda)$ and $\bar{x}^1 = r\sin\theta\cos\lambda$, $\bar{x}^2 = r\sin\theta\sin\lambda$, $\bar{x}^3 = r\cos\theta$, thus it is easy to see that (using (3.41) and (3.42))

$$[g_{ij}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{bmatrix} \tag{3.46}$$

Substituting the results from (3.44), (3.45) and (3.46) into (3.40), we get the Helmholtz equation in spherical coordinates

$$\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\cot\theta}{r^2}\frac{\partial u}{\partial \theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 u}{\partial \lambda^2} + ku = 0 \tag{3.47}$$

If $k = 0$, (3.47) reduces to Laplace's equation in spherical coordinates (Heiskanen and Moritz, 1967, p. 19). Substituting $u = R(r)\,T(\theta)\,S(\lambda)$ into (3.47), after some algebra we get

$$\frac{1}{R}(r^2R'' + 2rR') + kr^2 = -\frac{1}{T}(\cot\theta T' + T'') - \frac{1}{S\sin^2\theta}S'' \tag{3.48}$$

In (3.48), the expression on the left side depends only on r and the expression on the right-hand side on θ and λ, thus we can separate (3.48) in the same way as we do for Laplace's equation (ibid., p. 20). Let the separation constant be $n(n + 1)$, so that we have

$$r^2R'' + 2rR' + [kr^2 - n(n+1)]R = 0 \tag{3.49}$$

$$\frac{\sin\theta}{T}(\sin\theta T'' + \cos\theta T') + n(n+1)\sin^2\theta = \frac{-S''}{S} \tag{3.50}$$

The solutions for (3.50) are the usual Laplace surface spherical harmonics (ibid, p. 21) if n is an integer. For the solution R(r), we substitute

$$R = \frac{Z}{\sqrt{r}}$$

$$R' = \frac{Z'}{\sqrt{r}} - \frac{Z}{2r\sqrt{r}} \tag{3.51}$$

$$R'' = \frac{Z''}{\sqrt{r}} - \frac{Z'}{r\sqrt{r}} + \frac{3}{4}\frac{Z}{r^2\sqrt{r}}$$

into (3.49), then we get

$$Z'' + \frac{1}{r}Z' + \left[k - \frac{\left(n+\frac{1}{2}\right)^2}{r^2}\right]Z = 0 \tag{3.52}$$

Eq. (3.52) is equivalent to (C.2) in Appendix C (note: in (C.2) k^2 is used instead of k, but this is not important). All possible solutions for (3.52) can be found in (Morse and

Feshbach, 1953, p. 1465). One solution of (3.52) is the Bessel function of the first kind $J_{n+1/2}(\sqrt{k}r)$, $k \neq 0$. With this particular solution, the solution for R then becomes

$$R(r) = \frac{J_{n+1/2}(\sqrt{k}r)}{\sqrt{r}} \tag{3.53}$$

As shown in Appendix C, if $\sqrt{k_i}$, i = 1, ... , are the zeroes of $J_{n+1/2}(r)$, then $J_{n+1/2}(\sqrt{k_i}r)$ form an orthogonal series. Let $Y(\theta, \lambda)$ be the Laplace surface spherical harmonics, i.e. the solution for (3.50), we can show that (see also Appendix C)

$$\int_\sigma \int_0^1 \frac{J_{n+1/2}(\sqrt{k_i}r)}{\sqrt{r}} \frac{J_{n+1/2}(\sqrt{k_j}r)}{\sqrt{r}} Y^2 r^2 dr d\sigma \tag{3.54}$$

$$= \int_0^1 J_{n+1/2}(\sqrt{k_i}r) J_{n+1/2}(\sqrt{k_j}r)\, r dr \cdot \int_\sigma Y^2 d\sigma$$

$$= c\delta_{ij} \quad , \quad c \text{ a constant}$$

Thus the solution u_i for (3.47) (index i indicates the sequence of the eigenvalue k in (3.47)) are orthogonal over the domain (now the interior of the unit sphere). Note that in (3.54), σ is the surface of the unit sphere and $r^2 d\sigma$ is the surface element of the sphere.

It is also clear that the solution

$$u_i(r, \theta, \lambda) = \frac{J_{n+1/2}(\sqrt{k_i}r)}{\sqrt{r}} Y(\theta, \lambda) \tag{3.55}$$

satisfies the B.C., since

$$u_i(1, \theta, \lambda) = J_{n+1/2}(\sqrt{k_i}) Y(\theta, \lambda) = 0 \tag{3.56}$$

by the definition of $\sqrt{k_i}$ (the zeros of $J_{n+1/2}(r)$).

Although we have shown the technique of separation of variables for the Helmholtz equation in the above development, the example is probably too simple to bring out the difficulty that we will face in constructing the orthonormal functions over the oceans. The difficulty is of course the separation of variables itself. Morse and Feshbach (ibid.) tabulated 11 separable coordinates in E^3 for the Helmholtz equation. Page (1955, Chapter VI) also has a similar table. But none of the tables immediately gives the desired separation of variables for the extremely irregular oceanic domain. What we can do for this study is to take the ideas of the solutions for the Helmholtz equation to develop a set of orthonormal functions over a "regularized" domain on the surface of the earth, such as a spherical cap or a region bounded by two parallels and two meridians. This will be demonstrated shortly.

If the domains of interest lie on a plane, the solutions of the Helmholtz equation for a rectangle, circle or ellipse are readily shown in Courant and Hilbert (1953), Morse and Feshbach (1953), Kaplan (1981), etc. These solutions thus provide the set of orthonormal functions for the corresponding domains. For the solution of the Helmholtz equation in a relatively "irregular" domain, we consider the following example. Suppose now the domain is the isosceles shown in Figure 3.2. Using a rectangular coordinate system, (3.31) becomes

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + ku = 0 \tag{3.57}$$

The boundary condition is

$$u = 0 \ , \quad \text{on the boundary of the isosceles} \tag{3.58}$$

The dilemma in this case is that a separation of variables for such a domain is not possible, so a solution of the form $u(x,y) = f(x)h(y)$ satisfying the B.C. cannot be obtained. To overcome this difficulty, we may linearly combine the solutions for the square ($0 \le x \le a$, $0 \le y \le a$), i.e., $\sin\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{a}y\right)$, to get a desired solution. Morse and Feshbach (ibid., p. 756) show empirically that the function

$$u_{mn} = \sin\left[\frac{\pi}{a}(m+n)x\right]\sin\left[\frac{\pi}{a}ny\right] - (-1)^m \sin\left[\frac{\pi}{a}(m+n)y\right]\sin\left[\frac{\pi}{a}nx\right] \tag{3.59}$$

satisfies (3.57) and (3.58).

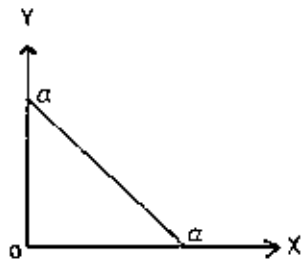


Figure 3.2 An Isosceles for the Helmholtz Equation

As one can verify, the functions u_{mn} are indeed orthogonal (with respect to the area integral) in the isosceles. Since u_{nm} are the solutions, we can find the eigenvalue k_{mn} in (3.57) from the relationship

$$k_{mn} = -\frac{\frac{\partial^2 u_{mn}}{\partial x^2} + \frac{\partial^2 u_{mn}}{\partial y^2}}{u_{mn}} = \left(\frac{\pi}{a}\right)^2\left[(m+n)^2 + n^2\right] \tag{3.60}$$

3.3.2 Eigenfunctions Orthogonal on a Portion of the Unit Sphere

3.3.2.1 Hypergeometric Functions and Legendre's Functions

In order to be prepared for the development of the spherical cap harmonics, we devote this section to the relationship between the hypergeometric functions and Legendre's functions.

The hypergeometric function is the solution of the second order differential equation (Lebedev, 1972, p. 162):

$$z(1 - z)\, u'' + [c - (a + b + 1)z]u' - abu = 0 \tag{3.61}$$

where a, b, and c are parameters which can take any real or complex values. The independent variable z can also be real or complex. To obtain a solution for (3.61), we assume the solution u can be expressed in a Gaussian series representation:

$$u = z^s \sum_{n=0}^{\infty} e_n z^n \tag{3.62}$$

where the constant s and the coefficients e_n are to be determined. Substituting (3.62) into (3.61), two possible solutions can be obtained (Lebedev, 1972, p. 162; Sommerfeld, 1949, p. 152):

$$u_1 = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{n!(c)_n} z^n \tag{3.63}$$

$$u_2 = z^{1-c} \sum_{n=0}^{\infty} \frac{(a - c + 1)_n (b - c + 1)_n}{n!(2 - c)_n} z^n \tag{3.64}$$

where $(a)_n$ is the "shifted factorial" defined as

$$(a)_0 = 1$$

$$(a)_n = a(a + 1) \ldots (a + n - 1),\ n = 1, 2, \ldots \tag{3.65}$$

Solutions u_1 and u_2 are usually denoted as

$$u_1 = F(a, b; c; z) \tag{3.66}$$

$$u_2 = z^{1-c} F(a - c + 1, b - c + 1;\ 2 - c; z) \tag{3.67}$$

Comparing (3.66) and (3.63), one immediately finds the property

$$F(a, b; c; z) = F(b, a; c; z) \tag{3.68}$$

Although two solutions for the hypergeometric equation are available, we shall just discuss the first one, i.e., u_1. Since

$$e_n = \frac{(a)_n (b)_n}{n!(c)_n} \tag{3.69}$$

one gets the recursive relationship for the coefficients as follows

$$e_{n+1} = \frac{(a + n)(b + n)}{(n + 1)(c + n)} e_n \tag{3.70}$$

In addition to the hypergeometric functions arising from the solutions for (3.61), one can also define the generalized hypergeometric function as (Miller, p. 271, 1977)

$${}_pF_q\left(\begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix} ; z \right) = \sum_{n=0}^{\infty} \frac{(a_1)_n \ldots (a_p)_n}{(b_1)_n \ldots (b_q)_n} \frac{z^n}{n!} \tag{3.71}$$

By such a definition, it is clear that

$$u_1 = {}_2F_1\left(\begin{matrix} a, b \\ c \end{matrix} ; z \right) \tag{3.72}$$

It is not our intention here to pursue the subject of the hypergeometric function, rather we would like to use the existing form of the hypergeometric function to obtain a general representation of the Legendre function of any "degree". To this end, we make the substitutions

$$\begin{aligned} z &= \frac{1 - t}{2} \\ a &= -\ell \\ b &= \ell + 1 \\ c &= 1 \\ P_\ell(t) &= u\left(\frac{1 - t}{2}\right) \end{aligned} \tag{3.73}$$

in (3.61). Thus we find

$$(1 - t^2)P''_\ell - 2tP'_\ell + \ell(\ell + 1)P_\ell = 0 \tag{3.74}$$

which is Legendre's equation (Hobson, 1965, p. 10). Note that ℓ does not need to be a nonnegative integer in (3.74). Since Legendre's equation is obtained by a simple change of variable in the differential equation associated with the hypergeometric function, its solution must assume the same form of the hypergeometric function. Using (3.66) and (3.73), we can express Legendre's function as

$$P_\ell(t) = F\left(-\ell,\ \ell + 1;\ 1\ ;\ \frac{1 - t}{2}\right) \tag{3.75}$$

When ℓ is an integer, one can easily verify that (using (3.65)) the coefficients e_n in (3.63) are zero if $n > \ell$. In this case, $P_\ell(t)$ becomes a polynomial of degree ℓ.

For the related representation of the associated Legendre function, we differentiate (3.74) m times with respect to t and apply the rule (an application of Leibniz's rule)

$$\begin{aligned}
\frac{d^m}{dt^m}\left(tP_\ell'\right) &= t\frac{d^{m+1}}{dt^{m+1}}(P_\ell) + m\frac{d^m}{dt^m}(P_\ell) \\
\frac{d^m}{dt^m}\left(t^2P_\ell''\right) &= t^2\frac{d^{m+2}}{dt^{m+2}}(P_\ell) + 2mt\frac{d^{m+1}}{dt^{m+1}}(P_\ell) + m(m-1)\frac{d^m}{dt^m}(P_\ell)
\end{aligned} \tag{3.76}$$

Thus we get

$$\left\{(1 - t^2)\frac{d^2}{dt^2} - 2(m + 1)t\frac{d}{dt} + \left[\ell(\ell + 1) - m(m + 1)\right]\right\}\frac{d^m}{dt^m}(P_\ell) = 0 \tag{3.77}$$

Now the differential equation corresponding to the associated Legendre function is (Hobson, 1965, p. 89)

$$\left\{(1 - t^2)\frac{d^2}{dt^2} - 2t\frac{d}{dt} + \ell(\ell + 1) - \frac{m^2}{1 - t^2}\right\}P_\ell^m(t) = 0 \tag{3.78}$$

If we make the substitution

$$P_\ell^m(t) = (1 - t^2)^{m/2}v(t) \tag{3.79}$$

in (3.78), then we have

$$\left\{(1 - t^2)\frac{d^2}{dt^2} - 2(m + 1)t\frac{d}{dt} + \left[\ell(\ell + 1) - m(m + 1)\right]\right\}v(t) = 0 \tag{3.80}$$

Clearly (3.80) is equivalent to (3.77). Thus we can say that v satisfies the differential equation of $P_\ell^{(m)}$, where $P_\ell^{(m)} = \frac{d^m}{dt^m}(P_\ell)$. In other words, v is equal to $P_\ell^{(m)}$ up to a constant factor. Thus we may define the associated Legendre function as

$$P_\ell^m(t) = (1 - t^2)^{m/2}\frac{d^m}{dt^m}(P_\ell) \tag{3.81}$$

The definition of P_ℓ^m according to (3.81) is consistent with Heiskanen and Moritz (1967, p. 23), except that P_ℓ^m in (3.81) could have a non-integer "degree" ℓ due to the use of the hypergeometric function.

Recalling the property of the hypergeometric function (Lebedev, 1972, p. 241)

$$\frac{d}{dz}F(a, b;\ c;\ z) = \frac{ab}{c}F(a + 1, b + 1;\ c + 1;\ z) \tag{3.82}$$

we finally get the desired form of P_ℓ^m:

$$P_\ell^m(t) = (1 - t^2)^{m/2}\frac{d^m}{dt^m}F\left(-\ell, \ell + 1;\ 1;\ \frac{1 - t}{2}\right)$$

$$= \frac{(-1)^m}{2^m}(1 - t^2)^{m/2}\frac{(-\ell)_m(-\ell + 1)_m}{m!}F\left(m - \ell, m + \ell + 1;\ m + 1;\ \frac{1 - t}{2}\right) \tag{3.83}$$

where the definition of $(-\ell)_m$, etc. can be found in (3.65). Furthermore, Lebedev (1972, p. 197) shows that

$$(-\ell)_m = (-1)^m\frac{\Gamma(\ell + 1)}{\Gamma(\ell - m + 1)} \tag{3.84}$$

$$(\ell + 1)_m = \frac{\Gamma(\ell + m + 1)}{\Gamma(\ell + 1)} \tag{3.85}$$

where $\Gamma(z)$ is the gamma function (1972, p. 1) defined by

$$\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt\ ,\quad z > 0 \tag{3.86}$$

Therefore, the associated Legendre function can be expressed as

$$P_\ell^m(t) = \frac{1}{2^m m!}\frac{\Gamma(\ell + m + 1)}{\Gamma(\ell - m + 1)}(1 - t^2)^{m/2}F\left(m - \ell, m + \ell + 1;\ m + 1;\ \frac{1 - t}{2}\right)$$

$$m = 0, 1, 2, \ldots, \infty\ ;\ \ell \text{ arbitrary} \tag{3.87}$$

Now we shall discuss two cases:

(1) When ℓ is a non-negative integer. In such a case $\Gamma(\ell + m + 1) = (\ell + m)!$, $\Gamma(\ell - m + 1) = (\ell - m)!$, and (3.87) reduces to

$$P_\ell^m(t) = \frac{1}{2^m m!}\frac{(\ell + m)!}{(\ell - m)!}(1 - t^2)^{m/2}F\left(m - \ell, m + \ell + 1;\ m + 1;\ \frac{1 - t}{2}\right) \tag{3.88}$$

In (3.88), if $m > \ell$, then according to (3.81) $P_\ell^m(t) = 0$. If $\ell > m$, $F(m - \ell, m + \ell + 1;\ m + 1;\ (1 - t)/2)$ is a polynomial of degree $(\ell - m)$. The second statement can be verified by using (3.63) and (3.69). Therefore, formula (3.88) is also valid for Legendre's function of the first kind in Heiskanen and Moritz (1967, p. 22) and holds for the closed interval [-1, 1].

(2) When ℓ is not an integer. In such a case we may express $P_\ell^m(t)$ as an infinite series in t using (3.87) and (3.63), namely

$$P_\ell^m(t) = \sum_{k=0}^{\infty} a_k \left(\frac{1-t}{2}\right)^k \tag{3.89}$$

where coefficients a_k can be found by (3.70):

$$a_o = \frac{1}{2^m m!} \frac{\Gamma(\ell + m + 1)}{\Gamma(\ell - m + 1)} (1 - t^2)^{m/2}$$

$$a_k = \frac{(m - \ell + k - 1)(m + \ell + k)}{k(m + k)} a_{k-1} \tag{3.90}$$

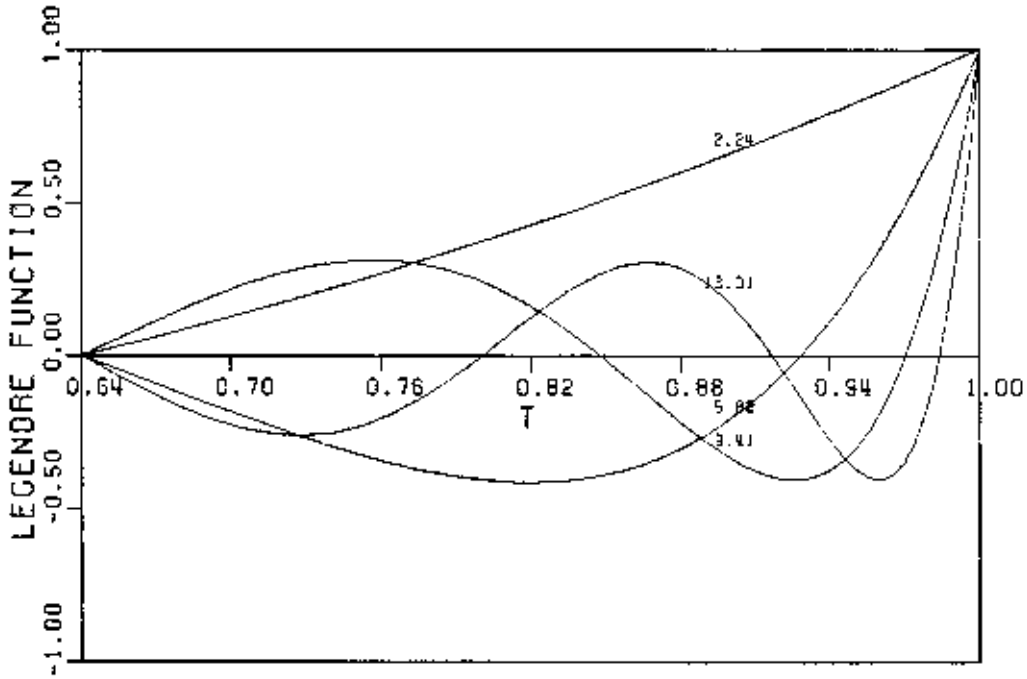
To numerically evaluate (3.89), we may define a small positive number ε, according to the desired accuracy, and set the truncated sum S_N for a fixed ℓ and a fixed m as

$$S_N = \sum_{k=0}^{N} a_k \left(\frac{1-t}{2}\right)^k \tag{3.91}$$

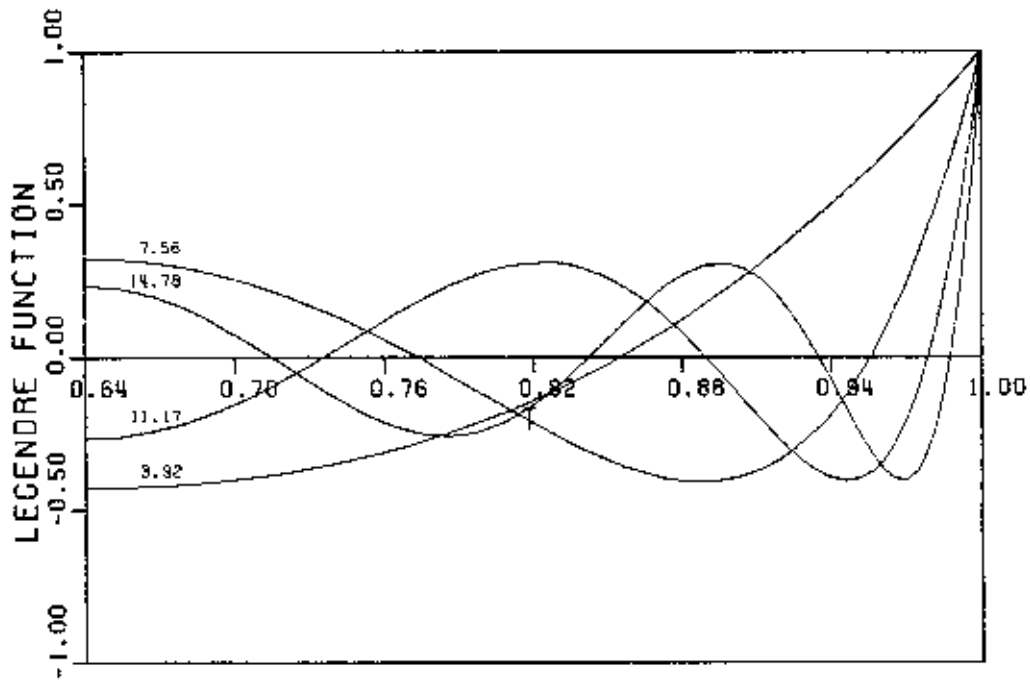
If $|S_{N+1} - S_N| \le \varepsilon$, then the summation is terminated at N.

In case of non-integer ℓ, an important issue concerning the convergence of the series in (3.89) at the two end points $t = \pm 1$ must be discussed. At $t = 1$, the series in (3.89) apparently converges and especially $P_\ell(1) = 1$. However, at $t = -1$, the series in (3.89) diverges. Therefore the associated Legendre function of non-integer degree cannot be regular throughout the closed interval [-1, 1]. In potential theory, normally we require that the solution for Laplace's equation be regular throughout the specified domain, thus the associated Legendre function of non-integer degree cannot satisfy the need for the entire sphere. This is an important point readily made in Courant and Hilbert (1953) and Sommerfeld (1949, p. 155). Nevertheless, for the spherical cap harmonic discussed in the next section, the domain is only extended from $t = 1$ to a point before $t = -1$, thus our study on that particular function will not be affected by the singularity at $t = -1$.

In connection with the discussion in the next section, we calculate two groups of P_ℓ^m using (3.89) and (3.90) and plot the functions in Figure 3.3. In this figure, $t_o \le t \le 1$, and $t_o = \cos(50°)$. The first group, i.e., the one corresponding to Figure 3.3(a), contains the P_ℓ^m which vanish at t_o, i.e., $P_\ell^m(t_o) = 0$. The second group, i.e., the one corresponding to Figure 3.3(b), contains the P_ℓ^m whose derivatives vanish at t_o, i.e., $\frac{d}{dt} P_\ell^m(t_o) = 0$. The ℓ values in Figure 3.3 have been chosen from Haines (1985a, Table 1). Within each group, the functions P_ℓ^m are orthogonal in $[t_o, 1]$. The orthogonality property will be addressed in the next section.



(a)



(b)

Figure 3.3 Legendre Functions of Non-integer Degree, m = 0 for all Curves, ℓ Plotted Next to the Curves, cos 50° ≤ t ≤ 1.

In these calculations, we have set $\varepsilon = 10^{-14}$. The N value in (3.91) varies from 1 to 50. Typically, 0.0005 CPU second is needed to evaluate one single $P_\ell^m(t)$ for such an ε value on the IBM 3081 machine. The gamma function is calculated by the IMSL routine GAMMA in a double precision mode.

To conclude this section, we remark that the substitution in (3.73) is not the unique one for the transformation between the hypergeometric equation and Legendre's equation. As shown in Lebedev (1972, p. 164), we will obtain Legendre's function of the second kind if we set $z = t^{-2}$, $a = \ell/2 + 1$, $b = \ell/2 + 1/2$, $c = \ell + 3/2$, $Q_\ell(t) = t^{-\ell-1}u(t^{-2})$ in (3.61). However, to get a convergent series representation for Legendre's function of the second kind, we need to have the condition $|t| > 1$, thus such a function is not suitable for the studies pursued here.

3.3.2.2 Spherical Cap Harmonics

In this section, we discuss two orthogonal systems over a spherical cap. The formulae developed in the previous section will be used here. These sets are solutions of Laplace's equation for the spherical cap.

Laplace's equation $\Delta u = 0$ in spherical coordinates is (Hobson, 1965, p. 9)

$$\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + \frac{1}{T\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dT}{d\theta}\right) + \frac{1}{\sin^2\theta S}\frac{d^2S}{d\lambda^2} = 0 \tag{3.92}$$

where, as usual, $u = R(r)T(\theta)S(\lambda)$ and r, θ, λ are the spherical coordinates. If only a separation of r is attempted, i.e., $u = R(r)Y(\theta, \lambda)$, we have another form of Laplace's equation (Heiskanen and Moritz, 1967, p. 20):

$$\frac{1}{R}\left(r^2R'' + 2rR'\right) = \frac{-1}{Y\sin\theta}\left[\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial Y}{\partial\lambda^2}\right] \tag{3.93}$$

The only chance for (3.93) to be true is that both sides of (3.93) are equal to a constant. We denote the constant as k and get

$$r^2R'' + 2rR' - kR = 0 \tag{3.94}$$

$$\frac{1}{\sin\theta}\left[\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y}{\partial\theta}\right) + \frac{1}{\sin\theta}\frac{\partial^2Y}{\partial\lambda^2}\right] + kY = 0 \tag{3.95}$$

The separation constant k is arbitrary. One choice is $k = \ell(\ell + 1)$ where ℓ can be any real number. In fact, if the solution for (3.94) has the form

$$R = Ar^{\alpha 1} + Br^{\alpha 2} \tag{3.96}$$

then α_1 and α_2 are the roots of the quadratic equation (Courant and Hilbert, 1953, p. 316)

$$\alpha(\alpha + 1) = k \tag{3.97}$$

What we are really interested in is the solution of (3.95), which, according to Appendix B, can be expressed as

$$\Delta^* Y + kY = 0 \tag{3.98}$$

The notation Δ^* is introduced by Courant and Hilbert (1953, p. 317). We shall call Δ^* the "Laplace surface operator" in accordance with Appendix B. Due to its similarity with (3.31), (3.98) may be regarded as the spherical "Helmholtz" equation on the surface of the unit sphere with k being the eigenvalue.

Assume now Y is a function of θ, λ, defined in a region (or domain) σ on the unit sphere. The relationship between σ and its boundary B and outer normal n is given in Figure 3.4. Note that the outer normal n, as shown in Figure B.1 (Appendix B), lies on the tangential plane passing a point located on the unit sphere. Furthermore, we assume that Y satisfies (3.98) and any of the B.C. described in Section 3.3.1. According to the proof in Appendix B, if Y_i and Y_j, together with their eigenvalues k_i and k_j, are two such functions, we have

$$\begin{aligned}(k_i - k_j)\iint_\sigma Y_i Y_j d\sigma &= \iint_\sigma \left(Y_i \Delta^* Y_j - Y_j \Delta^* Y_i\right) d\sigma \\ &= \oint_B \left(Y_i \frac{\partial Y_j}{\partial n} - Y_j \frac{\partial Y_i}{\partial n}\right) ds \\ &= 0\end{aligned} \tag{3.99}$$

Thus

$$\iint_\sigma Y_i Y_j d\sigma = 0, \quad \text{if } k_i \neq k_j \tag{3.100}$$

Eq. (3.100) states that, as long as a sequence of Y_i and k_i which satisfy (3.98) and the B.C. in a specified domain σ can be found, then we automatically find a set of orthogonal functions in the domain σ.

Now we consider a special case in which σ is a spherical cap, as shown in Figure 3.5. For such a domain we may separate Y in a usual manner, i.e. $Y = f(\theta)h(\lambda)$. From (3.95) we obtain

$$\frac{\sin\theta}{f} \frac{d}{d\theta}\left(f' \sin\theta\right) + k\sin^2\theta = -\frac{h''}{h} \tag{3.101}$$

Let $m^2 = -h''/h$, $t = \cos\theta$, $k = \ell(\ell + 1)$, and $v = v(t) = f(\theta)$, then

$$(1 - t^2)v'' - 2tv' + \left[\ell(\ell + 1) - \frac{m^2}{1 - t^2}\right]v = 0 \tag{3.102}$$

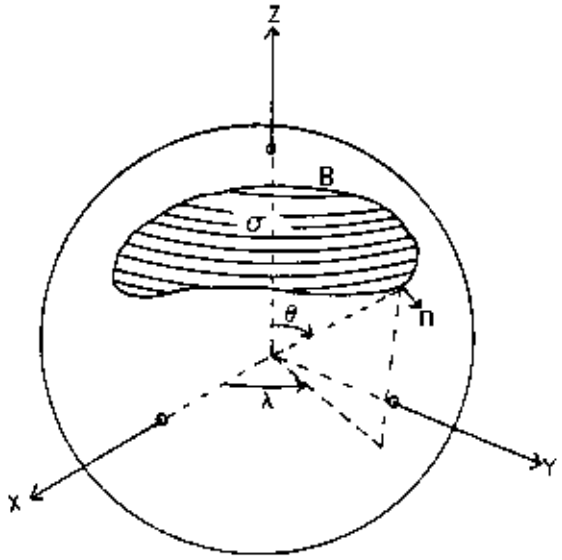


Figure 3.4 A Domain σ and its Boundary B and Outer Normal n on the Surface of the Unit Sphere.

$$h'' + m^2 h = 0 \tag{3.103}$$

Eq. (3.102) is equivalent to (3.78). Now our task is to find v(t) and h(λ) such that the product, i.e., Y will be orthogonal in the spherical cap. In finding such functions, one should always bear in mind what the B.C. is. If we write (3.102) as

$$\frac{d}{dt}\left[(1 - t^2)v'\right] + \left[\ell(\ell + 1) - \frac{m^2}{1 - t^2}\right]v = 0 \tag{3.104}$$

then (3.104) has the form of a one-dimensional Sturm-Liouville problem

$$L(v) + k\rho v = (pv')' - qv + k\rho v = 0 \tag{3.105}$$

Clearly for (3.104), we have

$$p(t) = 1 - t^2 \quad , \quad k = \ell(\ell + 1)$$
$$q(t) = \frac{m^2}{1 - t^2} \quad , \quad \rho = 1$$

However, when $\theta = 0°$, i.e., $t = 1$, then $p(t) = 0$ in (3.106), so (3.104) is a singular Sturm-Liouville problem. For some singular Sturm-Liouville problems certain solutions still exist. One obvious example is the Legendre polynomial $P_\ell(t)$ (Heiskanen and Moritz, 1967, p. 22) which satisfies (3.104) with $m = 0$ and $p(t) = 0$ on both sides of the boundary.

Assume that v_i and v_j, together with ℓ_i and ℓ_j, are two functions satisfying (3.104). Upon multiplying one such equation by the other function, we get a pair of equations

$$v_j \frac{d}{dt}\left[(1 - t^2)v_i'\right] + v_j\left[\ell_i(\ell_i + 1) - \frac{m^2}{1 - t^2}\right]v_i = 0 \tag{3.107}$$

$$v_i \frac{d}{dt}\left[(1 - t^2)v_j'\right] + v_i\left[\ell_j(\ell_j + 1) - \frac{m^2}{1 - t^2}\right]v_j = 0 \tag{3.108}$$

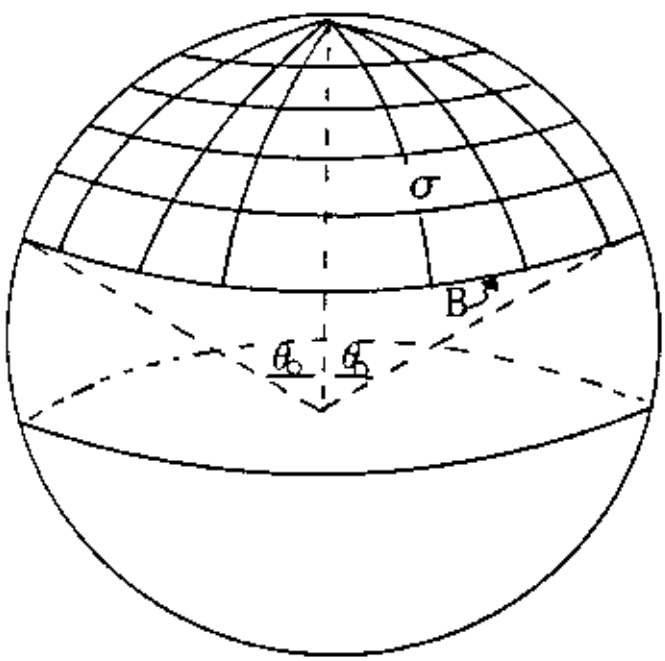


Figure 3.5 A Spherical Cap of Radius θ_0 with a Boundary Curve B

Subtracting (3.108) from (3.107) and integrating the resulting equation on both sides, we get

$$(\ell_j - \ell_i)(\ell_i + \ell_j + 1)\int_{t_o}^{t_1} v_i v_j dt = \int_{t_o}^{t_1} \left\{ v_j \frac{d}{dt}\left[(1 - t^2)v_i'\right] - v_i \frac{d}{dt}\left[(1 - t^2)v_j'\right]\right\} dt$$

$$= -(1 - t^2)\left(v_i v_j' - v_i' v_j\right)\Big|_{t_o}^{t_1} \tag{3.109}$$

In (3.109), integration by parts is used. Now, for a spherical cap,

$$t_o = \cos\theta^\circ$$

$$t_1 = 1 \tag{3.110}$$

If we change the notations to $P_{\ell_i}^m = v_i$, and $P_{\ell_j}^m = v_j$ we can write (3.109) as

$$(\ell_j - \ell_i)(\ell_i + \ell_j + 1)\int_{t_o}^{1} P_{\ell_i}^m P_{\ell_j}^m dt = \tag{3.111}$$

$$(1 - t_o^2)\left[P_{\ell_i}^m(t_o)\frac{dP_{\ell_j}^m(t_o)}{dt} - P_{\ell_j}^m(t_o)\frac{dP_{\ell_i}^m(t_o)}{dt}\right]$$

It is understood that, in (3.111),

$$\frac{dP_\ell^m(t_o)}{dt} = \frac{dP_\ell^m(t)}{dt}\Big|_{t=t_o} \tag{3.112}$$

Assuming that ℓ_i and ℓ_j are two distinct positive values, there are two cases in which the right side of (3.111) will vanish:

$$\text{Case 1 : } P_{\ell_i}^m(t_o) = 0 \text{ and } P_{\ell_j}^m(t_o) = 0 \tag{3.113}$$

$$\text{Case 2 : } \frac{dP_{\ell_i}^m(t_o)}{dt} = 0 \text{ and } \frac{dP_{\ell_j}^m(t_o)}{dt} = 0 \tag{3.114}$$

If ℓ_i and ℓ_j happen to be a pair of ℓ from (3.113) or (3.114) we get

$$(\ell_j - \ell_i)(\ell_i + \ell_j + 1)\int_{t_o}^{1} P_{\ell_i}^m(t) P_{\ell_j}^m dt = 0 \tag{3.115}$$

Thus $P_{\ell_i}^m$ and $P_{\ell_j}^m$ are orthogonal if $\ell_i \neq \ell_j$. Therefore, we have two groups of orthogonal functions in $[t_o, 1]$, arising from the choice of ℓ in case 1 and case 2. In other words, if we consider $P_\ell^m(t_o)$ and its derivative as two functions of ℓ, then by solving seperately the equations

$$P_\ell^m(t_o) = 0 \tag{3.116}$$

$$\frac{dP_\ell^m(t_o)}{dt} = 0 \tag{3.117}$$

we get two groups of orthogonal functions (the functions are orthogonal only within one group and with respect to the spherical cap integral). For this purpose, the representation of P_ℓ^m according to (3.87) or (3.89) is needed. Although analytical solutions for ℓ in (3.116) and (3.117) can be found in Hobson (1965, p. 409) and Pal (1920, p. 88 and p. 94), Haines (1985a) suggested a numerical method for finding the roots, namely, a root finding routine will be good enough to find the ℓ values. Haines (ibid., Table 1) also tabulated the first few ℓ values for m = 0, ... , 8, when $t_o = \cos 50°$. He also recommended a way of indexing the functions P_ℓ^m arising from the solutions (3.116) and (3.117). As stated before, the functions plotted in Figure 3.3 are based on the ℓ values in Haines' table, thus $P_\ell^m(t)$ vanish at t_o in Figure 3.3(a) and $dP_\ell^m(t_o)/dt$ vanish at t_o in Figure 3.3(b).

To find the normalizing factor for P_ℓ^m, we again apply L'Hospital's rule. By treating ℓ_j as a variable approaching ℓ_i, we have

Case 1 : ℓ such that $P_\ell^m(t_o) = 0$, m fixed

$$I_{\ell_i}^m = \int_{t_o}^{1} \left(P_{\ell_i}^m(t)\right)^2 dt = \lim_{\ell_j \to \ell_i} \frac{\left(1 - t_o^2\right)\left[P_{\ell_i}^m(t_o)\dfrac{dP_{\ell_j}^m(t_o)}{dt} - P_{\ell_j}^m(t_o)\dfrac{dP_{\ell_i}^m(t_o)}{dt}\right]}{\left(\ell_j - \ell_i\right)\left(\ell_i + \ell_j + 1\right)}$$

$$= \frac{\left(t_o^2 - 1\right)}{2\ell_i + 1}\frac{\partial}{\partial \ell}\left[P_{\ell_i}^m(t_o)\right]\frac{dP_{\ell_i}^m(t_o)}{dt} \tag{3.118}$$

Case 2 : ℓ such that $\dfrac{dP_\ell^m(t_o)}{dt} = 0$, m fixed

$$K_{\ell_i}^m = \int_{t_o}^{1} \left(P_{\ell_i}^m(t)\right)^2 dt = \frac{\left(1 - t_o^2\right)}{2\ell_i + 1} P_{\ell_i}^m(t_o)\frac{\partial}{\partial \ell}\left[\frac{dP_{\ell_i}^m(t_o)}{dt}\right] \tag{3.119}$$

To get $\dfrac{dP_\ell^m(t_o)}{dt}$, we can apply the recursive formulae for P_ℓ^m. These formulae are basically derived from the differential equation for P_ℓ^m, i.e., (3.78), thus they are applicable to any kind of ℓ values. For example, we can use (Lebedev, 1972, p. 195):

$$(t^2 - 1)\frac{dP_\ell^m(t)}{dt} = \ell t P_\ell^m(t) - (\ell + m)P_{\ell-1}^m(t) \tag{3.120}$$

Thus by this formula we need to evaluate P_ℓ^m and $P_{\ell-1}^m$ to get a derivative. To do this, the truncation series of (3.91) can still be applied.

It is relatively involved to get the $\frac{\partial}{\partial \ell}(P_\ell^m(t))$ and $\frac{\partial}{\partial \ell}\left(dP_\ell^m(t)/dt\right)$ values which are needed for the normalizing factors. If we take the partial derivative with respect to ℓ on both sides of (3.87) and (3.120), after considerable algebra we get

$$\left.\begin{aligned} \frac{\partial}{\partial \ell}(P_\ell^m(t)) &= \frac{(1 - t^2)^{m/2}}{2^m m!\Gamma(\ell - m + 1)\Gamma(m - \ell)} \sum_{k=0}^{\infty} \alpha_k \left(\frac{1 - t}{2}\right)^k \\ (t^2 - 1)\frac{\partial}{\partial \ell}\left(\frac{dP_\ell^m(t)}{dt}\right) &= tP_\ell^m - P_{\ell-1}^m + \ell t \frac{\partial}{\partial \ell}(P_\ell^m) - (\ell + m)\frac{\partial}{\partial \ell}(P_{\ell-1}^m) \end{aligned}\right\} \tag{3.121}$$

where

$$\alpha_k = \frac{\Gamma(m - \ell + k)\Gamma(m + \ell + 1 + k)}{k!(m + 1)_k}\left[\psi(m + \ell + 1 + k) - \psi(m - \ell + k) - \psi(\ell - m + 1) + \psi(m - \ell)\right] \tag{3.122}$$

and the ψ-function is (Lebedev, 1972)

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)} = \gamma + \sum_{n=0}^{\infty}\left(\frac{1}{n + 1} - \frac{1}{n + z}\right), \quad z > 0 \tag{3.123}$$

where $\gamma = 0.5772156$ is Euler's constant. If the argument of the gamma function or the ψ-function is negative, we should use the extended definitions:

$$\Gamma(z) = \frac{\Gamma(z + n)}{(z)_n}, \quad -n < z < -n + 1, \; n = 1, 2, \ldots \tag{3.124}$$

$$\psi(z) = \psi(z + n) - \sum_{k=o}^{n-1} \frac{1}{z + k}, \quad -n < z < -n + 1, \; n = 1, 2, \ldots \tag{3.125}$$

where $(z)_n$ can be found in (3.65). Eq. (3.124) is due to Widder (1989); (3.125) is due to the fact that $\psi(z + 1) = 1/z + \psi(z)$ and

$$\frac{d}{dz}[(z)_n] = (z)_n \sum_{k=1}^{n-1} \frac{1}{z+k}$$

The IMSL routines GAMMA and PSI may be used to evaluate the gamma function and the ψ-function, respectively.

If the related subroutines are hard to find, we may use numerical differentiations to get the desired derivatives. In Table 3.1, we compare the $\partial/\partial\ell(P_\ell^m(t_o))$ values at 4 pairs of ℓ, m using the analytical formulae in (3.121) and the numerical differentiation by the IMSL routine DERIV. The number of terms needed to achieve an accuracy of 10^{-14} is also listed in Table 3.1. These pairs of ℓ, m are taken from Haines (ibid.).

Table 3.1 $\partial/\partial\ell(P_\ell^m)$ From Analytic Formula and Numerical Differentiation at $t_o = \cos 50°$

ℓ	m	analytical	numerical	cpu1†	cpu2	N*
2.24	0	-0.480796	-0.480796	0.0095	0.0048	18
5.82	0	0.316496	0.316496	0.0095	0.0089	19
3.92	1	-1.659279	-1.659279	0.0092	0.0100	19
7.56	1	2.253507	2.253507	0.0093	0.0057	19

† cpu1 : analytical, cpu2 : numerical, unit in CPU seconds (IBM 3081)
* N : number of terms in (3.121) for an accuracy of 10^{-14}

From Table 3.1, we can see that the results are exactly the same up to the sixth decimal place. If we decrease the error tolerance in the numerical differentiation to 10^{-14}, one can expect the same results up to the 14th decimal place.

As far as the solution for (3.103) is concerned, we have the classical Fourier functions

$$h(\lambda) = \sin m\lambda \text{ or } h(\lambda) = \cos m\lambda \tag{3.127}$$

With $h(\lambda)$ in (3.127) and the associated Legendre functions of non-integer degree in the above development, we define two sets of normalized spherical cap harmonics as follows:

Set 1 : $P_\ell^m(t_o) = 0$

$$\begin{Bmatrix} \overline{R}_\ell^m(\theta, \lambda) \\ \overline{S}_\ell^m(\theta, \lambda) \end{Bmatrix} = \sqrt{\frac{(2 - \delta(m))(1 - \cos\theta_o)}{I_\ell^m}}\, P_\ell^m(\cos\theta) \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} \tag{3.128}$$

Set 2 : $dP_\ell^m(t_o)/dt = 0$

$$\begin{Bmatrix} \overline{U}_\ell^m(\theta,\lambda) \\ \overline{V}_\ell^m(\theta,\lambda) \end{Bmatrix} = \sqrt{\frac{(2-\delta(m))(1-\cos\theta_o)}{K_\ell^m}}\, P_\ell^m(\cos\theta) \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} \tag{3.129}$$

where $\delta(m)$ is the Dirac delta function defined as

$$\delta(m) = \begin{cases} 1, & m = 0 \\ 0, & m \neq 0 \end{cases} \tag{3.129}$$

The normalizing factors I_ℓ^m and K_ℓ^m can be found in (3.118) and (3.119). Within each set, the functions are orthonormal with respect to the integral over the domain $0 \le \theta \le \theta_o$, $0 \le \lambda \le 2\pi$. The area of this domain is $2\pi(1 - \cos\theta_o)$.

3.3.2.3 Generalized Fourier-Bessel Series on the Unit Sphere

In a spherical coordinate system, r, θ, λ, the surface of a sphere becomes "flat" when θ is relatively small. In such a "flat" area of a sphere, or, in the "polar" region, the Legendre functions behave like functions on a plane, such as the Bessel functions. To see this, we list Mehler's formula (Hobson, 1965, p. 317 and Miller, 1977):

$$\lim_{n\to\infty} P_n\left(\cos\frac{\theta}{n}\right) = J_o(\theta) \tag{3.130}$$

Another example is Macdonald's formula (Hobson, 1965, p. 406):

$$\begin{aligned} P_n^{-m}(\cos\theta) = \frac{1}{\left(n\cos\frac{1}{2}\theta\right)^m} \Big\{ & J_m(x) - \sin^2\tfrac{1}{2}\theta J_{m+1}(x) - \sin^2\tfrac{1}{2}\theta\left[\tfrac{1}{2}J_{m+2}(x) - \right. \\ & \left.\tfrac{1}{6}xJ_{m+3}(x)\right] - \sin^2\tfrac{1}{2}\theta\left[\tfrac{2}{x}J_{m+2}(x) - \tfrac{3}{2}J_{m+3}(x) + \tfrac{1}{6}xJ_{m+4}(x)\right] \\ & + \sin^4\tfrac{1}{2}\theta\left[\tfrac{1}{72}x^2J_{m+6}(x) - \tfrac{17}{60}xJ_{m+5}(x) + \tfrac{11}{8}J_{m+4}(x) - \right. \\ & \left.\tfrac{4}{3x}J_{m+3}(x)\right] - \ldots \Big\} \end{aligned} \tag{3.131}$$

where

$$P_n^{-m}(\cos\theta) = \frac{\Gamma(n-m+1)}{\Gamma(n+m+1)} P_n^m(\cos\theta) \tag{3.132}$$

and m is a postive integer, n is an arbitrary real number, and $x = 2n\sin\theta/2$.

Motivated by these two examples, we shall discuss an orthonormal system for a spherical cap, as shown in Figure 3.5 and another system for a domain bounded by two parallels and two meridians on a sphere. Again, we start with the Helmholtz equation. The

suitable coordinates for the Helmholtz equation in a unit circle are the polar coordiantes r, λ, as shown in Figure 3.6. By separation of variables in r, λ and by $u = f(r)h(\lambda)$ in (3.31), we get two equations in analogy to (3.102) and (3.103) (Kaplan, 1981, p. 448):

$$f'' + \frac{1}{r}f' + \left(k - \frac{m^2}{r^2}\right)f = 0 \tag{3.133}$$

$$h'' + m^2 h = 0 \tag{3.134}$$

where k is the eigenvalue as in (3.31) and m is an arbitrary number. For the entire circle, we may require the periodicity in λ, i.e., $h(\lambda + 2\pi) = h(\lambda)$, thus m must be an integer. The solutions for (3.134) then become the classical Fourier basis functions $\cos m\lambda$ and $\sin m\lambda$. The solutions for (3.133) are the Bessel functions shown in Appendix C.

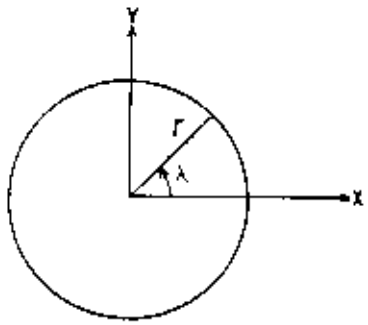


Figure 3.6 Polar Coordinate for a Unit Circle

Now, for the moment, we will treat (3.133) as an ordinary differential equation, regardless of whatever the definition of r is. In this manner, we could define a function of $t = \cos\theta$ which, after properly changing the variables, satisfies (3.133) and a necessary B.C.. Variable θ is the colatitude shown in Figure 3.4.

It is known from Appendix C that the set of functions
$J_m(r_{mp}r)$, $p = 1, 2, \ldots, \infty$; m fixed (3.135)

are orthogonal in [0, 1]. The r_{mp} values are the zeros of $J_m(r)$. If the eigenvalue k in (3.133) is chosen to be

$$k = r_{mp}^2 \tag{3.136}$$

then $J_m(r_{mp}r)$ satisfies (3.133) and the eigenvalues are simply the squares of the zeros. Given such an existing orthogonal system, we may construct an orthogonal system in the

domain $[t_o, 1]$ by the substitution $r = \dfrac{1 - t}{1 - \cos\theta_o}$ in (3.135). Thus we have, according to Appendix C,

$$\int_{r=0}^{1} rJ_m(r_{mp}r)J_m(r_{mq}r)dr$$

$$= \int_{\theta=0}^{\theta_o} \frac{1 - \cos\theta}{(1 - \cos\theta_o)^2} J_m\left(r_{mp}\frac{1 - \cos\theta}{1 - \cos\theta_o}\right)J_m\left(r_{mq}\frac{1 - \cos\theta}{1 - \cos\theta_o}\right)\sin\theta d\theta$$

$$= \int_{t_o}^{1} \frac{1 - t}{(1 - t_o)^2} J_m\left(r_{mp}\frac{1 - t}{1 - t_o}\right)J_m\left(r_{mq}\frac{1 - t}{1 - t_o}\right)dt$$

$$= \begin{cases} 0, & p \neq q \\ \dfrac{1}{2}J_{m+1}^2(r_{mp}), & p = q \end{cases} \tag{3.137}$$

Eq. (3.137) shows that the functions

$$\sqrt{1 - t}\, J_{mp}\left(r_{mp}\frac{1 - t}{1 - t_o}\right) \qquad p = 0, 1, \cdots, \infty \tag{3.138}$$

are orthogonal with respect to integration over $t_o \le t \le 1$. Now we can define the normalized Bessel function in $0 \le \theta \le \theta_o$ as

$$\bar{J}_{mp}(t) = \bar{J}_{mp}(\cos\theta) = \frac{\sqrt{2(2 - \delta(m))(1 - t)}}{\sqrt{1 - t_o}\,|J_{m+1}(r_{mp})|} J_m\left(r_{mp}\frac{1 - t}{1 - t_o}\right) \tag{3.139}$$

and the solutions for $h(\lambda)$, i.e., $\cos m\lambda$ and $\sin m\lambda$, form the orthonormal Fourier-Bessel functions defined as

$$\begin{Bmatrix} \bar{B}_{mp}(\theta, \lambda) \\ \bar{C}_{mp}(\theta, \lambda) \end{Bmatrix} = \bar{J}_{mp}(\cos\theta)\begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} \tag{3.140}$$

where $0 \le \theta \le \theta_o$, $0 \le \lambda \le 2\pi$. For a proof of the completeness and the convergence (in the mean) of the Fourier-Bessel series, Tolstov (1976, Chapter 8) can be consulted. A square integrable function $f(\theta, \lambda)$ in $\{0 \le \theta \le \theta_o, 0 \le \lambda \le 2\pi\}$ may be expanded into the Fourier-Bessel series as follows:

$$f(\theta, \lambda) = \sum_{m=0}^{\infty} \sum_{p=0}^{\infty} (\bar{a}_{mp}\bar{B}_{mp} + \bar{b}_{mp}\bar{C}_{mp}) \tag{3.141}$$

Note that the first positive root of $J_m(r)$ corresponds to $p = 0$. The coefficients $\bar{a}_{mp}$, $\bar{b}_{mp}$ are found by

$$\begin{Bmatrix} \bar{a}_{mp} \\ \bar{b}_{mp} \end{Bmatrix} = \frac{1}{A} \int_{\theta=o}^{\theta_o} \int_{\lambda=0}^{2\pi} f(\theta, \lambda) \begin{Bmatrix} \bar{B}_{mp} \\ \bar{C}_{mp} \end{Bmatrix} \sin\theta d\theta d\lambda \tag{3.142}$$

where $A = 2\pi(1 - \cos\theta_o)$ is the area of the spherical cap. For a numerical evaluation of $J_m(r)$, the IMSL routine BSJNS may be used. Figure 3.7 shows the normalized Bessel functions ((3.139)) for $m = 0$ and $p = 0, 1, 2, 3$. The roots, i.e., r_{mp} in Figure 3.7, are taken from Beyer (1987, p. 352) and their values are plotted next to the curves.

It is clear that, if a system $\{f_i(x)\}$ is orthogonal w.r.t integration over interval $a \le x \le b$ and the other system $\{g_j(y)\}$ is orthogonal w.r.t integration over the interval $c \le y \le d$, then the products $\phi_{ij}(x, y) = f_i(x)f_j(y)$ are orthogonal w.r.t integration over the domain $a \le x \le b$ and $c \le y \le d$. Using such a concept, we now shall propose an orthogonal series for a region bounded by two parallels and two meridians, as shown in Figure 3.8. To this end, we make the following substitution in the Bessel function (see (3.135)) for the latitude (or colatitude) dependent part:

$$r = \frac{t_1 - t}{t_1 - t_2} = \frac{\cos\theta_1 - \cos\theta}{\cos\theta_1 - \cos\theta_2} \tag{3.143}$$

Then the normalized Bessel Function in $\theta_1 \le \theta \le \theta_2$ can be defined as

$$\bar{K}_{mp}(t) = \bar{K}_{mp}(\cos\theta) = \frac{\sqrt{2(2 - \delta(m))(t_1 - t)}}{\sqrt{t_1 - t_2}\,|J_{m+1}(r_{mp})|} J_m\left(r_{mp}\frac{t_1 - t}{t_1 - t_2}\right) \tag{3.144}$$

For the longitude dependent part, we may take $\cos m\frac{2\pi}{\Delta\lambda}\lambda$, $\sin m\frac{2\pi}{\Delta\lambda}\lambda$, with $\Delta\lambda = \lambda_2 - \lambda_1$. Thus the following system is orthonormal in $\{\theta_1 \le \theta \le \theta_2$ and $\lambda_1 \le \lambda \le \lambda_2\}$:

$$\begin{Bmatrix} \bar{E}_{mp} \\ \bar{F}_{mp} \end{Bmatrix} = \bar{K}_{mp}(\cos\theta) \begin{Bmatrix} \cos m\frac{2\pi}{\Delta\lambda}\lambda \\ \sin m\frac{2\pi}{\Delta\lambda}\lambda \end{Bmatrix} \tag{3.145}$$

It should be noted that the functions $\bar{K}_{mp}(\cos\theta)$ and $\cos m(2\pi/\Delta\lambda)\lambda$, $\sin m(2\pi/\Delta\lambda)\lambda$ are not the solutions for the Helmholtz equation as given by the equivalent system (3.133) and (3.134). To obtain solutions for such a domain, it would require the use of the Bessel

functions of non-integer order. For a solution of the Helmholtz equation in such a domain, Sommerfeld (1949, p. 167-168) may be consulted. Our main goal here is just to find an orthonormal set in such a particular domain and the set in (3.145) has fulfilled the need.

For a square integrable function in such a domain, the expansion formulae similar to (3.141) and (3.142) can be used. Since Bessel function is a "straight" function, it is expected that the separation between the two meridians in Figure 3.7 is not too large.

By far we have presented various sets of eigenfunctions for various domains of concern. Each set of functions form a function space with countably infinite number of dimensions. Each of these spaces fulfills all the requirements for a Hilbert space and, as a matter of fact, it is a Hilbert space. For an extensive discussion on the function space spanned by the eigenfunctions, see Page (1955, Chapter 5).

3.3.3 Frequency Classification of Eigenfunctions

One important issue in developing the orthogonal functions for this study is the frequency classification, since eventually we will be dealing with discrete data for which a sampling theorem concerning the data interval needs to be applied. The frequency of an eigenfunction is usually measured by the density of node. A node is a collection of points where the corresponding eigenfunction vanishes. Thus, in a one-dimensional case a node is one single point; in a two-dimensional case a node is a curve, or more specifically, a zero-contour; in a three-dimensional case a node is a surface. Unlike the nodes of a sine or cosine function which is also an eigenfunction for some particular problem, the nodes of an eigenfunction we have discussed in this chapter are usually not evenly distributed on the interval where the function is defined. For example, for function sinx at $0 \le x \le 2\pi$, the separation between two nodes is precisely π. This is not the case for functions $P_\ell^m(t)$ and $J_m(r_{mp}r)$ that we have studied. See also Figures 3.3 and 3.7.

According to the definition of node, one node corresponds to a change of sign of an eigenfunction. Thus the denser the distribution of nodes is, the faster the eigenfunction oscillates, or the higher the frequency is. It is also clear that a node is a zero of a function. For a self-adjoint Sturm-Liouville problem, if we arrange the eigenvalues in the order of increasing magnitude, i.e.,

$$k_1 < k_2 < \ldots \tag{3.146}$$

Figure 3.8 A Region Bounded by Two Parallels and Two Meridians with the corresponding eigenfunctions

$$v_1, v_2, \ldots \tag{3.147}$$

we shall show that if $k_2 > k_1$, then the number of zeros of v_2 is greater than that of v_1, or the frequency of v_2 is higher than that of v_1. For simplicity we discuss a one-dimensional case. With proper changes of notation and use of an arbitrary t as the upper limit for the integration in (3.109), we have (note: this holds for any kind of Sturm-Liouville problem, not only for P_ℓ^m)

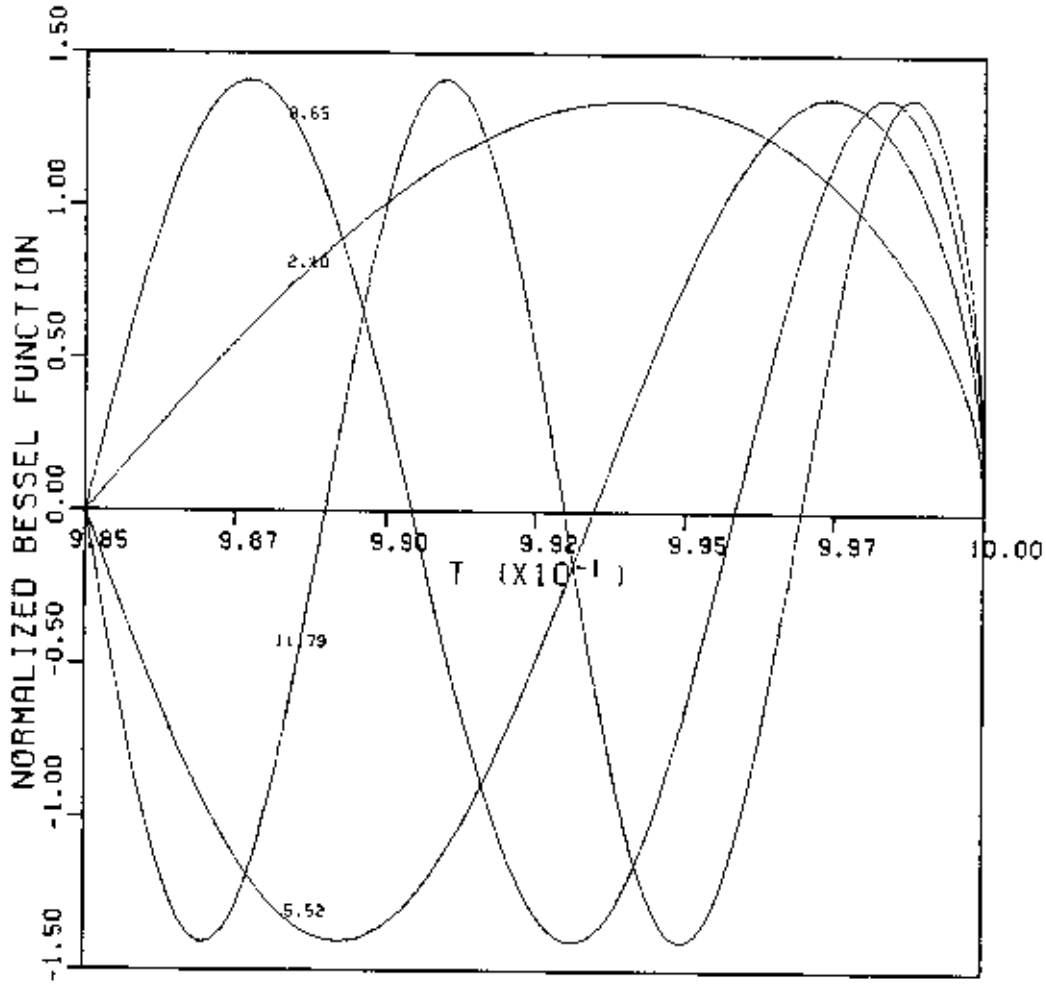


Figure 3.7 Normalized Bessel Function in $\cos 10° \le t \le 1$, m = 0, Roots r_{mp} Plotted Next to the Curves

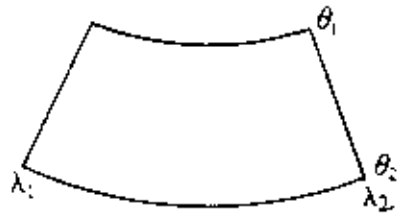


Figure 3.8 A Region Bounded by Two Parallels and Two Meridians

$$(k_2 - k_1)\int_{t_o}^{t} v_1 v_2 dt = -p(t)(v_1 v_2' - v_1' v_2)\Big|_t + p(t)(v_1 v_2' - v_1' v_2)\Big|_{t=t_o}$$

$$= -p(t)(v_1 v_2' - v_1' v_2)\Big|_t \quad , \quad p(t) > 0 \tag{3.148}$$

where $t_o \le t \le t_1$, t_o and t_1 being two boundary points, see also Figure 3.9. Assuming that t is the first zero of v_1, then (3.148) becomes (since $v_1(t) = 0$):

$$(k_2 - k_1)\int_{t_o}^{t} v_1 v_2 dt = p(t)v_1'(t)v_2(t) \tag{3.149}$$

If $v_1 > 0$ in (t_o, t), then $v_1'(t)$ must be negative (since at t, v_1 changes from a positive number to a negative number). Since $k_2 > k_1$, v_2 must change sign in (t_o, t) to make (3.149) true; if $v_1 < 0$, v_2 must also change sign for the same reason. Thus a zero is encountered by v_2 in (t_o, t) (note: an open interval, not a closed interval).

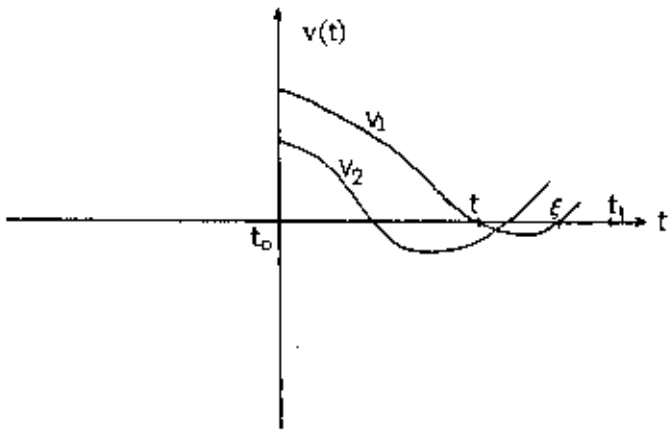


Figure 3.9 Zeros of Two Eigenfunctions, $k_2 > k_1$

Assuming that ξ is the next zero of v_1, we now take integration from t to ξ in (3.109) to get (since $v_1(t) = v_1(\xi) = 0$)

$$(k_2 - k_1)\int_{t}^{\xi} v_1 v_2 dt = p(t)\left(v_1'(\xi)v_2(\xi) - v_1'(t)v_2(t)\right) \tag{3.150}$$

Using exactly the same argument, we can show that v_2 must change sign in (t, ξ) to make (3.150) true, thus a zero is encountered by v_2 in (t, ξ). In this way, we thus prove that the number of zeros of v_2 is greater than that of v_1 if $k_2 > k_1$. Further, a theorem by Courant and Hilbert (1953, p. 454) states that:

> The n-th eigenfunction for a Sturm-Liouville problem divides the fundamental domain into precisely n parts by means of its nodes.

In the above theorem, the eigenvalues must be in the order shown in (3.146). One could inspect Figures 3.3 and 3.7 to verify this theorem. (Note: if we write the Bessel equation as $(rJ_m')' - m^2J_m/r + krJ_m = 0$, then clearly it is a Sturm-Liouville form).

Courant and Hilbert's theorem (1953) can be applied to the Legendre polynomials since we have already shown that the Legendre equation can be written in a Sturm-Liouville form. Specifically, the eigenvalues of Legendre polynomial are $k = n(n + 1)$, with n being degree. Explicitly, k are

$$0, 2, 6, 12, \ldots \tag{3.151}$$

Thus $P_0(t)$ (the first eigenfunction) has the same sign over [-1, 1], $P_1(t)$ (the second eigenfunction) changes sign once over [-1, 1], $\cdots$. One could verify this statement by looking at Figure 1-8 in Heiskanen and Moritz (1967, p. 24).

We shall interpret the eigenvalue in a different way. We start the discussion with the two-dimensional Helmholtz equation on a plane over a domain σ. It can be shown that the solutions of the Helmholtz equation may be generated through a variational problem with a constraint (Dettman, 1988) as follows:

$$I(u) = \iint_\sigma (\nabla u \cdot \nabla u)dxdy = \iint_\sigma \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right]dxdy = \text{a minimum} \tag{3.152}$$

subject to

$$\iint_\sigma u^2 dxdy = 1 \tag{3.153}$$

To see this, we employ the calculus of variation and set

$$L(x, y;\ u, u_x, u_y) = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 - ku^2(x, y) \tag{3.154}$$

where k is the Lagrange multiplier, and $u_x = \partial u/\partial x$, $u_y = \partial u/\partial y$. Then Euler's equation for this case is (Gelfand and Fomin, 1962):

$$\frac{\partial L}{\partial u} - \left[\frac{\partial}{\partial x}\left(\frac{\partial L}{\partial u_x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial L}{\partial u_y}\right)\right] = 0 \Leftrightarrow \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + ku = 0 \tag{3.155}$$

Clearly (3.155) states that the eigenvalue k is the Lagrange multiplier. The constraint (3.153) implies a normalized eigenfunction. This short account has shown how the variational problem can be transformed into the Helmholtz equation. This result can be immediately extended to a sphere where a coordinate free expression for $\nabla u \cdot \nabla u$ needs to be used and Euler's equation still maintains the same form. It is remarkable that the eigenvalue k is exactly the required minimum value in (3.152) (see also Morse and Feshbach, 1953, for a one-dimensional case). This can be shown using integration by parts in two-dimensions. We have

$$I(u) = \iint_\sigma \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right] dxdy$$

$$= \oint_B u\frac{\partial u}{\partial x}\, n \cdot i ds + \oint_B u\frac{\partial u}{\partial y}\, n \cdot j ds - \iint_\sigma u\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) dxdy$$

$$= \oint_B u\left(\frac{\partial u}{\partial x} i + \frac{\partial u}{\partial y} j\right) \cdot n ds - \iint_\sigma u(-ku) dxdy$$

$$= \oint_B u\frac{\partial u}{\partial n} ds + k\iint_\sigma u^2 dxdy = k \qquad (3.156)$$

where we have assumed $u = 0$ or $\partial u/\partial n = 0$ on the boundary B, i and j being unit vectors along x and y directions, respectively, and n the outer normal. The relationship in (3.155) is also used in deriving (3.156).

What this variational principle tells us is that a solution u is a function which makes the "sum" of squares of the magnitude of gradient minimum, or u is the "smoothest" one among all the possible functions. The eigenvalue k is just the sum. For the next eigenfunction, as its eigenvalue increases, the sum I(u) then increases, hence this particular eigenfunction becomes "rougher". Such an argument can be carried over to the third eigenfunction with a larger (than second one) eigenvalue. In this way, we thus show that, a sequence of eigenfunctions has increasing frequency in the order of increasing eigenvalue. The interpretation of frequency in this way is particularly useful for a set of orthogonal functions constructed over a domain whose boundary cannot be described by a simple geometry, since the above result is valid for any kind of continuous boundary.

3.4 Method of Conformal Mapping for Constructing Complex Orthogonal Functions

3.4.1 Transformation Between Two Domains and Finding the Orthonormal Functions

In this section, we will try the third method of constructing orthogonal functions. The algebra of complex variables will be used here. The notation i is reserved for the imaginary number $\sqrt{-1}$. Let us start with Riemann's mapping theorem (Dettman, 1965, p. 256):

> Let R be a simply connected domain on the z-plane with at least two boundary points. Then there exists a simple function $w = f(z)$ which maps R onto the unit disk R, $|w| < 1$. If we specify that a point z_0 in R is mapped into the origin O in R′ and a given direction at z_0 is mapped into a given direction at O, then the mapping is one-to-one and unique. See also Figure 3.10.

The proof of this theorem can be found in Dettman (1965). A simply connected domain R is a domain such that every simple closed curve (such a curve does not intersect

itself) within it encloses on points of R (Churchill and Brown, 1984). In other words, if C is a curve in R, then it can be shrunk to a point in R, cf. Figure 3.11. Riemann's mapping theorem was presented in Riemann's famous dissertation "Foundations for a general theory of functions of a single complex variable" and was proved only for a simply connected domain. However, it is possible to extend this theorem to a case where a domain is bounded by two closed curves, one inside the other (Spiegel, 1964). While Riemann's mapping theorem guarantees the existence of such a mapping, it does not provide a concrete way of constructing a specific mapping.

With Riemann's mapping theorem as backbone, we now present a theorem related to the construction of complex orthogonal functions. Let

$$w = f(z) = u(x, y) + iv(x, y) \tag{3.157}$$

be an analytic function mapping a domain B_z on the z-plane onto a domain B_w on the w-plane, as shown in Figure 3.12. The Jacobian of the transformation between B_z and B_w is

$$\begin{aligned} J_w &= \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x} \\ &= \frac{\partial u}{\partial x}\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial u}{\partial y} = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 \\ &= \left|\frac{\partial u}{\partial x} - i\,\frac{\partial u}{\partial y}\right|^2 = |f'(z)|^2 \end{aligned} \tag{3.158}$$

If $J_w \neq 0$ for all the points in B_z, then the inverse funtion of f

$$z = f^{-1}(w) = g(w) \tag{3.159}$$

exists and the mapping w = f(z) is one-to-one (see Churchill and Brown, 1984, pp. 220-223).

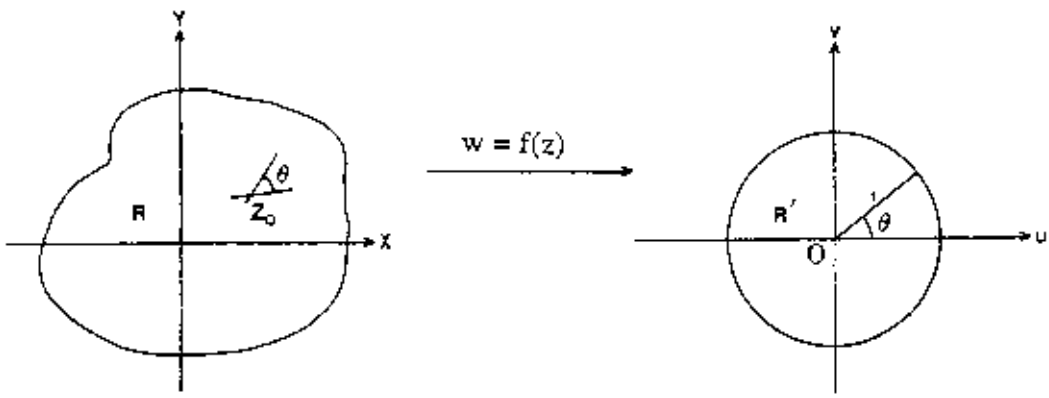


Figure 3.10 Riemann's Mapping Theorem

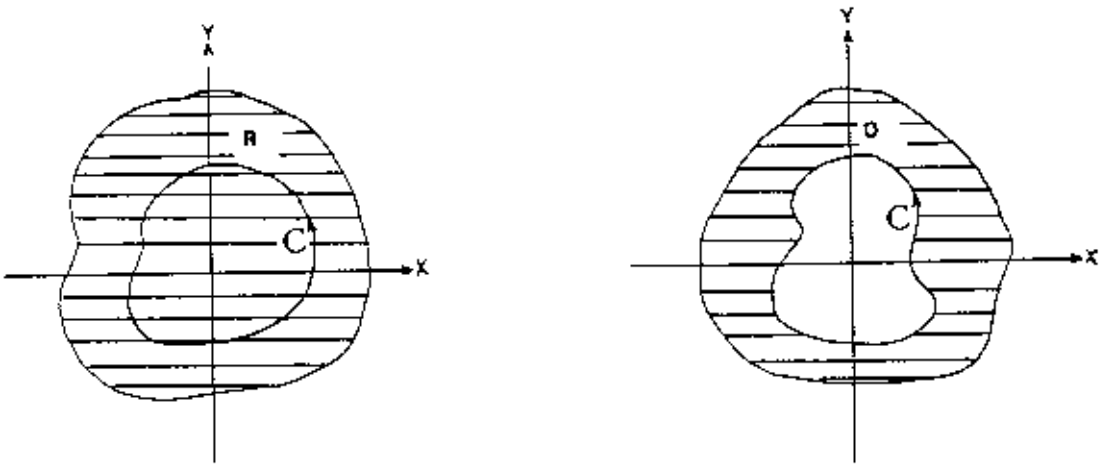


Figure 3.11 A Simply Connected Domain R, a Multiply Connected Domain D and a Curve C

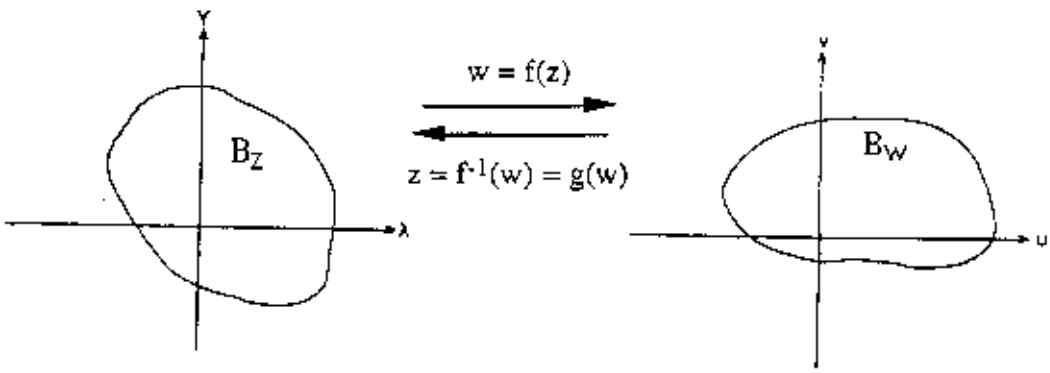


Figure 3.12 Transformation Between Domains B_z and B_w

In (3.158) we have used the Cauchy-Riemann equations for the analytical function f(z) (Churchill and Brown, p. 40):

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad , \quad \frac{\partial u}{\partial y} = - \frac{\partial v}{\partial x} \tag{3.160}$$

and it is understood that, since f(z) is analytic (or holomorphic), we can approach z from any directions on the z-plane when evaluating the derivative of f, namely

$$\begin{aligned} f'(z) &= \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} = \lim_{\substack{\Delta x \to 0 \\ \Delta y = 0}} \frac{f(z + \Delta z) - f(z)}{\Delta x} \\ &= \lim_{\substack{\Delta y \to 0 \\ \Delta x = 0}} \frac{f(z + \Delta z) - f(z)}{i\Delta y} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y} \end{aligned} \tag{3.161}$$

Thus by (3.161) the last identity in (3.158) holds. It can be shown that the inverse Jacobian J_z is $1/J_w$, or (Spiegel, 1964)

$$J_z = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{1}{J_w} = \left|\frac{dg}{dw}\right|^2 = |g'(w)|^2 \tag{3.162}$$

Using the Jacobian, the relationship between a surface element $d\sigma_z$ and a surface element $d\sigma_w$ is

$$d\sigma_z = |g'(w)|^2 d\sigma_w \tag{3.163}$$

With the above development, we now show that (cf. Smirnov and Lebedev, 1968, p. 225):

If $\{\varphi_j(z)\}$ is a complex orthonormal systems in domain B_z, then the functions

$$\psi_j(w) = \sqrt{\frac{A_w}{A_z}}\, \varphi_j(z)g'(w) = \sqrt{\frac{A_w}{A_z}}\, \varphi_j(g(w))g'(w) \tag{3.164}$$

form an orthonormal system in B_w.

A_z and A_w are the areas of B_z and B_w, respectively, defined as

$$A_z = \iint_{B_z} d\sigma_z = \iint_{B_z} dxdy \tag{3.165}$$

$$A_w = \iint_{B_w} d\sigma_w = \iint_{B_w} dudv \tag{3.166}$$

The theorem in (3.164) can be proved as follows: Since φ_j are orthonormal, we have

$$\begin{aligned}
\delta_{nm} &= \frac{1}{A_z}\iint_{B_z} \varphi_n(z)\varphi_m^*(z)d\sigma_z \\
&= \frac{1}{A_z}\iint_{B_w} \varphi_n(g(w))\varphi_m^*(g(w))\left|g'(w)\right|^2 d\sigma_w \\
&= \frac{1}{A_w}\iint_{B_w} \left[\sqrt{\frac{A_w}{A_z}}\varphi_n(g(w))g'(w)\right]\left[\sqrt{\frac{A_w}{A_z}}\varphi_m(g(w))g'(w)\right]^* d\sigma_w \\
&= \frac{1}{A_w}\iint_{B_w} \psi_n(w)\psi_m^*(w)d\sigma_w
\end{aligned} \tag{3.167}$$

where

$$\delta_{nm} = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases} \tag{3.168}$$

There are two important related theorems:

- If the system $\{\varphi_j(z)\}$ is complete in B_z, then the system $\{\psi_j(w)\}$ is complete in B_w. The relation of completeness can be found in Section 2.4.
- For every simply connected domain, there is a complete orthonormal system.

For the proof of the first theorem, see Smirnov and Lebedev (1968, p. 226). The second theorem is true if we can find a complete orthonormal system in the unit disk. Since, by Riemann's mapping theorem, we are able to find a mapping function between a simply connected domain and the unit disk, the orthonormal system in the simply connected domain is then constructed by means of (3.164). In Table 1.1, we have pointed out that the polynomials z^n, $n = 1, 2, ...$, are orthogonal with respect to the area integral over the unit disk. To see this, we use the polar coordinate r, λ (see Figure 3.6) and let $z = x + iy = re^{i\lambda}$, then

$$(z^n, z^m) = \int_{r=0}^{1}\int_{\lambda=0}^{2\pi} e^{i\lambda(n-m)}r^{n+m}rdrd\lambda$$

$$= \begin{cases} \dfrac{\pi}{n+1} & , \quad n = m \\ 0 & , \quad n \neq m \end{cases} \tag{3.169}$$

Therefore, the series (since the area of the unit disk is π)

$$\sqrt{n+1}\, z^{n+1} \quad , \quad n = 0, 1, \cdots \tag{3.170}$$

forms an orthonormal system in the unit disk $|z| \leq 1$.

To construct a complex orthonormal system over the oceans, the only thing we have not done in this approach is the finding of the transformation function between the oceans and the unit disk, namely the function g(w) in (3.164). As stated in Riemann's mapping theorem, a domain must be simply connected in order to have a transformation between itself and the unit disk. The extension of Riemann's theorem also has a strict limitation. Since the real oceans are by no means a simply connected domain or a domain bounded by two closed curves, it will be impossible to find a transformation between the real oceans and the unit disk. So again we have to idealize the oceans. But this time we have more freedom of "designing" the oceans, at least the oceans need not be restricted to a spherical cap or a region bounded by two parallels and two meridians. We will discuss this shortly.

Having found a complete and countable orthonormal system $\{\psi_j(w)\}$ in domain B_w, we can expand an arbitrary function $\zeta \in L^2(B_w)$ into $\psi_j(w)$, as in the case of real orthonormal function expansion. For the practical case in this study, ζ will be a real-valued function $\zeta(u, v)$, which can still be expanded into a complex orthonormal series. The expansion has the form

$$\zeta(u, v) = \sum_{j=1}^{\infty} c_j \psi_j(w) \tag{3.171}$$

and the coefficients c_j are found by

$$c_j = \frac{1}{A_w} \iint_{B_w} \zeta(u, v) \psi_j^*(w) d\sigma_w \tag{3.172}$$

3.4.2 Schwarz-Christoffel Transformation Between the Oceans and the Unit Disk

We now turn to the realization of the mapping between the oceans and the unit disk. For this purpose, we first introduce the Schwarz-Christoffel transformation which maps the interior of a polygon onto the half plane $\mathrm{Im}(z) \geq 0$, where Im(z) indicates the imaginary part of z. Then, by finding another transformation which maps the half plane $\mathrm{Im}(z) \geq 0$ onto the unit disk, we will complete our task. The S-C transformation is defined by (see Churchill and Brown, 1984, Dettman, 1965):

$$w = f(z) = A\int_0^z (s - x_1)^{-k_1}(s - x_2)^{-k_2}\cdots(s - x_n)^{-k_n}ds + B \tag{3.173}$$

where

k_j : $\frac{1}{\pi}(\pi - \alpha_j)$, α_j is the interior angle at vertex w_j (given)
x_j : point j on the x-axis, corresponding to the image of w_j of the polygon (unknown)
A : A rotation and scale factor of the polygon (unknown)
B : Translation factor (unknown)

Note that the given polygon is on the w-plane and its image is on the z-plane. Written down in the form (3.173), the S-C transformation is from the real axis (z-plane) to the polygon (w-plane), not in a direct form we desire (from the polygon to the real axis). However, if we can find the inverse of the S-C transformation (the existence will be shown later), our goal can still be achieved. The validity of S-C transformation can be shown below.

Differentiating (3.173) with respect to z, we have

$$f'(z) = \frac{dw}{dz} = A(z - x_1)^{-k_1}(z - x_2)^{-k_2}\cdots(z - x_n)^{-k_n} \tag{3.174}$$

Thus

$$\arg\left(f'(z)\right) = \arg(A) - k_1\arg(z - x_1) - \cdots - k_n\arg(z - x_n) \tag{3.175}$$

where $\arg(z_0)$ is the angle between the x-axis and the vector from the origin to a point z_0, being positive if counter clockwise. From (3.174) to (3.175) we have used the basic properties of complex functions which can be easily found in a textbook of complex variable. From (3.175) and Figure 3.13, we can see the following results if z is on the real axis:

$$\left.\begin{aligned}
&(1)\ z < x_1 \Rightarrow \arg\left(f'(z)\right) = \arg(A) - (k_1 + k_2 \cdots + k_n)\pi \\
&(2)\ x_1 < z < x_2 \Rightarrow \arg\left(f'(z)\right) = \arg(A) - (k_2 + \cdots + k_n)\pi \\
&\qquad\vdots \\
&(n)\ x_{n-1} < z < x_n \Rightarrow \arg\left(f'(z)\right) = \arg(A) - k_n\pi
\end{aligned}\right\} \tag{3.176}$$

Thus, when a point z moves from the interval (x_{j-1}, x_j) to the next interval (x_j, x_{j+1}), the argument of $f'(z)$ changes by an angle of $k_j\pi$, which is the exterior angle of the polygon at vertex w_j (see Figure 3.13). While z remains on any of the intervals (x_j, x_{j+1}), $\arg(f'(z))$ is some constant. If a point w on the boundary of the polygon travels in the counterclockwise direction, the interior of the polygon is on the left hand side of the boundary.

Eq. (3.175) and (3.176) describe how the angle of a point on a curve on the z-plane will be changed on the w-plane by the conformal mapping w = f(z) (see also Churchill and Brown, 1984, p. 217). The way the angles change as described in (3.176) can only be possible for a line on the z-plane and a polygon on the w-plane. Since we request that the line be on the x-axis, the transformation in (3.173) holds. There are many issues to be clarified. We will address them when we discuss the next mapping.

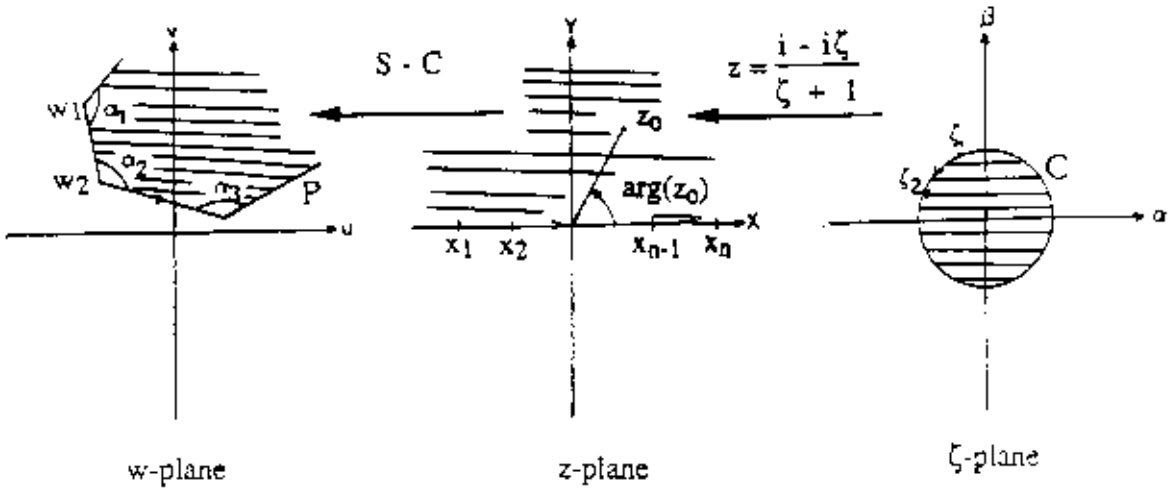


Figure 3.13 Mapping the Unit Disk Onto a Polygon

The author does not intend to repeat the already good documentation about the S-C transformation which can be found in textbooks such as Churchill and Brown (1984), Dettmann (1965), Spiegel (1964), and Bieberbach (1953). Instead, the author would like to apply such a transformation to the construction of orthonormal functions and point out some possible problems when using them.

The next mapping to be used is the one that transforms the unit disk $|\zeta| \le 1$ to the half plane $\mathrm{Im}(z) \ge 0$, where $\zeta = \alpha + i\beta$, see also Figure 3.13. The inverse of this mapping can be found in Churchill and Brown (1984), namely, $\zeta = \frac{1 - z}{i + z}$. Thus the desired mapping is

$$z = g(\zeta) = \frac{i - i\zeta}{\zeta + 1} \tag{3.177}$$

Then

$$z' = g'(\zeta) = \frac{-2i}{(\zeta + 1)^2} \tag{3.178}$$

From (3.173), the mapping from the unit disk to the interior of a polygon is

$$w = f(z) = f(g(\zeta)) \tag{3.179}$$

Thus

$$\begin{aligned}
\frac{dw}{d\zeta} &= \frac{df}{dz}\frac{dz}{d\zeta} \\
&= A(z - x1)^{-k1}(z - x2)^{-k2} \cdots (z - xn)^{-kn} \cdot \left[\frac{-2i}{(\zeta + 1)^2}\right] \\
&= A\prod_{j=1}^{n}\left(\frac{i - i\zeta}{\zeta + 1} \cdot \frac{i - i\zeta_j}{\zeta_j + 1}\right)^{-k_j} \cdot \left[\frac{-2i}{(\zeta + 1)^2}\right] \\
&= -i2A(-2i)^{-\Sigma k_j} \cdot (1 + \zeta)^{\Sigma k_j} \cdot \left[\prod_{j=1}^{n}(1 + \zeta_j)^{k_j}\right] \cdot \left[\prod_{j=1}^{n}(\zeta - \zeta_j)^{-k_j}\right] \cdot \frac{1}{(1 + \zeta)^2}
\end{aligned} \tag{3.180}$$

Since the sum of the exterior angles of a polygon is 2π, and $k_j\pi + \alpha_j = \pi$, we have

$$2\pi = \sum_{j=1}^{n}(\pi - \alpha_j) = \pi\sum_{j=1}^{n} k_j \Leftrightarrow \sum_{j=1}^{n} k_j = 2 \tag{3.181}$$

Substituting (3.181) into (3.180) and grouping all the constants into one single constant, we get

$$\begin{aligned}
\frac{dw}{d\zeta} &= \overline{A}\prod_{j=1}^{n}(\zeta - \zeta_j)^{-k_j} \\
&= \overline{A}(\zeta - \zeta_1)^{-k_1}(\zeta - \zeta_2)^{-k_2} \cdots (\zeta - \zeta_n)^{-k_n} \\
&= 1/(d\zeta/dw)
\end{aligned} \tag{3.182}$$

with

$$\overline{A} = \frac{iA}{2}\prod_{j=1}^{n}(1 + \zeta_j)^{k_j}$$

Thus the desired transformation from the unit disk to the interior of a given polygon is:

$$w = h(\zeta) = \overline{A} \int_o^{\zeta} \prod_{j=1}^{n} (\eta - \zeta_j)^{-k_j} \, d\eta + \overline{B} \tag{3.183}$$

Eq. (3.183) has exactly the same form as (3.173). Important issues related to the transformation in (3.183) and to the problem of finding orthonormal functions in the interior of a polygon are:

(1) The transformation is one-to-one between points inside and on the polygon and points inside and on the unit circle, except at the singular points ζ_j. To see the one-to-one correspondence, we recall the Cauchy integral formula

$$\phi(\xi_o) = \frac{1}{2\pi i} \oint_B \frac{\phi(\xi)}{\xi - \xi_o} \, d\xi \tag{3.184}$$

where $\phi(\xi)$ is an analytical function defined in the domain enclosed by the boundary curve B and ξ_o is an arbitrary point inside B. Letting $\phi(\xi) = 1$ everywhere, we have

$$1 = \frac{1}{2\pi i} \oint_B \frac{d\xi}{\xi - \xi_o} \tag{3.185}$$

Furthermore, according to Churchill and Brown (1984, p. 169), if N is the number of zeros of function $r(\zeta)$ inside B, then

$$\frac{1}{2\pi i} \oint_B \frac{r'(\zeta)}{r(\zeta)} \, d\zeta = N \tag{3.186}$$

Now we let P be the given polygon, C be the boundary of the unit disk, w_o be a point interior to P and the image of a point b in the unit disk. Furthermore, in (3.185) we let $\xi = h(\zeta)$, $\xi_o = w_o = h(b)$, we have

$$1 = \frac{1}{2\pi i} \oint_P \frac{d\xi}{\xi - \xi_o} = \frac{1}{2\pi i} \oint_C \frac{h'(\zeta) d\zeta}{h(\zeta) - w_o}$$

$$= \frac{1}{2\pi i} \oint_C \frac{[h(\zeta) - w_o]'}{h(\zeta) - w_o} \, d\zeta \tag{3.187}$$

Thus, according to (3.186), the function $(h(\zeta) - w_o)$ has only one zero inside C, namely, $h(b) = w_o$. From this derivation we verify the one-to-one correspondence in (3.183). A similar proof can be found in (Spiegel, 1964, p. 223).

(2) The function $h(\zeta)$ in (3.183) can hardly be an elementary function. The inverse function of $h(\zeta)$, i.e.,

$$\zeta = h^{-1}(w) = g(w) \tag{3.188}$$

that is needed in (3.164) is even more difficult to find. However, the function $g'(w) = d\zeta/dw$ does have a simple form, as given in (3.182). The polygons that will yield elementary functions $h(\zeta)$ can be found in Churchill and Brown (1984), but normally they are triangles or "degenerate" polygons (open polygons). If there is no island in the open oceans and some regularization process is made, then "the oceans" may be a polygon with an extremely large number of vertices. To find the explicit form of $h(\zeta)$ for such a polygon, we need to first do some series expansion inside the integral in (3.183), then perform the inverse series expansion to get $g(w)$ after the integration is done. Therefore, the function $g(w)$ will only be approximately obtained.

(3) The polygon is given, thus we know the exterior angles, or the k_j values. However, the constants $|\bar{A}|$, $\arg(\bar{A})$, $\bar{B}$ and all the points ζ_j on C need to be determined. The solution of these unknowns can be found by setting up n equations using (3.183) and employing the conditions $w_j = h(\zeta_j)$. Recalling that $\bar{A}$ has the responsibility of scaling and rotating the polygon and $\bar{B}$ is used to transfer the polygon, these constants (three in total, since we split $\bar{A}$ into $|\bar{A}|$ and $\arg(\bar{A})$) must be determined after ζ_j are decided. Since we only have n equations, three of ζ_j can be chosen arbitrarily. An alternative explanation on the determination of constants can be found in Dettman (1965).

(4) Finally, we have to settle the problem of working on the surface of a sphere. Since the previous section, we have used the variables u, v for the w-plane on which the polygon of the oceans is supposed to be defined. The vertices of the polygon of the oceans are normally given in the θ, λ coordinates. In order to have a surface integral over the oceans, it is necessary to use

$$\begin{aligned} u &= \lambda \\ v &= t = \cos\theta \end{aligned} \tag{3.190}$$

as the two real variables on the w-plane.

3.5 Some Comments on the Proposed Methods

In this chapter, three methods of constructing a set of orthonormal functions over a given domain have been presented. They are the Gram-Schmidt orthonormalizing process, the method of eigenvalue-eigenfunction and the method of conformal mapping. The first method is purely "mechanical", requiring only a set of independent functions pre-defined over the given domain. There is essentially no restriction on the shape of the domain, thus the construction of orthonormal functions is possible even for a domain such as the oceans which has an extremely irregular boundary. Two detailed algorithms have been worked

out for this method. What is left is just the calculation of the scalar inner products of the predefined functions over the oceans. Such a calculation can always be carried out using a numerical method.

The second method theoretically can lead to an orthonormal (ON) system over the oceans, as long as we can solve the spherical Helmholtz equation $\Delta^*Y + kY = 0$ or the Sturm-Liouville-type of eigenvalue problem analytically. Unfortunately, the ON systems by this approach are found only for two kinds of domains on the sphere: a spherical cap and a region bounded by two parallels and two meridians. It should be emphasized that only the analytical solution will lead to an ON system, since for any kind of domain, a numerical method for the solution of the Helmholtz equation is always possible, yielding some approximation for the Y function. The numerical method can be characterized as the finite element method in which certain "shape functions" are chosen to approximate Y according to the partitioning of the original domain. A useful textbook dealing with this subject is, for example, Zienkiewicz and Morgan (1983).

The third method involves the transformation between the given domain and some other "regularized" domain where an ON system has already existed. The crucial part is the transformation function which must be also analytical. We have introduced the Schwarz-Christoffel transformation which maps an ocean-like polygon onto the unit disk. However, due to its non-elementary form, the S-C transformation will only create an ON system over the oceans theoretically. Unlike the numerical solution for the Helmholtz equation, this time the approximation is needed for the transformation function. Since the transformation of domains is a necessary step, the use of an approximated transformation function will create only an "approximated" ON system, namely a system which cannot be exactly orthonormal between any pair of the members. A number of other applications of conformal mapping in physical problems, such as fluid flow and heat conduction, can be found in Churchill and Brown (1984).

During the discussion of the second method, some ON systems have already been presented analytically. It is also possible to get the analytical forms of some ON systems using the third method, provided that the domains are such that the S-C transformation has an elementary form. Therefore, the second and the third methods can provide complete ON systems for some "regularized" domains where spectral analyses based on these systems can be made. For example, one could expand the sea surface topography in the Antarctic area into the spherical cap harmonic series. On the other hand, the first method, i.e., the Gram-Schmidt process is very flexible in constructing the ON functions, therefore it will be chosen for the numerical experiments concerning the expansion of the Levitus sea surface topography and the simultaneous geoid-SST solution in satellite altimetry.

The requirement of completeness of the ON systems in the sense of (2.20) is also very important in all the discussions made in this Chapter. The completeness will ensure the series expansions of functions in terms of these ON systems to be convergent in the mean.

4. Functions Orthogonal on an Entire Sphere

In this chapter, functions that are orthogonal on an (entire) sphere are studied. The relationship between the surface spherical harmonics and a 2-D Fourier functions is also addressed. Two global alternative orthonormal sets of basis functions are suggested separately in the two sections that follow. If one does not quite care whether the functions used are orthogonal with respect to integration over the oceans, these two orthogonal systems (on an entire sphere) are highly recommended, due to their simplicity and their theoretical justification.

4.1 Spherical Harmonics and 2-D Fourier Series

This subject has been pursued by Colombo (1981), but it is felt that some important points have not been made clear. Thus the relationship between spherical harmonics and 2-D Fourier series will be discussed here. The associated Legendre function of integer degree and order can be written in an explicit form as (Heiskanen and Moritz, 1967, p. 24):

$$P_n^m(t) = 2^{-n}(1-t^2)^{m/2}\sum_{k=0}^{r}(-1)^k\frac{(2n-2k)!}{k!(n-k)!(n-m-2k)!}t^{n-m-2k}$$
$$= (\sin\theta)^m\sum_{k=0}^{r}a_k^{nm}(\cos\theta)^{n-m-2k} \qquad (4.1)$$

where $r = \left[\frac{n-m}{2}\right]$, $t = \cos\theta$, and a_k^{nm} is a simplified notation for the coefficients of the power series in t. The square bracket $[\cdot]$ in the upper limit of a summation stands for "the integer part" and it will be used throughout this report. Using Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$, we have:

$$(\cos\theta)^m = \frac{(e^{i\theta}+e^{-i\theta})^m}{2^m} = 2^{-m}\sum_{s=0}^{m}\binom{m}{s}e^{-i(m-2s)\theta} = 2^{-m}\sum_{s=0}^{m}\binom{m}{s}\cos(m-2s)\theta \qquad (4.2)$$

Similarly,

$$(\sin\theta)^m = \frac{(e^{i\theta}-e^{-i\theta})^m}{2^m i^m} = 2^{-m}\sum_{s=0}^{m}(-1)^{m-s}\binom{m}{s}\cos\left[(m-2s)\theta+\frac{m\pi}{2}\right]$$
$$= \begin{cases} (-1)^{m/2}2^{-m}\sum_{s=0}^{m}(-1)^s\binom{m}{s}\cos(m-2s)\theta, & m \text{ even} \\ (-1)^{(m-1)/2}2^{-m}\sum_{s=0}^{m}(-1)^s\binom{m}{s}\sin(m-2s)\theta, & m \text{ odd} \end{cases} \qquad (4.3)$$

In developing (4.2) and (4.3), the conjugate parts vanish due to the fact that $\binom{m}{s} = \binom{m}{m-s}$ and the sine function is an odd function. Note that the terms of the summations in (4.2) and (4.3) can be reduced to half their original ones if one wishes to get a compact form. The forms of (4.2) and (4.3) have the advantage of clear expression in the sine or cosine functions. Furthermore,

$$\sin\alpha\cos\beta = \frac{1}{2}\sin(\alpha+\beta) + \frac{1}{2}\sin(\alpha-\beta) \tag{4.4}$$

$$\cos\alpha\cos\beta = \frac{1}{2}\cos(\alpha+\beta) + \frac{1}{2}\cos(\alpha-\beta) \tag{4.5}$$

Therefore, $P_n^m(t)$ can be expressed in the following forms:

(1) When m is even

$$\begin{aligned} P_n^m(\cos\theta) &= (\sin\theta)^m \sum_{k=0}^{r} a_k^{nm} \sum_{j=0}^{n-m-2k} b_j^k \cos(n-m-2k-2j)\theta \\ &= \frac{1}{2}\sum_{k=0}^{r}\sum_{s=0}^{m}\sum_{j=0}^{n-m-2k} D_{ksj}^{nm}\big[\cos(n-2m-2k+2s-2j)\theta \\ &\quad + \cos(n-2k-2s-2j)\theta\big] = \sum_{q=0}^{[n/2]} T_q^{nm}\cos(n-2q)\theta \end{aligned} \tag{4.6}$$

(2) When m is odd

$$\begin{aligned} P_n^m(\cos\theta) &= \frac{1}{2}\sum_{k=0}^{r}\sum_{s=0}^{m}\sum_{j=0}^{n-m-2k} E_{ksj}^{nm}\big[\sin(n-2m-2k+2s-2j)\theta \\ &\quad + \sin(n-2k-2s-2j)\theta\big] = \sum_{q=0}^{[n/2]} S_q^{nm}\sin(n-2q)\theta \end{aligned} \tag{4.7}$$

In developing (4.6) and (4.7), the coefficients a, b, D, T, E and S are not explicitly given since the goal of the derivations is mainly to get the final form of the associated Legendre function in terms of the classical Fourier functions. The final identities in (4.6) and (4.7) are made by finding the upper bounds and lower bounds of $\ell_1 = n - 2(k+s+j)$ and $\ell_2 = n - 2m + 2s - 2(k+j)$. Apparently we get a maximum $\ell_1 = n$ when $k = s = j = 0$ and a minimum $\ell_1 = 0$ when $s = m$ and $(k+j) = \left[\frac{n-m}{2}\right]$; we get a maximum $\ell_2 = n$ when $s = m$, $k = j = 0$ and a minimum $\ell_2 = -n$ when $s = 0$ and $(k+j) = \left[\frac{n-m}{2}\right]$. Note that the maximum possible value of $(k+j)$ is $r = \left[\frac{n-m}{2}\right]$. If a negative ℓ_2 value is encountered, we may change the sign of the coefficient of the sine function to get a positive ℓ_2. Thus $P_n^m(\cos\theta)$ is a sum of the cosine functions or the sine functions, depending on whether m is even or odd. Such a result was stated in Ricardi and Burrows (1972), but no proof was given. A real function ζ defined on the sphere (at least it must be square integrable) can be expanded into spherical harmonics as:

$$\zeta(\theta,\lambda) = \sum_{n=0}^{Nmax} \sum_{m=0}^{n} (a_{nm}\cos m\lambda + b_{nm}\sin m\lambda) P_n^m(\cos\theta)$$
$$= \sum_{n=0}^{Nmax} \sum_{m=-n}^{n} F_{nm} e^{im\lambda} P_n^{|m|}(\cos\theta) \qquad (4.8)$$

with

$$F_{nm} = \frac{1}{2}(a_{nm} - ib_{nm}),\ F_{n,-m} = F_{n,m}^* = \frac{1}{2}(a_{nm} + ib_{nm}) \qquad (4.9)$$

where Nmax is the maximum degree of expansion, possibly infinite. In (4.8), the complex form is used in order to simplify the notations. Using Euler's formula and the results in (4.6) and (4.7), we may write (4.8) as:

$$\zeta(\theta,\lambda) = \sum_{n=0}^{Nmax} \sum_{m=-n}^{n} F_{nm} e^{im\lambda} \begin{cases} \sum_{q=0}^{[n/2]} T_q^{nm} \dfrac{e^{i(n-2q)\theta} + e^{-i(n-2q)\theta}}{2}, & m \text{ even} \\ + \\ \sum_{q=0}^{[n/2]} S_q^{nm} \dfrac{e^{i(n-2q)\theta} - e^{-i(n-2q)\theta}}{2i}, & m \text{ odd} \end{cases}$$

$$= \sum_{n=-Nmax}^{Nmax} \sum_{m=-Nmax}^{Nmax} C_{nm} e^{i(n\theta + m\lambda)}$$

$$= \sum_{n=0}^{Nmax} \sum_{m=0}^{Nmax} (\alpha_{nm}\cos n\theta\cos m\lambda + \beta_{nm}\sin n\theta\cos m\lambda$$

$$+ \gamma_{nm}\cos n\theta\sin m\lambda + \varepsilon_{nm}\sin n\theta\sin m\lambda) \qquad (4.10)$$

where C_{nm} are complex coefficients, α_{nm}, β_{nm}, γ_{nm} and ε_m are real coefficients. The relationships between the complex coefficients and the real coefficients are:

$$\left.\begin{aligned} C_{n,m} &= 1/4\,[\alpha_{nm} - \varepsilon_{nm} - i(\beta_{nm} + \gamma_{nm})] \\ C_{n,-m} &= 1/4\,[\alpha_{nm} + \varepsilon_{nm} - i(\beta_{nm} - \gamma_{nm})] \\ C_{-n,-m} &= 1/4\,[\alpha_{nm} - \varepsilon_{nm} + i(\beta_{nm} + \gamma_{nm})] \\ C_{-n,m} &= 1/4\,[\alpha_{nm} + \varepsilon_{nm} + i(\beta_{nm} - \gamma_{nm})] \end{aligned}\right\} \qquad (4.11)$$

where $0 \le n \le Nmax$, $0 \le m \le Nmax$. Therefore

$$C_{nm} = C_{-n,-m}^*,\ C_{n,-m} = C_{-n,m}^* \qquad (4.12)$$

In the above development, we have shown that a spherical harmonic expansion is equivalent to a 2-D trigonometric series. We must emphasize that the form in (4.10) is merely a 2-D trigonometric series, not a classical 2-D Fourier series. We will explain this below.

First of all, let us introduce the concept of surface integral, since almost all the geodetic practices are carried out on the surface of the Earth. The precise definition of surface integral may be found in Widder (1989, p. 232). Or, we may define that

$$\iint_S P(x, y, z)dS$$

is the surface integral of P(x, y, z) over the surface S. In the formulations of physical problem, the surface integral is a very important concept. For example, if we are given a surface S with a coating of density $\sigma = dM/dS$, then the surface integrals

$$\frac{1}{M}\iint_S (x, y, z)\sigma(S)dS$$

give the center of mass of the surface mass M with respect to the x, y, z coordinate system, where M is evaluated by

$$M = \iint_S \sigma(S)dS$$

Similarly, the evaluations of the coefficients for a 2-D Fourier series expansion of a periodic signal and the 2-D Fourier transform of a non-periodic signal are two examples of surface integral in which the surface is a plane. However, for the spherical coordinates (r, θ, λ), the integral

$$\iint K(\theta, \lambda)d\theta d\lambda \qquad (4.13)$$

is not a surface integral with respect to the kernel $K(\theta, \lambda)$ and has no physical meaning. Therefore, one should not state that the series in (4.10) is a 2-D Fourier series expansion. It merely happens to consist of the sine and cosine functions. Secondly, even if we treat θ and λ as planar coordinates, the functions $e^{i(n\theta+m\lambda)}$ are not orthogonal over $0 \le \theta \le \pi$, $0 \le \lambda \le 2\pi$ (the domain of a sphere). A 2-D Fourier series over that domain will be based on $e^{i(2n\theta+m\lambda)}$. The classical definition of a 2-D Fourier system would require the orthogonality relationship over the given domain, thus we again verify that the form in (4.10) is not a Fourier series. This important point is not made clear in Colombo's treatment (ibid).

According to the above discussion, we conclude that a spherical harmonic expansion can be transformed to a 2-D trigonometric series, but the corresponding 2-D trigonometric functions could not form an orthogonal basis over the entire sphere (but they

can be orthonormalized on a sphere). Furthermore, an orthogonal series such as the spherical harmonic series may be called a generalized Fourier series. We then ask ourself what the 2-D Fourier series over the entire sphere will be if it exists. From surface theory (Lipschutz, 1969, p. 174), it is clear that

$$\iint K(t, \lambda)dtd\lambda = \iint K(\theta, \lambda)\sin\theta d\theta d\lambda,\ t = \cos\theta \tag{4.14}$$

is a surface integral with respect to the kernel $K(\theta, \lambda)$ on the sphere. Based on such a concept, we now propose a 2-D Fourier system on the sphere defined as:

$$X_{nm}(\theta,\lambda) = e^{i(n\pi t+m\lambda)},\ n, m = \cdots, -2, -1, 0, 1, 2, \cdots, -1 \le t \le 1, 0 \le \lambda \le 2\pi \tag{4.15}$$

The system in (4.15) is orthonormal on the sphere since

$$(X_{nm}, X_{pq}) = \frac{1}{4\pi}\int_{t=-1}^{1}\int_{\lambda=0}^{2\pi} e^{i(n\pi t+m\lambda)} \cdot e^{-i(p\pi t+q\lambda)}dtd\lambda = \delta_{np}\delta_{mq} \tag{4.16}$$

Let us interpret the meaning of the system $\{X_{nm}(\theta, \lambda)\}$. Written in products of sine and cosine functions, the system $\{X_{nm}(\theta, \lambda)\}$ is equivalent to

$$\cos n\pi t \cos m\lambda,\ \sin n\pi t \cos m\lambda,\ \cos n\pi t \sin m\lambda,\ \sin n\pi t \sin m\lambda \tag{4.17}$$

The discussion is mainly related to the colatitude dependent function $\sin n\pi t$ and $\cos n\pi t$, since the longitude-dependent functions $\cos m\lambda$ and $\sin m\lambda$ have been treated in the case of spherical harmonic expansion and no special attention needs to be paid to them. Recalling the series expansion form of trigonometric functions, we have

$$\left.\begin{aligned} \sin n\pi t = \sin(n\pi\cos\theta) &= \sum_{k=0}^{\infty}(-1)^k\frac{(n\pi t)^{2k+1}}{(2k+1)!} \\ \cos n\pi t = \cos(n\pi\cos\theta) &= \sum_{k=0}^{\infty}(-1)^k\frac{(n\pi t)^{2k}}{(2k)!} \end{aligned}\right\} \tag{4.18}$$

Or, using the Bessel coefficients, we have (Tranter, p. 12, 1968)

$$\left.\begin{aligned} \sin n\pi t = \sin(n\pi\cos\theta) &= 2\sum_{k=0}^{\infty}(-1)^k J_{2k+1}(n\pi)\cos(2k+1)\theta \\ \cos n\pi t = \cos(n\pi\cos\theta) &= J_0(n\pi) + 2\sum_{k=1}^{\infty}(-1)^k J_{2k}(n\pi)\cos 2k\theta \end{aligned}\right\} \qquad (4.19)$$

where $J_k(n\pi)$ is the Bessel function of integer order defined in (C.7). Thus the functions sin nπt and cos nπt are two polynomials of infinite degree in $t = \cos\theta$ (obviously (4.19) can be transformed to polynomials of t). It is possible to have an orthogonal system formed by polynomials of infinite degree. One example is the Bessel function (see (C.7) in Appendix C), another one is the Legendre function of non-integer degree (see (3.89)). Therefore, one should not have too much difficulty accepting the orthogonal system formed by sin nπt and cos nπt. Now the theoretical questions are: What is the differential equation associated with $\sin n\pi t$ and $\cos n\pi t$? Does the series converge? It is easy to answer these two questions if we treat the functions sin nπt and cos nπt just as the regular sine and cosine functions. The differential equation associated with them is $h''(t) + (n\pi)^2 h(t) = 0$ which describes a special type of eigenvalue-eigenfunction problem. Since the eigenfunctions form a complete system (or in this case a separable Hilbert space), convergence in the mean is guaranteed.

Next we have to settle the sampling problem for the expansion using the system $\{X_{nm}(\theta, \lambda)\}$. Let a function $f(\theta, \lambda)$ be expanded into $X_{nm}(\theta, \lambda)$ as

$$f(\cos\theta,\lambda) = f(t,\lambda) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} A_{nm} e^{i(n\pi t + m\lambda)} \qquad (4.20)$$

then the coefficients A_{nm} are found by

$$A_{nm} = \frac{1}{4\pi}\int_{t=-1}^{1}\int_{\ell=0}^{2\pi} f(\theta,\lambda)e^{-i(n\pi t + m\lambda)}\,dt\,d\lambda \qquad (4.21)$$

If the data are sampled at intervals $\Delta t = 2/N$ and $\Delta\lambda = 2\pi/M$, then the coefficients A_{nm} are found by

$$A_{nm} = \frac{1}{MN}\sum_{k=0}^{N-1}\sum_{\ell=0}^{M-1} f_{k\ell} e^{-2\pi i(nk/N + m\ell/M)} \qquad (4.22)$$

Thus A_{nm} can be calculated by a regular two-dimensional FFT routine. The highest n and m, or the Nyquist frequencies in co-latitude and longitude, will be N/2 and M/2, respectively. The expansion in the form in (4.22) is not without problem. First of all, in the standard sampling process on the sphere such as that in Rapp (1986), the constant interval along latitudinal direction is measured in θ, not t, namely Δθ is a constant. Such a sampling process has the advantage of easily dividing the sphere into sub-domains

bounded by constantly increasing parallels and meridians. A constant interval Δt will imply that

$$\Delta t \doteq -\sin\theta\Delta\theta = \cos\phi\Delta\phi \tag{4.23}$$

where ϕ is the geocentric latitude. Thus the value $\Delta\phi$ must be adjusted according to the latitude ϕ to get the constant value Δt. This will create problems when using the existing data bases which are constructed basically using the sampling strategy in Rapp (1986).

The second problem is related to the frequency interpretation. The use of a constant $\Delta\theta$ in data sampling yields a meaningful interpretation of frequency for a signal on the sphere. As a signal travels on the sphere in the north-south direction, the cycles per unit distance are equivalent to the cycles per unit angle subtended at the center of the sphere. Thus a constant $\Delta\theta$ in a spherical case is analogous to a constant distance in a planar case. On the other hand, the use of t cannot justify a good frequency interpretation, at least geometrically. It can only be regarded as a transformed coordinate by the mapping $t = \cos\theta$.

One way to overcome the sampling problem is to sacrifice the convenient FFT form in (4.22) and use the classical sampling form in a spherical harmonic expansion, namely, we could write a discrete form of (4.21) as

$$\begin{aligned} A_{nm} &= \frac{1}{4\pi}\sum_{k=0}^{N-1}\sum_{\ell=0}^{2N-1}\iint_{\Delta\sigma_{k\ell}} f(\theta,\lambda)e^{-i(n\pi t+m\lambda)}dtd\lambda \\ &\doteq \frac{1}{4\pi}\sum_{k=0}^{N-1} U_k^n V_m \sum_{\ell=0}^{2N-1} \bar{f}_{k\ell}e^{-im\ell\Delta\lambda} \end{aligned} \tag{4.24}$$

where

$$\left.\begin{aligned} U_n^k &= \begin{cases} \cos k\Delta\theta - \cos(k+1)\Delta\theta\ , & n = 0 \\ \frac{i}{n\pi}\left(e^{-in\pi\cos k\Delta\theta} - e^{-in\pi\cos(k+1)\Delta\theta}\right), & n \neq 0 \end{cases} \\ V_m &= \begin{cases} \Delta\lambda = \lambda_{\ell+1} - \lambda_\ell\ , & m = 0 \\ \frac{i(e^{im\Delta\lambda} - 1)}{m}\ , & m \neq 0 \end{cases} \end{aligned}\right\} \tag{4.25}$$

The expansion method in (4.24) is completely analogous to Rapp's method for the global geopotential solution using mean gravity anomalies (Rapp, 1986, p. 368). The quantity $\bar{f}_{k\ell}$ is the mean value of f in an equiangular block $\Delta\theta$ x $\Delta\lambda$. In such a sampling strategy, the sphere is divided into N x 2N equiangular blocks in latitudinal and longitudinal directions. The indices k, ℓ show the locations of the blocks and the sampled values $\bar{f}$. However, as pointed out by Bath (1974, p.112), "it is only in the case of the Fourier series and Fourier transform and the spherical harmonic analysis that we are justified in talking about frequency and spectra in the usual sense of these terms", thus the problem of

frequency interpretation for the system $\{X_{nm}(\theta,\lambda)\}$ remains regardless of the sampling strategies used. Needless to say, the orthogonal systems we have discussed in Chapter 3 all suffer the same problem in interpreting the frequencies.

However, in the approximation theory, a more important issue is the convergence of the series expansion in the mean, not too much attention is paid to frequency. For some signals, the Fourier expansion or the surface spherical harmonic expansion cannot guarantee the fastest convergence, examples can be found in Bath (1974). To compromise the requirements of frequency interpretation and convergence, one could use a different measure for frequency, for example, the "density of nodes" or eigenvalue such as in the Sturm-Liouville problem is one way to measure the frequency (see Section 3.3.3). Having such a "generalized" frequency, we now can enjoy a good frequency interpretation and a possibly faster convergence using an orthogonal series other than the classical Fourier series. Further, in the study of membrane vibrations in physics, the eigenvalues of the Helmholtz equation are associated with the "eigenfrequencies". Physicists adopt the idea that the higher the frequency, the finer the subdivision of the membrane into regions of alternating signs (Sommerfield, 1949). The interpretation of frequency in this way indeed makes sense to us, since by such a interpretation we will know the degree of oscillations of the orthogonal functions just as we know the number of cycles of the Fourier functions in some unit distance (or time).

4.2 Are the Spherical Harmonics the Unique Set of Orthogonal Functions on a Sphere?

We have essentially answered this question by introducing the system $\{X_{nm}(\theta,\lambda)\}$ on a sphere in the previous section. Therefore, the surface spherical harmonics are not the unique set of orthogonal functions on a sphere. We shall treat this problem in a more systematic way. We must first recognize that basically, we are dealing with a set of two-dimensional orthogonal functions. The natural choice of the two variables is the spherical coordinates θ, λ. Since the boundary of the domain formed by $0 \le \theta \le \pi$ and $0 \le \lambda \le 2\pi$ is rectangular, one can follow the usual technique of searching a 2-D orthogonal function and separate the desired orthogonal functions into a θ-dependent orthogonal function and a λ-dependent orthogonal function. Further, due to the consideration of the integral on the sphere, it is natural to transform the θ-dependent function to the t-dependent function, where $t = \cos\theta$. In order to maintain the periodicity on the sphere, at least one function must be periodic. In accordance with the use of the spherical harmonics, the λ-dependent functions can be chosen to be $\cos m\lambda$ and $\sin m\lambda$, where m must be integer, either negative or positive. So what is left is the t-dependent function. Now we may also simplify the question to: Can we find any t-dependent orthogonal functions other than the associated Legendre function $P_n^m(t)$ in the interval $-1 \le t \le 1$?

In searching the t-dependent orthogonal functions over $-1 \le t \le 1$, one also has to keep in mind that they must form a complete set (with respect to the topology in a separable Hilbert space), a requirement stated in (2.20). If the completeness in the t-dependent functions is ensured, then the products of the t-dependent functions and the functions $\cos m\lambda$, $\sin m\lambda$ form a complete set in $-1 \le t \le 1$ and $0 \le \lambda \le 2\pi$, since the functions $\cos m\lambda$, $\sin m\lambda$ have already formed a complete set over $0 \le \lambda \le 2\pi$. This follows the principle stated in Kaplan (1981, p. 177). Checking whether a set is complete is equivalent to checking whether a set is closed. A set of orthonormal functions $\{\psi_j\}$ is said to be closed if no normalized function is orthogonal to every function in the set. In other words, for a set $\{\psi_j\}$, if we can find at least one function $f \neq 0$ such that

$$(f, \psi_j) = 0 \quad \text{for all } j \tag{4.26}$$

then the set $\{\psi_j\}$ is not closed; if no such function exists, it is closed or complete. The completeness is a necessary property for a separable Hilbert space. If we have an incomplete set such as

$$\{\psi_j\} = \{1, \sin x, \cos x, \cos 2x, \sin 3x, \ldots\}, \; -\pi \le x \le \pi$$

then the series expansion into ψ_j will never converge for a function $f = \sin 2x$, since the expansion coefficients (f, ψ_j) will be always zero and the expansion will be $f = 0 + 0 + \cdots$.

Another thing we must remember is that our orthogonal functions, either in a one-dimensional form or in a two-dimensional form, have no obligation to fulfill any differential equations nor must they be solutions in the potential theory or the quantum theory (such as Schrödinger's equation). This somewhat gives us more flexibility in finding the orthogonal functions. With these understandings in mind, we can discuss the tools of finding the t-dependent orthogonal functions and answer our question through the following discussion.

(1) Find orthogonal functions from orthogonal polynomials and their weight function. We recall that the scalar inner product with respect to a weight function is

$$(P_n(t), P_m(t)) = \int_{-1}^{1} P_n(t) P_m(t) w(t) d(t) \;, \; w(t) \ge 0 \tag{4.27}$$

where $w(t)$ is a weight function, which could be infinite at the endpoints $t = \pm 1$. $P_n(t)$ and $P_m(t)$ in this case are polynomials of degree n and m, respectively. If the weight function $w(t)$ is equal to 1, then the orthogonal functions for which we are looking become orthogonal polynomials; if it is not equal to 1, then the desired orthogonal functions are

$$f_n(t) = \sqrt{w(t)}\, P_n(t) \tag{4.28}$$

Since

$$(f_n(t), f_m(t)) = \int_{-1}^{1} \left(\sqrt{w(t)}\, P_n(t)\right)\left(\sqrt{w(t)}\, P_m(t)\right) dt \tag{4.29}$$

Thus $f_n(t)$ is no more a polynomial if $\sqrt{w(t)}$ is not a polynomial. From this discussion, we thus know that it is very important to specify the weight function when we are talking about a polynomial. For example, for the Legendre polynomials, $w(t) = 1$; for the Tschebyscheff polynomials of the first kind, $w(t) = (1 - t^2)^{-1/2}$; for the Tschebyscheff polynomial of the second kind, $w(t) = (1 - t^2)^{1/2}$. If we require that all the inner products possess $w(t) = 1$, then only the Legendre polynomial is really a polynomial. It appears that all the polynomials (with respect to their weights) we have mentioned can be formed by the Gram-Schmidt process (Davis, 1975, p. 246) using the sequence $1, t, t^2, \cdots$ and one weight function corresponds to one type of polynomial. The orthogonal functions are obtained by just multiplying the square of the weight and the polynomials.

Even the associated Legendre function $P_n^m(t) = (1 - t^2)^{m/2} d^m P_n / dt^m$ falls into this category. Now the weight function is $(1 - t^2)^m$ and the polynomial is $d^m P_n / dt^m$, and amazing enough, the Gram-Schmidt process in this case indeed leads to a polynomial which is just the derivative of the Legendre polynomial, up to a constant factor (see Davis, 1975, p. 246, or Morse and Feshbach, 1953, p. 782). In fact, the (associated) Legendre function and Tschebyscheff polynomials belong separately to two families of the Jacobi polynomials defined by Davis (1975, p. 246)

$$P_n^{\alpha,\beta}(t) = \frac{(-1)^n}{2^n n!}(1-t)^{-\alpha}(1-t)^{-\beta}\frac{d^n}{dt^n}\left\{(1-t)^{n+\alpha}(1-t)^{n+\beta}\right\} \tag{4.30}$$

with the weight function

$$w(t) = (1-t)^{\alpha}(1+t)^{\beta} \tag{4.31}$$

where α and β are the parameters that define a family. For the (associated) Legendre function, $\alpha = \beta = m$; for the Tschebyscheff polynomial of the first kind, $\alpha = \beta = -1/2$; for the Tschebyscheff polynomial of the second kind, $\alpha = \beta = 1/2$.

We are now in a position to propose an orthonormal system on the sphere based on the Tschebyscheff polynomials and the functions $\cos m\lambda$, $\sin m\lambda$. The explicit form of the Tschebyscheff polynomial of the first kind is (Beyer, 1987)

$$T_n(t) = \frac{n}{2}\sum_{j=0}^{[n/2]}(-1)^j\frac{(n-j-1)!}{j!(n-2j)!}(2t)^{n-2j} = \cos\left(n\cos^{-1}t\right) \tag{4.32}$$

or

$$T_n(\cos\theta) = \cos n\theta \tag{4.33}$$

and the Teschebyscheff polynomial of the second kind is (ibid.)

$$U_n(t) = \sum_{j=0}^{[n/2]}(-1)^j\frac{(n-j)!}{j!(n-2j)!}(2t)^{n-2j} \tag{4.34}$$

or

$$U_n(\cos\theta) = \frac{\sin\left[(n+1)\theta\right]}{\sin\theta},\ U_{-1}(\cos\theta) = 0 \tag{4.35}$$

Now, the following Fourier base functions are orthogonal with respect to integration over $0 \le \theta \le \pi$:

$$\{\cos 2n\theta,\ \sin 2n\theta\},\quad n = 0,1,2,\ldots \tag{4.36}$$

Further,

$$T_{2n}(t) = \cos 2n\theta \tag{4.37}$$

$$U_{2n-1}(t) = \sin 2n\theta/\sin\theta \tag{4.38}$$

Recalling the weight functions for $T_{2n}(t)$ and $U_{2n-1}(t)$ and using the orthogonal system in (4.36), we get the following complete orthogonal system with respect to $w(t) = 1$ over $-1 \le t \le 1$:

$$\overline{T}_n(t) = (1 - t^2)^{-1/4}\, T_{2n}(t) = \cos 2n\theta/\sqrt{\sin\theta} \tag{4.39}$$

$$\overline{U}_n(t) = (1 - t^2)^{1/4}\, U_{2n-1}(t) = \sin 2n\theta/\sqrt{\sin\theta} \tag{4.40}$$

The functions $\overline{T}_n(t)$ and $\overline{U}_n(t)$ are plotted in Figure 4.1. Further, the normalizing factors are

$$\left.\begin{aligned}
&\int_{-1}^{1} \overline{T}_n^2(t)dt = \begin{cases} \pi, & n = 0 \\ \pi/2, & n > 0 \end{cases} \\
&\int_{-1}^{1} \overline{U}_n^2(t)dt = \pi/2,\ n > 0 \\
&\int_{0}^{2\pi} \cos^2 m\lambda d\lambda = \begin{cases} 2\pi, & m = 0 \\ \pi, & m > 0 \end{cases} \\
&\int_{0}^{2\pi} \sin^2 m\lambda d\lambda = \pi,\ m > 0
\end{aligned}\right\} \tag{4.41}$$

Combining $\overline{T}_n(t)$, $\overline{U}_n(t)$ with $\cos m\lambda$, $\sin m\lambda$ and taking into account the normalizing factors, we finally obtain a complete orthonormal system on a sphere $-1 \le t \le 1, 0 \le \lambda \le 2\pi$ as follows:

$$\left.\begin{aligned}
S^1_{nm} &= \alpha_{nm}\overline{T}_n \cos m\lambda,\ S^2_{nm} = \alpha_{nm}\overline{U}_n \cos m\lambda \\
S^3_{nm} &= \alpha_{nm}\overline{T}_n \sin m\lambda,\ S^4_{nm} = \alpha_{nm}\overline{U}_n \sin m\lambda
\end{aligned}\right\} \tag{4.42}$$

where

$$\alpha_{nm} = \begin{cases} \sqrt{2}/\sqrt{\pi}, & \text{if } n = m = 0 \\ 2/\sqrt{\pi}, & \text{if } n > 0,\ m = 0 \text{ or } n = 0,\ m > 0 \\ 2\sqrt{2}/\sqrt{\pi}, & \text{if } n > 0,\ m > 0 \end{cases} \tag{4.43}$$

The system $\{S^{\gamma}_{nm}(\theta,\lambda)\}$ is orthonormal on a sphere in the sense

$$(S^k_{nm},\ S^\ell_{rs}) = \frac{1}{4\pi}\int_{\theta=0}^{\pi}\int_{\lambda=0}^{2\pi} S^k_{nm}S^\ell_{rs}\sin\theta\, d\theta d\lambda = \delta_{k\ell}\delta_{nr}\delta_{ms} \tag{4.44}$$

A square integrable function $f(\theta,\lambda)$ defined on a sphere can be expanded into S^γ_{nm} as:

$$f(\theta,\lambda) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sum_{\gamma=1}^{4} a^\gamma_{nm}S^\gamma_{nm}(\theta,\lambda) \tag{4.45}$$

and the coefficients are found by

$$a^\gamma_{nm} = (f, S^\gamma_{nm}) \tag{4.46}$$

where the inner product is defined as in (4.44). The series in (4.45) is designated as 2-D Fourier-Tschebyscheff series.

Using the results in (4.39), (4.40) and (4.43), we can rewrite (4.46) in the following explicit form:

$$\begin{Bmatrix} a^1_{nm} \\ a^2_{nm} \\ a^3_{nm} \\ a^4_{nm} \end{Bmatrix} = \frac{\alpha_{nm}}{4\pi}\int_0^{\pi}\int_0^{2\pi} f(\theta,\lambda)\begin{Bmatrix} \cos 2n\theta\cos m\lambda \\ \sin 2n\theta\cos m\lambda \\ \cos 2n\theta\sin m\lambda \\ \sin 2n\theta\sin m\lambda \end{Bmatrix}\sqrt{\sin\theta}\, d\theta d\lambda \tag{4.47}$$

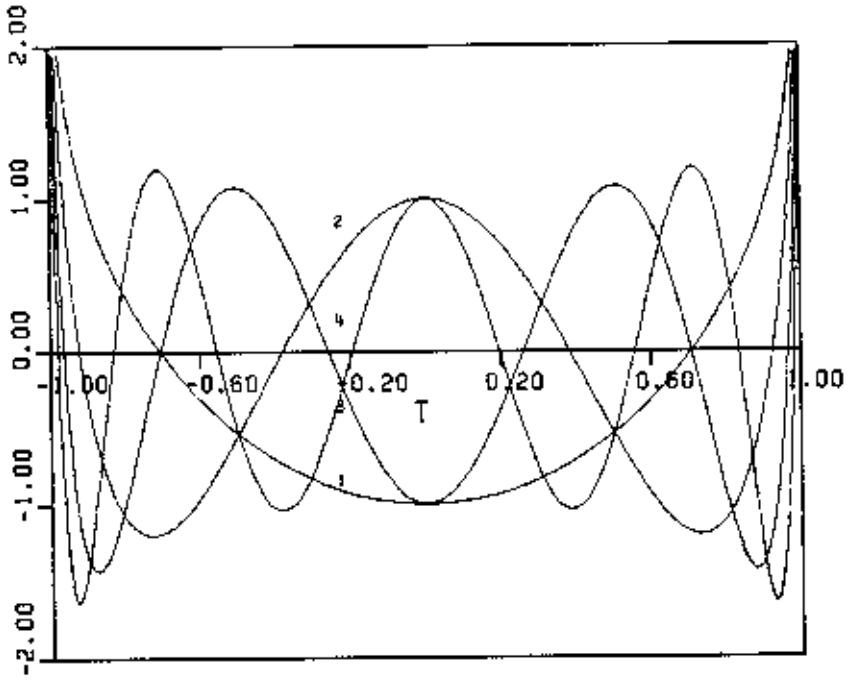
The simple form in (4.47) is due to the use of $t = \cos\theta$ in the Tschebyscheff polynomials. It is also necessary to point out that the function $T_n(t)$ has two singular points $t = \pm 1$, but the expansion in (4.45) is still valid and will be convergent (in the mean but not uniformly) due to the fact that $T_n(t)$ is square integrable over $-1 \le t \le 1$ (see Sansone, 1959, Chapter 1).

By the use of orthogonal polynomials and their weights, we have successfully constructed an orthonormal system on a sphere. Now we turn to the second tool of finding orthogonal functions over $-1 \le t \le 1$.

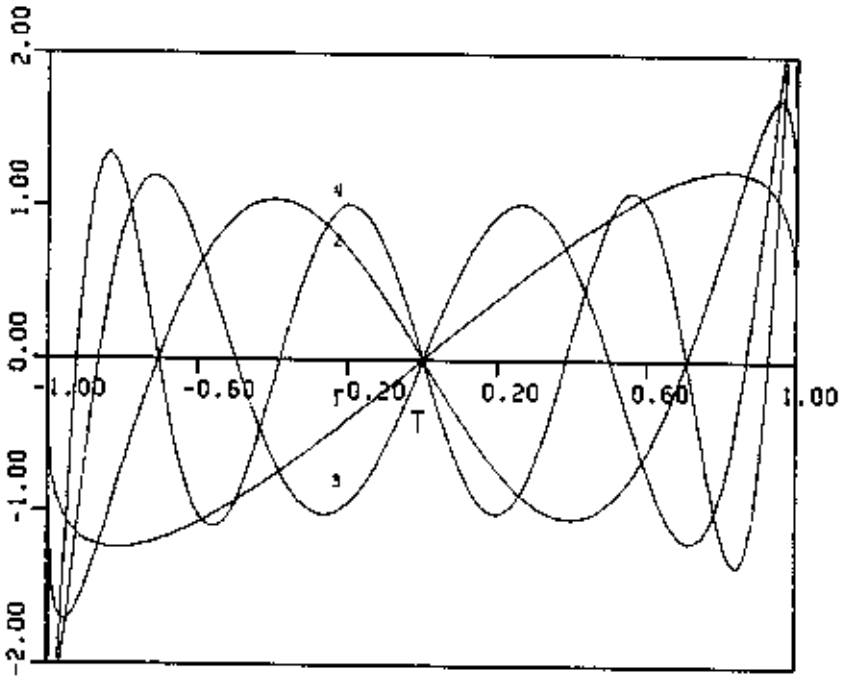
(2) Find orthogonal functions by solving the self-adjoint Sturm-Liouville problem. We recall the one-dimensional Sturm-Liouville problem from Chapter 3:

$$\frac{d}{dt}\left(p(t)\frac{d}{dt}u(t)\right) - q(t)u(t) + kw(t)u(t) = 0 \tag{4.48}$$

where $u(t)$ is the desired function, $p(t)$ and $q(t)$ are two given functions of t, $w(t)$ is the weight function and k is the eigenvalue. If $u_1(t)$ and $u_2(t)$, together with k_1 and k_2, are two functions satisfying (4.48), then



(a)



(b)

Figure 4.1 (a) Function $\overline{T}_n(t)$ (b) Function $\overline{U}_n(t)$

$$(k_1 - k_2)\int_{-1}^{1} u_1(t)u_2(t)w(t)dt = p(t)\left(u_1u_2' - u_1'u_2\right)\Big|_{-1}^{1} \tag{4.49}$$

as demonstrated in Chapter 3. Suppose k_1 and k_2 are two distinct values, then we can show that $u_1(t)$ and $u_2(t)$ are mutually orthogonal if the right side of (4.49) vanishes. We have done this in finding the spherical cap harmonics in Section 3.3.2.1. We shall apply the same technique over $-1 \le t \le 1$ and point out more theoretical problems.

Apparently, in the following two cases, the right side of (4.49) vanishes:

(a) $u(k, t)$ such that $\alpha u(k, t) + \beta u'(k, t) = 0$ at $t = \pm 1$,

α and β not simultaneously zero, $p(t) \neq 0$ over $-1 \le t \le 1$.

(b) $p(t) = 0$ at $t = \pm 1$.

In the first case, we essentially are specifying the boundary conditions. The function $u(k, t)$ is so written that its association with the eigenvalue is clear. One choice of functions $p(t)$, $q(t)$ and $w(t)$ then corresponds to one family of orthogonal functions. For the solution $u(k, t)$, the most general technique is the method of series solution in which the function $u(k, t)$ is assumed to take the form (Dettman, 1988, p. 196)

$$u(k,t) = \sum_{j=0}^{\infty} a_j t^j \tag{4.50}$$

Then the series form is substituted into the differential equation to find the coefficients. We may sometimes obtain a finite series, depending on functions $p(t)$, $q(t)$, $w(t)$ and the eigenvalue. Since we impose the condition that $p(t) \neq 0$ over $-1 \le t \le 1$, all the points over $-1 \le t \le 1$ are regular points and two independent solutions for $u(k, t)$ can be obtained (Kaplan, 1981, p. 694). For a convenient discussion, we assume the boundary condition is $u(k, \pm 1) = 0$. Without loss of generality, the solution $u(k, t)$ may be a linear combination of the two independent solutions $v_1(k, t)$, $v_2(k, t)$ in the form

$$u(k, t) = v_1(k, t) + bv_2(k, t) \tag{4.51}$$

Now we can treat the eigenvalue k and the coefficient b as two unknowns with which the B.C. $u(k, \pm 1) = 0$ can be satisfied. Therefore,

$$\left.\begin{aligned} v_1(k,1) + bv_2(k, 1) &= 0 \\ v_1(k,-1) + bv_2(k, -1) &= 0 \end{aligned}\right| \tag{4.52}$$

Multiplying one equation by the other function, then subtracting, one gets

$$f(k) = v_1(k, 1)\, v_2(k, -1) - v_1(k, -1)\, v_2(k, 1) = 0$$

So the problem becomes the finding of the roots of f(k). Once k is found, then b is found from (4.52). Each k corresponds to one eigenfunction u(k,t) and the orthogonal system is formed by all possible eigenfunctions.

Now we discuss the solution for u(t) in Case b. In this case, $t = \pm 1$ are two singular points of the differential equation. Denoting the singular points as t_0, we can say that t_0 is a regular singular point if $p(t_0) = 0$ and $p'(t_0) \neq 0$, or $p(t_0) = 0$, $p'(t_0) \neq 0$ and $p''(t_0) = 0$ (Kaplan, 1981). At the neighborhood of a regular singular point, one series solution as discussed in Case a will always exist (ibid.). Normally, the method to be used in such a case is Frobenius' method. Since this method is quite involved and the purpose here is just to show that we can find the orthogonal system through this approach, no further elaboration will be made.

From the discussion in (1) and (2), we have presented here one orthonormal system explicitly on the (entire) sphere. If one is interested in a global expansion of a function on a sphere, then the system $\{X_{nm}(\theta,\lambda)\}$ in the previous section or the system $\{S^{Y}_{nm}(\theta,\lambda)\}$ in this current section may serve as an alternative for the spherical harmonics. Also, we again conclude that the spherical harmonics are not the unique set of orthogonal functions on a sphere.

In addition to the two ON systems (on the entire sphere) developed in this chapter, some other systems of functions are worthy to be recommended. If the data are discretely given on a sphere, the spherical spline functions introduced by Freeden (1981, eqs. (5.1) and (5.2)) can be used for the purpose of interpolation and approximation of the data. Freeden's spherical spline functions represent combinations of spherical harmonics and Green's kernel functions (see Freeden, 1981, Section 4 for the definitions) and are suitable for both global and local interpolation and approximation. Furthermore, if we approximate the earth's surface by an ellipsoid of revolution, then the normalized spheroidal surface harmonics developed by Thong and Grafarend (1989, eqs. 3(9i) and 3(9ii)) can be used for expanding a function defined on such a surface. Also, Thong and Grafarend's functions could be the candidate of a function system that can be used for the orthonormalizing process over the oceans discussed in Chapter 3.

This chapter has been devoted to finding orthogonal functions on a sphere. Currently global functions such as the spherical harmonics are widely used in modeling the oceanic signal such as SST. Problems about the correlations between the SST harmonic coefficients and satellite orbit errors have been reported when a simultaneous geoid-SST estimation scheme in altimetry is used (e.g., Denker and Rapp, 1990). If one is still trying to use global functions to model the SST in the simultaneous scheme, the global functions proposed in this chapter could be used and hopefully the correlations can be reduced.

5. Gram-Schmidt Orthonormalizing Process Over the Oceans Using Spherical Harmonics

As suggested in the last section of Chapter 3, we will construct a set of orthonormal functions using the Gram-Schmidt process. The functions to be used for such a process are chosen to be the surface spherical harmonics. Some properties of the constructed orthonormal functions will also be presented.

5.1 The Inner Products of Spherical Harmonics over the Oceans

5.1.1 The Complete Formulae of Inner Products

In Chapter 3, we have listed two algorithms for the construction of orthonormal functions over the oceans using the Gram-Schmidt process. It remains to calculate the inner products of the given functions from which the orthonormal functions are constructed. We will choose the spherical harmonics to be the given functions. This means that we need to calculate the inner products of the spherical harmonics for our purpose. As pointed out in Chapter 3, we will be using the elevation data to define the oceans. In practice, the ocean is subdivided into equiangular blocks along the latitudinal and longitudinal directions. Once certain criteria for defining the oceans are determined, we can have an index function $w_{k\ell}$ for land and oceans defined as

$$w_{k\ell} = \begin{cases} 1, & \text{ocean} \\ 0, & \text{land} \end{cases} \tag{5.1}$$

where k is an index along latitude, and ℓ an index along longitude. Let functions f, g be two elements of a function space defined on the entire sphere. The inner product of f, g over the oceans (a subdomain of a sphere) is

$$(f, g) = \frac{1}{\sigma} \iint_{\text{oceans}} fg^* \, d\sigma = \frac{1}{\sigma} \sum_{k=0}^{N-1} \sum_{\ell=0}^{2N-1} w_{k\ell} \iint_{\Delta\sigma_{k\ell}} fg^* d\sigma \tag{5.2}$$

where σ is the area of the oceans. The number of blocks on the earth will be always N x 2N if equiangular blocks are used. The formula in (5.2) is rigorous according to the fundamental theorems of calculus, provided that the ocean is really a collection of equiangular blocks. (This is only an ideal case, see Chapter 3). In (5.2), f and g can be complex functions, and * is the conjugate operator.

Let us use the notations

$$\bar{R}_{nm} = \bar{P}_n^m(\cos\theta)\cos m\lambda, \; \bar{S}_{nm} = \bar{P}_n^m(\cos\theta)\sin m\lambda \tag{5.3}$$

for the surface spherical harmonics, where $\bar{P}_n^m(\cos\theta)$ denotes the fully normalized Legendre function. Further, for the reason that will be clear in the following development, we define a complex form of the surface spherical harmonics:

$$\bar{Y}_{nm} = \bar{P}_n^{|m|}(\cos\theta)e^{im\lambda} \tag{5.4}$$

Now our goal is to evaluate four kinds of inner products over the oceans A_{nm}^{rs}, B_{nm}^{rs}, C_{nm}^{rs} and D_{nm}^{rs} as follows:

$$\begin{Bmatrix} A^{rs}_{nm} \\ B^{rs}_{nm} \\ C^{rs}_{nm} \\ D^{rs}_{nm} \end{Bmatrix} = \begin{Bmatrix} (\bar{R}_{nm}, \bar{R}_{rs}) \\ (\bar{S}_{nm}, \bar{S}_{rs}) \\ (\bar{R}_{nm}, \bar{S}_{rs}) \\ (\bar{S}_{nm}, \bar{R}_{rs}) \end{Bmatrix} = \frac{1}{\sigma} \iint_{oceans} \bar{P}^m_n(t)\bar{P}^s_r(t) \begin{Bmatrix} \cos m\lambda \cos s\lambda \\ \sin m\lambda \sin s\lambda \\ \cos m\lambda \sin s\lambda \\ \sin m\lambda \cos s\lambda \end{Bmatrix} dtd\lambda \tag{5.5}$$

where $t = \cos\theta$. We will try to use the inner products of $\bar{Y}_{nm}$ in (5.4) and develop a compact and efficient formula for computing the desired inner products.

Using the formulae

$$\left.\begin{aligned} e^{i\theta} &= \cos\theta + i\sin\theta \\ \cos(\alpha - \beta) &= \cos\alpha\cos\beta + \sin\alpha\sin\beta \\ \sin(\alpha - \beta) &= \sin\alpha\cos\beta - \cos\alpha\sin\beta \end{aligned}\right\} \tag{5.6}$$

it is easy to see that

$$a = (\bar{Y}_{nm}, \bar{Y}_{rs}) = \frac{1}{\sigma} \iint_{oceans} P^{|m|}_n(t)P^{|s|}_r(t)e^{i(m-s)\lambda} dtd\lambda$$

$$= A + B - i(C - D) \tag{5.7}$$

where we have dropped the subscripts and superscripts for A, B, C and D in order to economize the notations. They will be present in the final equation at the end of this section. Similar to (5.7), we also have

$$\begin{aligned} b &= (\bar{Y}_{nm}, \bar{Y}_{r,-s}) = A - B + i(C+D) \\ c &= (\bar{Y}_{n,-m}, \bar{Y}_{r,s}) = A - B - i(C+D) \\ d &= (\bar{Y}_{n,-m}, \bar{Y}_{r,-s}) = A + B + i(C-D) \end{aligned} \tag{5.8}$$

Regarding (5.7) and (5.8) as a system of equations, we may solve for A, B, C and D in terms of a, b, c and d:

$$\begin{aligned} A &= \frac{1}{4}(a+b+c+d) \\ B &= \frac{1}{4}(a-b-c+d) \\ C &= \frac{1}{4i}(-a+b-c+d) \\ D &= \frac{1}{4i}(a+b-c-d) \end{aligned} \tag{5.9}$$

Thus, once a, b, c and d are found, our goal is achieved. It turns out that a very simple formula can be developed for a, b, c and d. To see this, we rewrite $a = (Y_{nm}, Y_{rs})$ using the form in (5.2):

$$a = \frac{1}{\sigma}\sum_{k=0}^{N}\sum_{\ell=0}^{2N-1} w_{k\ell}\left(\int_{\theta_k}^{\theta_{k+1}} \overline{P}_n^{|m|}(\cos\theta)\overline{P}_r^{|s|}(\cos\theta)\sin\theta d\theta\right)\left(\int_{\lambda_\ell}^{\lambda_{\ell+1}} e^{i(m-s)\lambda}d\lambda\right)$$

$$= \frac{1}{\sigma}\sum_{k=0}^{N} I^k_{nmrs} f(m,s) \sum_{\ell=0}^{2N-1} w_{k\ell} J_\ell(m,s) \qquad (5.10)$$

where

$$I^k_{nmrs} = \int_{\theta_k}^{\theta_{k+1}} \overline{P}_n^{|m|}(\cos\theta)\,\overline{P}_r^{|s|}(\cos\theta)\sin\theta d\theta \qquad (5.11)$$

(Note: the indices n, m, r, s in I^k_{nmrs} are all positive)

$$f(m,s) = \begin{cases} \Delta\lambda = \lambda_{\ell+1} - \lambda_\ell\ , & m = s \\ i\left(1 - e^{i(m-s)\Delta\lambda}\right)/(m-s)\ , & i = \sqrt{-1}\ ,\quad m \neq s \end{cases} \qquad (5.12)$$

$$J_\ell(m,s) = e^{i(m-s)\ell\Delta\lambda} = e^{2\pi i\left[(m-s)\ell/2N\right]} \qquad (5.13)$$

If we write

$$F_k(m,s) = \sum_{\ell=0}^{2N-1} w_{k\ell} J_\ell(m,s) = \sum_{\ell=0}^{2N-1} w_{k\ell} e^{2\pi i\left[(m-s)\ell/2N\right]} \qquad (5.14)$$

which is the discrete Fourier transform of $w_{k\ell}$ at frequency (m-s), then

$$a = \frac{1}{\sigma}\sum_{k=0}^{N-1} I^k_{nmrs} f(m,s) F_k(m,s) \qquad (5.15)$$

Following the development for a, we can easily verify that

$$\begin{Bmatrix} b \\ c \\ d \end{Bmatrix} = \frac{1}{\sigma}\sum_{k=0}^{N-1} I^k_{nmrs} \begin{Bmatrix} f(m,-s)\,F_k(m,-s) \\ f(-m,s)\,F_k(-m,s) \\ f(-m,-s)\,F_k(-m,-s) \end{Bmatrix} \qquad (5.16)$$

Due to the definitions of $F_k(m,s)$ and $f(m,s)$, we can see that

$$\begin{aligned} F_k(-m,-s) &= F_k^*(m,s),\ F_k(-m,s) = F_k^*(m,-s) \\ f(-m,-s) &= f^*(m,s),\ f(-m,s) = f^*(m,-s) \end{aligned} \qquad (5.17)$$

Therefore

$$a+b = (c+d)^*,\ a-b = (d-c)^* \tag{5.18}$$

If we write

$$\begin{aligned} U_k &= f(m,s)F_k(m,s) + f(m,-s)F_k(m,-s) \\ V_k &= f(m,s)F_k(m,s) - f(m,-s)F_k(m,-s) \end{aligned} \tag{5.19}$$

then, using (5.9) and (5.18), finally we get

$$\begin{aligned} \begin{Bmatrix} A^{rs}_{nm} \\ B^{rs}_{nm} \\ C^{rs}_{nm} \\ D^{rs}_{nm} \end{Bmatrix} &= \frac{1}{4\sigma} \sum_{k=0}^{N-1} I^k_{nmrs} \begin{Bmatrix} (U_k + U_k^*) \\ (V_k + V_k^*) \\ (-V_k + V_k^*)/i \\ (U_k - U_k^*)/i \end{Bmatrix} \\ &= \frac{1}{2\sigma} \sum_{k=0}^{N-1} I^k_{nmrs} \begin{Bmatrix} \mathrm{Re}(U_k) \\ \mathrm{Re}(V_k) \\ \mathrm{Im}(-V_k) \\ \mathrm{Im}(U_k) \end{Bmatrix} \end{aligned} \tag{5.20}$$

where Re(·) indicates the real part of a complex number and Im(·) the imaginery part. Eq. (5.20) is similar to equation (6.11) of Mainville (1987, p. 64), except that the index function $w_{k\ell}$ of Mainville's formula has a different definition than that in (5.1) and he did not use the complex expressions for his results. The form in (5.20) has the advantage that it is compact and easy to follow when programming.

Let us interpret (5.20) and summarize the computational formulae. First of all, not all A^{rs}_{nm}, B^{rs}_{nm}, C^{rs}_{nm}, and D^{rs}_{nm} values are the desired inner products, only those contained in the following matrix are needed:

$$G = \begin{bmatrix} (\overline{R}_{00}, \overline{R}_{00}) & & & \\ (\overline{R}_{10}, \overline{R}_{00}) & (\overline{R}_{10}, \overline{R}_{10}) & & \\ (\overline{R}_{11}, \overline{R}_{00}) & (\overline{R}_{11}, \overline{R}_{10}) & (\overline{R}_{11}, \overline{R}_{11}) & \\ (\overline{S}_{11}, \overline{R}_{00}) & \cdots\cdots & \cdots\cdots & (\overline{S}_{11}, \overline{S}_{11}) \\ & \vdots & & \\ (\overline{S}_{\mu\mu}, \overline{R}_{00}) & \cdots\cdots & \cdots\cdots & (\overline{S}_{\mu\mu}, \overline{S}_{\mu\mu}) \end{bmatrix} \quad \text{symmetry} \tag{5.21}$$

where G is the Gram matrix (see Chapter 3), and μ is the maximum degree of spherical harmonics used. The values of B^{rs}_{nm}, C^{rs}_{nm}, D^{rs}_{nm} will be zero whenever the argument of sine function is zero or $\overline{S}_{n0}$ is present, and these values will not enter the G matrix. However, due to the use of FFT method for computing $F_k(m, \pm s)$ and convenient programming, all the A^{rs}_{nm}, B^{rs}_{nm}, C^{rs}_{nm}, and D^{rs}_{nm} values are first computed using (5.20), then a selection routine

is designed to obtain the desired elements in the G matrix. The needed nm, rs pairs fall into the following FORTRAN loops:

```
      do 1 n = 0, μ
      do 1 r = 0, n
      do 1 m = 0, n
      rr = r                                          (5.22)
      if (r. eq. n) rr = m
      do 1 s = 0, rr
      : (nm, rs pairs)
    1 continue
```

The second issue is the use of the FFT for computing $F_k(m, \pm s)$. The form in (5.14) is already an FFT form which is defined in some routines such as IMSL's FFTCF, except that the complex exponential function has a different sign convention. If the maximum degree of the spherical harmonics used is μ, then from (5.14) we know that the lowest frequency is $-\mu$ for $F_k(m, s)$ and the highest frequency is 2μ for $F_k(m, -s)$. Thus, in order to compute U_k and V_k in (5.19), we must have the values $F_k(m, \pm s)$ ready at frequencies $-\mu$ to 2μ. To get these values, we first perform an FFT for $w_{k\ell}$ to get $F_k(\ell)$ at $\ell = 0, \ldots, 2N - 1$. Then the desired $F_k(m, \pm s)$ values are obtained from $F_k(\ell)$ by using $\ell = m - s$ or $\ell = m + s$. For negative indices, we can obtain the desired values using the relationship $F_k(-\ell) = F_k(2N - \ell)$. One can consult, e.g., Press et al. (1989, p. 397), to understand better the above statement.

The final, yet the most important, computation is I^k_{nmrs}, which is the integration of the product of two associated Legendre functions in the sub-interval $\theta_k \le \theta \le \theta_{k+1}$. Two formulae for I^k_{nmrs} will be presented in the following sections. Further, due to the fact that (see equation (4.1) for such a reasoning)

$$T(t) = \bar{P}^m_n(t)\bar{P}^s_r(t) = (\cos\theta)^{n-m+r-s}(\sin\theta)^{m+s}\sum_k e^{nmrs}_k(\cos\theta)^{-2k} \tag{5.23}$$

we have

$$T(-t) = T(\cos(\pi - \theta)) = (-1)^{n-m+r-s}T(t) \tag{5.24}$$

for $\sin\theta$ and $(\cos\theta)^{-2k}$ are always nonnegative over $0 \le \theta \le \pi$. Therefore, I^k_{nmrs} are required for only one hemisphere. The values on the other hemisphere can be obtained by (5.24). The computation of I^k_{nmrs} will be carried out in a separate program and the values I^k_{nmrs} will be input to the program for computing the inner products.

The following computational procedure summarizes the above development:

1. Perform the FFT for w_{kl} to obtain $F_k(m, \pm s)$ at the northern-most latitude belt and the southern-most latitude belt.

2. Read I^k_{nmrs} for all possible n, m, r, and s at the northern latitude belt.

3. Within the loops shown in (5.22), accumulate A^{rs}_{nm}, B^{rs}_{nm}, C^{rs}_{nm}, and D^{rs}_{nm} values by first forming U_k and V_k values and then forming the products between I^k_{nmrs} and $Re(U_k)$, $Re(V_k)$, etc., at these two latitude belts.

4. Repeat 1 to 3 at the next northern and southern latitude belts, until all latitude belts are exhausted.

5. Select the desired inner products from the final A^{rs}_{nm}, B^{rs}_{nm}, C^{rs}_{nm} and D^{rs}_{nm} values.

5.1.2 Recursive Formulae for Integrating Products of Two Associated Legendre Functions

Now we present a method for computing I^k_{nmrs} in (5.11). This method is based on the recursive relationships between $P^m_n(t)$ and itself, and between $P^m_n(t)$ and its derivative $\frac{d}{dt} P^m_n(t)$. Detailed derivations for the desired integration I^k_{nmrs} have been given by Mainville (1987, Chapter 7), thus no re-derivation is attempted here. We now summarize his results and list the fomulae that are needed in this study. Mainville's notations will be used in the following discussion and the typographical errors (the author believes) will be corrected based on the author's verification of his formulae. Also, the index "k" will be dropped in the following formulae.

Mainville used the following formulae for his derivations:

$$(1 - t^2)\frac{dP^m_n(t)}{dt} = (n + 1)\, tP^m_n(t) - (n - m + 1)\, P^m_{n+1} \tag{5.25}$$

$$= -nt\, P^m_n(t) + (n + m)\, P^m_{n-1}(t) \tag{5.26}$$

$$(2n + 1)\, t\, P^m_n(t) = (n - m + 1)\, P^m_{n+1}(t) + (n + m)\, P^m_{n-1}(t) \tag{5.27}$$

$$\bar{P}^m_n(t) = H_{nm} P^m_n(t) \tag{5.28}$$

$$\bar{P}^n_n(t) = \frac{(2n)!}{n!\, 2^n}(1 - t^2)^{n/2} = \frac{(2n)!}{n!\, 2^n}(\sin\theta)^n \tag{5.29}$$

where

$$H_{nm} = \left[\frac{(2 - \delta(m))(2n + 1)(n - m)!}{(n + m)!}\right]^{1/2} \tag{5.30}$$

Equation (5.28) shows the relations between the fully normalized Legendre function and the Legendre function. The final result should be in the "fully normalized" form as requested in I_{nmrs}, but the derivation can be carried out using the relationships for the (unnormalized) Legendre function presented in (5.25) to (5.27).

The idea to get the recursive relationships for I_{nmrs} is to multiply both sides of (5.25), (5.26), and (5.27) by another function $P^s_r(t)$ and then perform integrations over the interval $\theta_k \le \theta \le \theta_{k+1}$ or $t_{k+1} \le t \le t_k$, depending on the variable used. A necessary technique is the integration by parts. Let $t_{k+1} = t_s$, $t_k = t_n$, where s denotes "south" and n denotes "north" (not to be confused by the degree n in $\bar{P}^m_n$), Mainville obtained the formulae (see equations (7.24), (7.25) and (7.33) in ibid.)

$$I_{nmrs} = \frac{a(n, m)}{(n + r + 1)}\left[\frac{(n - r - 2)}{a(n - 1, m)} I_{n-2, mrs} + \frac{(2r + 1)}{a(r, s)} I_{n-1, m, r-1, s} - (1 - t^2)\overline{P}^m_{n-1}(t)\,\overline{P}^s_r(t)\Big|_{t_a}^{t_n}\right], \quad n \neq m \tag{5.31}$$

$$I_{nnrs} = \frac{a(r, s)}{(n + r + 1)}\left[\frac{(r - n - 2)}{a(r - 1, s)} I_{nn, r-2, s} - (1 - t^2)\overline{P}^n_n(t)\,\overline{P}^s_{r-1}(t)\Big|_{t_a}^{t_n}\right], \quad r \neq s \tag{5.32}$$

$$I_{nnrr} = \frac{1}{(n + r + 1)}\left[(n + r)\, b(n)\, b(n - 1)\, I_{n-2, n-2, rr} + t\,\overline{P}^n_n(t)\,\overline{P}^r_r(t)\Big|_{t_a}^{t_n}\right], \quad n \neq 0, 1 \tag{5.33}$$

where $a(\cdot,\cdot)$ and $b(\cdot)$ are two functions defined as

$$a(n, m) = \left[\frac{(2n + 1)(2n - 1)}{(n + m)(n - m)}\right]^{1/2}, \quad n \neq m \tag{5.34}$$

$$b(n) = \left(\frac{2n + 1}{2n}\right)^{1/2}, \quad n > 1;\ b(1) = \sqrt{3} \tag{5.35}$$

Similar to the integration of one associated Legendre function that has been studied by Paul (1978) and Gerstl (1980), the recursive formula in (5.33) for computing I_{nnrr} becomes unstable in the polar region, as verified by Mainville. One way to resolve this problem is to use a backward formula instead of the forward formula in (5.33). Such a backward formula can be obtained by merely re-arranging (5.33) and appropriately changing the indices:

$$I_{nnrr} = \frac{1}{(n + r + 2)\, b(n + 2)\, b(n + 1)}\left[(n + r + 3)\, I_{n+2, n+2, rr} - t\,\overline{P}^{n+2}_{n+2}\,\overline{P}^r_r(t)\Big|_{t_a}^{t_n}\right] \tag{5.36}$$

The use of (5.36) would require the starting values of $I_{\mu\mu\mu\mu}$, $I_{\mu-1,\mu-1,\mu-1,\mu-1}$, and $I_{\mu-1,\mu-1,\mu\mu}$, where μ is the maximum degree. To compute these starting values, we use the formulae in (5.29) and the normalizing factor H_{nm} to obtain (see also equation (7.36) in Mainville (1987))

$$I_{nnrr} = \prod_{\ell=1}^{n} b(\ell) \prod_{k=0}^{r} b(k) \int_{t_a}^{t_n} (1 - t^2)^{n+r/2} dt \tag{5.37}$$

However, direct integration in (5.37) cannot yield a closed form for I_{nnrr}. To get the value of I_{nnrr}, one way is to perform a series expansion of the kernel of the integral. Observing that θ or $y = \sin\theta = (1 - t^2)^{1/2}$ is small in the polar region, we may change the variable inside the integral from t to y. Then $(1 - t^2)^{n+r/2} = y^{n+r}$, $dt = -y(1 - y^2)^{-1/2}dy$. Now we may expand $(1 - y^2)^{-1/2}$ into a binomial series such as the one in Rapp (1989a, Vol. I, p. 7). Using such an expansion, we may get I_{nnrr} through term-by-term integration with the result (see also equation (7.37) in Mainville (ibid.)):

$$I_{nnrr} = -\prod_{\ell=1}^{n} b(\ell) \prod_{k=0}^{r} b(k) y^{n+r+2}\left(\frac{1}{n+r+2} + \frac{y^2}{2(n+r+4)} + \frac{1 \cdot 3}{2 \cdot 4}\,\frac{y^4}{(n+r+6)} + \cdots\right)\Bigg|_{y_a}^{y_n} \tag{5.38}$$

where $y_n = (1 - t_n^2)^{1/2}$, $y_s = (1 - t_s^2)^{1/2}$. The number of terms in (5.38) for a sufficient accuracy can be determined by Gerstl's formula (Gerstl, 1980, p. 193 and Mainville, 1980, p. 71). As we can expect, a smaller θ will yield a faster convergence hence a smaller number of terms in (5.38).

Other important issues in computing I_{nmrs} are:

(1) When using the recursive formulae, it is required to have the values of the Legendre functions at $t = t_n$ and $t = t_s$. These values can be obtained using the recursive formula in (5.27). A fully normalized version of this recursive formula is:

$$\bar{P}_n^m(t) = a(n,m)\left[t\,\bar{P}_{n-1}^m(t) - \frac{1}{a(n-1,m)}\,\bar{P}_{n-2}^m(t) \right], \quad m \neq n \tag{5.39}$$

Using this formula, we are basically computing $\bar{P}_n^m(t)$ in a degreewise manner. The first step is to compute all the sectorial terms by another formula:

$$\bar{P}_n^n(t) = b(n)(1 - t^2)^{1/2}\,\bar{P}_{n-1}^{n-1}(t) \tag{5.40}$$

which can be easily derived from (5.29). Then, for a fixed order m, we need to have another value $\bar{P}_{m+1}^m(t)$, to start the recurrence. $\bar{P}_{m+1}^m(t)$ can be obtained by just setting $n=m+1$ and letting $\bar{P}_{m-1}^m(t) = 0$ in (5.39). Thus

$$\bar{P}_{m+1}^m(t) = a(m+1,m)\,t\,\bar{P}_m^m(t) \tag{5.41}$$

Therefore, the only required starting value in $\bar{P}_0(t) = 1$, for all $\bar{P}_n^m(t)$.

(2) The required I_{nmrs} values fall into the loops shown in (5.22).

(3) Due to the relationship shown in (5.24), I_{nmrs} values are required only on one hemisphere.

We will postpone the discussion of the numerical results from the recursive formulae until we propose another method of computing I_{nmrs} in the next section.

5.1.3 Product-Sum Formulae for Integrating Products of Two Associated Legendre Functions

We now present the second method for integrating products of two associated Legendre functions. It is known that the associated Legendre functions, $P_n^m(t)$, as shown in (4.1), is basically the product of $(1 - t^2)^{m/2}$ and a polynomial of degree (n-m). Thus the product of two associated Legendre functions $P_n^m(t)$ and $P_r^s(t)$ will be the product of $(1 - t^2)^{(m+s)/2}$ and a polynomial of degree (n + r - m - s). Such a product can be regarded as a summation of several associated Legendre functions of order (m + s), namely,

$$P_n^m(t)\,P_r^s(t) = \sum_{\ell=0}^{[v/2]} D_{n+r-2\ell} P_{n+r-2\ell}^{m+s}(t) \tag{5.42}$$

where

$$v = n + r - m - s \tag{5.43}$$

and $D_{n+r-2\ell}$ are the combination constants. For example,

$$P_2^1P_1^1 = \frac{1}{5}P_3^2,\ P_2^2P_1^1 = \frac{1}{5}P_3^3,\ P_2^2P_2^1 = \frac{35}{3}P_4^3,\ P_3^1P_1^1 = \frac{1}{7}P_4^2 - \frac{1}{7}P_2^2,\ P_3P_1 = \frac{4}{7}P_4 + \frac{3}{7}P_2$$

Thus the integration of $P_n^m(t)P_r^s(t)$ is equivalent to the sum of integrations of $P_{n+r-2\ell}^{m+s}(t)$ multiplied by constants $D_{n+r-2\ell}$, $\ell = 0, \ldots, [v/2]$. The algorithm and software have been well-developed for integrating one associated Legendre function (for example, Paul (1978)). Hence, we will be able to find the integration of $P_n^m(t)P_r^s(t)$ if the constants $D_{n+r-2\ell}$ are known.

Now let us design a systematic approach of finding the needed constants. Firstly, we express $P_n^m(t)P_r^s(t)$ in terms of t based on the formula given in (4.1):

$$P_n^m(t)P_r^s(t) = 2^{-(n+r)}(1-t^2)^{m+s/2}\sum_{\ell=0}^{[v/2]} C_{v-2\ell}t^{v-2\ell} \tag{5.44}$$

where

$$C_{v-2\ell} = \sum_{k=0}^{\ell} a_{2k}^{nm} b_{2(\ell-k)}^{rs} \tag{5.45}$$

with

$$\left.\begin{aligned} &a_{2k}^{2m} = 0,\ n-m-2k<0 \\ &b_{2k}^{rs} = 0,\ r-s-2k<0 \\ &a_{2k}^{nm} = (-1)^k \frac{(2n-2k)!}{k!(n-k)!\,(n-m-2k)!},\ n-m-2k \ge 0 \\ &b_{2k}^{rs} = (-1)^k \frac{(2r-2k)!}{k!(r-k)!\,(r-s-2k)!},\ r-s-2k \ge 0 \end{aligned}\right\} \tag{5.46}$$

Similarly, the right side of (5.42) is:

$$\sum_{\ell=0}^{[v/2]} D_{n+r-2\ell}\,P_{n+r-2\ell}^{m+s}(t) = 2^{-(n+r)}(1-t^2)^{m+s/2}\sum_{\ell=0}^{[v/2]} 2^{2\ell} D_{n+r-2\ell} \sum_{j=0}^{[v/2]-\ell} E_{v-2\ell}^{2j} t^{v-2\ell-2j} \tag{5.47}$$

where

$$E^{2j}_{\nu-2\ell} = (-1)^j \frac{[2(n+r-2\ell)-2j]!}{j![(n+r-2\ell-j)]![\nu-2\ell-2j]!} \tag{5.48}$$

Comparing the coefficients of the power t in (5.44) and (5.47), we get

$$C_{\nu-2\ell} = \sum_{k=0}^{\ell} 4^k D_{n+r-2k} E^{2(\ell-k)}_{\nu-2k}, \ \ell = 0, \cdots, [\nu/2] \tag{5.49}$$

or

$$D_{n+r-2\ell} = \frac{1}{4^\ell E^0_{\nu-2\ell}} \left[C_{\nu-2\ell} - \sum_{k=0}^{\ell-1} 4^k D_{n+r-2k} E^{2(\ell-k)}_{\nu-2k} \right] \tag{5.50}$$

which is a recursive relationship for the desired constants. It is cumbersome to apply (5.49) directly. To facilitate the computation, we perform the following manipulations. First of all, we can find the recursive relationships

$$\begin{aligned} a^{nm}_{2(k+1)} &= \alpha_{2k}\, a^{nm}_{2k} \\ b^{rs}_{2(k+1)} &= \beta_{2k}\, b^{rs}_{2k} \end{aligned} \tag{5.51}$$

where

$$\alpha_{2k} = \frac{-(n-m-2k)(n-m-2k-1)}{2(2k+1)(2n-2k-1)}$$

$$\beta_{2k} = \frac{-(r-s-2k)(r-s-2k-1)}{2(2k+1)(2r-2k-1)} \tag{5.52}$$

Thus

$$C_{\nu-2\ell} = \frac{(2n)!(2r)!}{n!(n-m)!r!(r-s)!} \overline{C}_{\nu-2\ell} \tag{5.53}$$

where

$$\overline{C}_{\nu-2\ell} = \sum_{k=0}^{\ell} \bar{a}_k \bar{b}_{\ell-k} \tag{5.54}$$

with

$$\bar{a}_k = \begin{cases} 1, & k = 0 \\ 0, & n-m-2k < 0 \\ \displaystyle\prod_{j=0}^{k-1} \alpha_{2j}, & \text{otherwise} \end{cases} \tag{5.55}$$

$$\bar{b}_k = \begin{cases} 1\ ,\ k = 0 \\ 0\ ,\ r - s - 2k < 0 \\ \prod_{j=0}^{k-1} \beta_{2j}\ ,\ \text{otherwise} \end{cases} \tag{5.56}$$

Similarly,

$$E_{\nu-2\ell}^{2j} = 2^{-2\ell-j}\frac{[2(n+r)]!}{(n+r)!\ \nu!} f_\ell^j h_\ell \tag{5.57}$$

where

$$f_\ell^j = \begin{cases} 1\ ,\ j = 0 \\ \prod_{p=0}^{j-1} e_{2\ell}^{2p}\ ,\ j > 0 \end{cases} \tag{5.58}$$

$$h_\ell = \begin{cases} 1\ ,\ \ell = 0 \\ \prod_{q=0}^{\ell-1} d_{2q}\ ,\ \ell > 0 \end{cases} \tag{5.59}$$

and

$$e_{2\ell}^{2p} = \frac{(\nu - 2\ell - 2p)(\nu - 2\ell - 2p - 1)}{(p+1)[2(n+r-2\ell) - 2p - 1]} \tag{5.60}$$

$$d_{2q} = \frac{(\nu - 2q)(\nu - 2q - 1)}{[2(n+r-2q)-1][2(n+r-2q)-3]} \tag{5.61}$$

Further, let the fully normalized form of the product-sum formula be

$$\bar{P}_n^m(t)\bar{P}_r^s(t) = \sum_{\ell=0}^{[\nu/2]} \bar{D}_{n+r-2\ell}\bar{P}_{n+r-2\ell}^{m+s}(t) \tag{5.62}$$

then

$$D_{n+r-2\ell} = \frac{H_{n+r-2\ell,m+s}}{H_{nm}\ H_{rs}}\bar{D}_{n+r-2\ell} \tag{5.63}$$

where H_{nm} is defined in (5.3). Further, we may express $H_{n+r-2\ell,m+s}$ in a recursive form as

$$H_{n+r-2\ell,\ m+s} = g_\ell\ H_{n+r,\ m+s} \tag{5.64}$$

where

$$g_\ell = \begin{cases} 1\ , \ \ell = 0 \\ \prod_{j=0}^{\ell-1} X_{2j}\ , \ \ell > 0 \end{cases} \tag{5.65}$$

with

$$X_{2j} = \sqrt{\frac{[2(n+r-2j)-3](n+r+m+s-2j)(n+r+m+s-2j-1)}{[2(n+r-2j)+1](\nu-2\ell)(\nu-2\ell-1)}} \tag{5.66}$$

With the above considerations, (5.49) becomes

$$\overline{G}\,\overline{C}_{\nu-2\ell} = \sum_{k=0}^{\ell} \left(2^{k-\ell}\,\overline{D}_{n+r-2k}\,g_k f_k^{\ell-k} h_k\right) \tag{5.67}$$

where

$$\overline{G} = \sqrt{\frac{2(2n+1)(2r+1)}{2(n+r)+1}}\sqrt{\binom{n-m+r-s}{n-m}\binom{n+m+r+s}{n+m}}\frac{\binom{n+r}{n}}{\binom{2n+2r}{2n}}\frac{1}{\sqrt{1+\delta_{(m\,\cdot\,s)}}} \tag{5.68}$$

From (5.67), we finally obtain the desired formula for $\overline{D}_{n+r-2\ell}$:

$$\overline{D}_{n+r-2\ell} = \frac{1}{g_\ell h_\ell}\left(\overline{G}\,\overline{C}_{\nu-2\ell} - \sum_{k=0}^{\ell-1} 2^{k-\ell}\overline{D}_{n+r-2k} g_k f_\ell^{\ell-k} h_k\right),\ \ell = 0,\ 1, \cdots, [\nu/2] \tag{5.69}$$

with the starting value $\overline{D}_{n+r} = \overline{G}$.

To compute $\overline{D}_{n+r-2\ell}$ based on the above development, the first step is to calculate $\overline{C}_{\nu-2\ell}$, g_k, $f_k^{\ell-k}$ and h_k, which can be easily programmed using their corresponding recursive formulae, for instance

$$h_k = d_{2(k-1)}\ h_{k-1},\ g_k = X_{2(k-1)}\ g_{k-1} \tag{5.70}$$

These values are stored in separate arrays and then are used in (5.69) later. The maximum length of these arrays will be just $\mu + 1$, where μ is the maximum degree of the spherical harmonics. It took 50 CPU seconds to compute all the needed constants $\overline{D}_{n+r-2\ell}$ for $\mu = 36$ on the CRAY Y-MP/864 machine. The total numbers of combination constants needed for all possible nm, rs pairs in the loops given in (5.22) for various maximum degrees are shown in Table 5.1.

Table 5.1 Total Numbers of Combination Constants Required for Various Maximum Degrees of Spherical Harmonics

maximum degree	number	maximum degree	number
8	3545	24	463593
12	19901	28	956285
16	71697	32	1799841
20	198781	36	3155197

From (5.42), it is evident that if m = s = 0, then the number of combination constants is the maximum among all possible products with fixed n and r values. It is also predictable that instability could occur when we are computing the combination constants for a product of two Legendre polynomials (i.e., m = s = 0) due to the accumulated error arising from the maximum number of $\overline{D}_{n+r-2\ell}$ used. In the experiments conducted on the CRAY-YMP/864 machine, the computed $\overline{D}_{n+r-2\ell}$ becomes unreasonably large after degree 17. Similar phenomenon occurred on the IBM 3081 machine when performing the same experiments. The problem was then overcome by using the double-precision mode on the CRAY-YMP/864. The tradeoff is the increase of computer time since vectorization in a supercomputer cannot be active in a double-precision mode and longer word-length is used in the double-precision calculations.

Having these constants, the next step is to obtain the I_{nmrs} values from the formula

$$I_{nmrs} = \int_{t_n} \overline{P}_r^m(t)\overline{P}_r^s(t)dt = \sum_{\ell=0}^{[v/2]} \overline{D}_{n+r-2\ell} \int_{t_n} \overline{P}_{n+r-2\ell}^{m+s} dt \tag{5.71}$$

To verify the formulae derived in this section, we may check the I_{nmrs} values in (5.71) against the values from Mainville's formulae. In fact, the algorithm developed in this section for computing I_{nmrs} is totally different from Mainville's algorithm which has been discussed in the previous section. With the help of Professor Rapp, the program for calculating I_{nmrs} based on Mainville's algorithm was made available to the author. Also, the program for computing the integrations of one Legendre function that are needed in (5.71) is available in Professor Rapp's program library (based on the Paul (1978) procedure).

The comparison of I_{nmrs} values up to n = m = r = s = 36 computed from the product-sum formulae (the formulae developed in this section) and the recursive formulae (the formulae developed by Mainville (1987) and described in the previous section) is shown in Table 5.2. In Table 5.2, $\cos\theta_n = t_n$ and we have used $\Delta\theta = 1°$. It can be seen from Table 5.2 that the discrepancy increases as θ_n approaches the polar region. The maximum difference occurred at n = r = 36 and m = s = 0 for almost all the latitude belts. When $\theta_n > 45°$, the maximum difference has dropped to a value less than 10^{-10}. The large discrepancies in the polar region may be due to the instability of the integration of $\overline{P}_n^m(t)$ in such a region, as pointed out by Paul (1978). The combination constants were appropriately obtained by the formulae developed here since good agreement of I_{nmrs} values from the two methods was found at most of the latitude belts (if the constants are wrong, no match will happen at all!).

Table 5.2 Comparison of I_{nmrs} Values from the Product-Sum Formulae and the Recursive Formulae at the nm, rs Pairs Where Maximum Differences Occur

θ_n	n	m	r	s	product-sum	recursive	max. difference
0	36	0	36	0	$1.003995 \cdot 10^{-2}$	$1.00461 \cdot 10^{-2}$	$6.111658 \cdot 10^{-6}$
9	36	0	36	0	$6.033331 \cdot 10^{-3}$	$6.093284 \cdot 10^{-3}$	$5.995270 \cdot 10^{-5}$
18	36	0	36	0	$7.198512 \cdot 10^{-4}$	$7.377833 \cdot 10^{-4}$	$1.793202 \cdot 10^{-5}$
27	36	0	36	0	$6.438187 \cdot 10^{-3}$	$6.439329 \cdot 10^{-3}$	$1.142596 \cdot 10^{-6}$
36	36	0	36	0	$1.720875 \cdot 10^{-2}$	$1.720877 \cdot 10^{-2}$	$1.644023 \cdot 10^{-8}$
45	36	0	36	0	$2.135710 \cdot 10^{-2}$	$2.135710 \cdot 10^{-2}$	$4.905332 \cdot 10^{-11}$
54	36	15	36	5	$1.742747 \cdot 10^{-2}$	$1.742747 \cdot 10^{-2}$	$1.976197 \cdot 10^{-13}$
63	36	21	36	1	$-3.794209 \cdot 10^{-2}$	$-3.794209 \cdot 10^{-2}$	$8.271161 \cdot 10^{-13}$
72	36	21	36	4	$-2.912995 \cdot 10^{-4}$	$-2.9122995 \cdot 10^{-4}$	$1.614411 \cdot 10^{-12}$
81	36	20	36	1	$-7.644171 \cdot 10^{-3}$	$-7.644171 \cdot 10^{-3}$	$3.863600 \cdot 10^{-12}$
89	36	25	36	1	$4.822763 \cdot 10^{-3}$	$4.822763 \cdot 10^{-3}$	$3.271521 \cdot 10^{-12}$

To further compare the two methods, we can perform the inner products in (5.20) using these two kinds of I_{nmrs} values and assume that $w_{k\ell} = 1$ for all k and ℓ, namely, all the blocks are oceans. Due to the orthonormality of the fully normalized spherical harmonics, we should expect all the inner products to be either 1 or 0 in such a case. The deviations of the inner products from 1 or 0 are mainly caused by the I_{nmrs} values used. As a result, the use of I_{nmrs} from the recursive formulae creates a maximum deviation of 10^{-13}, while the use I_{nmrs} from the product-sum formulae yields a slightly larger deviation. Such a experiment shows that the recursive formulae have a better accuracy. Further, the use of I_{nmrs} from the product-sum formulae will worsen the mutual dependence problem of the spherical harmonics over the oceans (a topic that will be discussed in later sections).

Based on the aforementioned comparisons, we decided to use the recursive formulae for the computation of the I_{nmrs} values that are needed in calculating the inner products of spherical harmonics over the oceans.

To conclude this section, we shall list some of the papers that have dealt with the product-sum formula similar to the one developed here. In Hobson (1965, p. 86), a product-sum formula has been derived for only the Legendre polynomials. Hobson's results is based on Legendre's differential equation (see (3.74)), thus his idea is totally different from the one given here. One can also find a product-sum formula for the Legendre polynomials in Banerji (1920, p. 179). As the last example, we can find a product-sum formula in Giacaglia (1980, p. 3). Giacaglia's formula is derived using the orthogonality of the Legendre functions $P_n^m(t)$ of the same order and does not have too much similarity with the one given in this section.

5.2 Construction of Orthonormal Functions over the Oceans

Up to this point, we have discussed the methods of computing the inner products, defined as integrals of surface spherical harmonics over the oceans. As pointed out in Section 3.2.2, the inner products are the necessary quantities for finding the combination coefficients c_{ij} in (3.2). We also talked about the definitions of the oceans in Section 3.1. It is now necessary to present the detailed process that is related to the use of surface

spherical harmonics in the construction of the orthonormal functions. For brevity, we shall call the orthonormal functions as ON functions from now on.

The first issue to be addressed is the definition of the oceans. Apparently, the "oceans" should be unique in the real world. But, as the elevation data or shoreline data are not accurate enough, we simply cannot get the real oceans. Moreover, if we restrict our study of some phenomena only over a portion of the real oceans, then the part of oceans outside the interested area really does not concern us. One example is given in Haines (1985b). Haines has studied Magsat vertical field anomalies available above 40°N. He then treated the magnetic field as a local phenomenon and expanded these available data into spherical cap harmonics (note: his definition of spherical cap harmonics is different from the one given in Chapter 3). The earth's magnetic field, like the gravity field, is a global phenomenon and should not be expanded into local functions such as the spherical cap harmonics. This is true from the physical point of view. Yet it is an acceptable idea that these magnetic anomalies are just some signals without any pre-assumed physical properties and they can be "fitted" by any local functions. Further, the reality is that we cannot get data outside the interested area so that any attempt to perform spectral analysis using global functions will end up with false results. Therefore, when Haines used the spherical cap harmonics to fit his data, he was doing a "local" spectral analysis with respect to a set of "local" harmonics. This is completely justifiable from the mathematical point of view if we introduce the concept of generalized Fourier series and generalized Fourier analysis. Based on this example, it is understood why we need to "choose" oceans: we cannot have the desired signal over the entire oceans and we just want to construct a set of local harmonics for data representation and spectral analysis.

Due to the properties of the ON functions constructed by the Gram-Schmidt process, such a "choice" of oceans is meaningful. These properties will be discussed in Section 5.4. Although the oceans defined in this way depends on the signal distribution, we should choose the one which fits the two most important purposes in this study, namely, the expansion of Levitus SST and simultaneous estimation of geoid and SST from satellite altimetry. Let us now look at the oceans implied by the NWL 80,000-point shoreline data and the edited TUG 1° x 1° mean elevation data (Wieser, 1987). In Figure 5.1, the shaded area implies $\overline{H} \leq 0$ where $\overline{H}$ is the mean elevation in a 1° x 1° block. In the original TUG 1° x 1° mean elevation data, some in-land blocks also receive negative $\overline{H}$ values. This happens in the western part of China and in Australia and some other in-land basins. It was also found that the mean elevation data do not quite match the 80,000-point shoreline data in the Antarctic area, especially near the Antarctic peninsula. These in-land blocks and the blocks with $\overline{H} \leq 0$ inside the shoreline of the Antarctic continent are then removed. The oceans presented in Figure 5.1 are based on such edited elevation data.

In Figure 5.2, "the oceans" are defined to be the area where the modified SST of Levitus exist. This modified SST data set is described by Engelis (1987, p. 3) and consists of 30,922 estimated 1° x 1° mean values, including data in the Mediterranean Sea and the Black Sea. The oceans in Figure 5.2 have a minimum depth of 2250 meters, except in the Mediterranean Sea and the Black Sea. In the analysis performed in the next chapter, the oceans will be the one given in Figure 5.2.

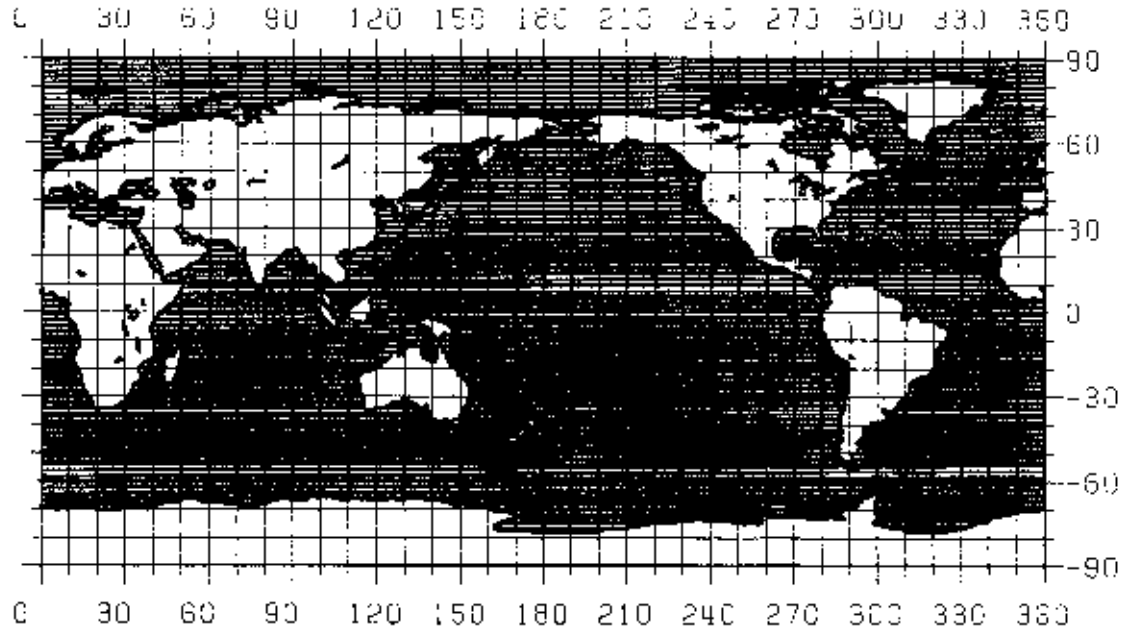


Figure 5.1 Oceans Implied by the Edited TUG87 1° x 1° Elevation Data (see text).

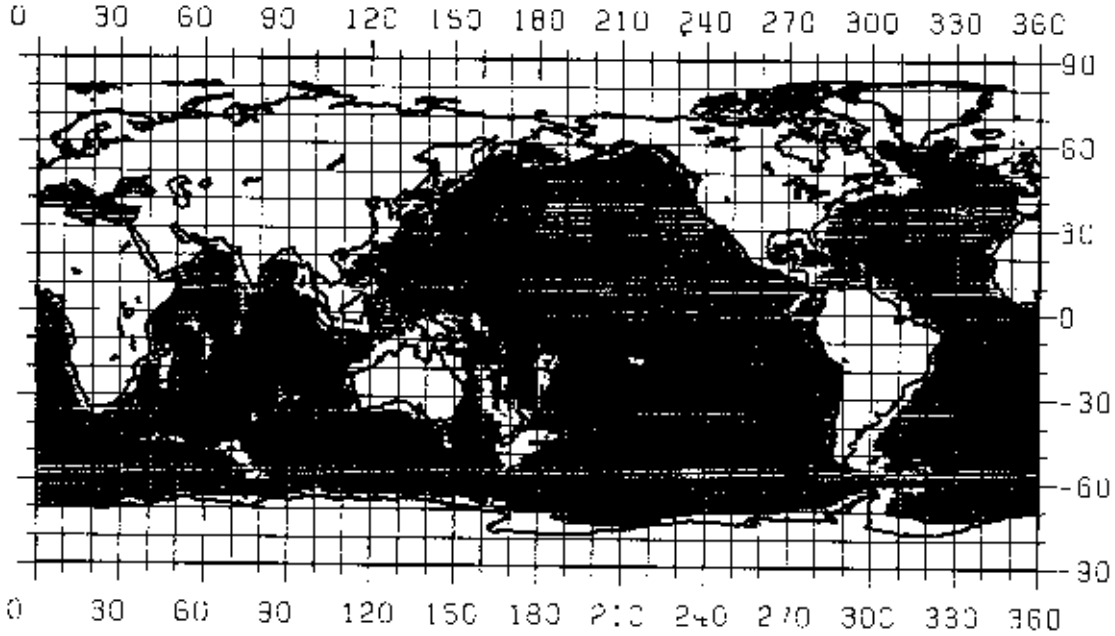


Figure 5.2 Oceans as the Area Where the Modified SST of Levitus Exists (Set 3 of Engelis, 1987b), $\overline{H} \leq -2250$ Meters.

The next issue is the linear dependence of spherical harmonics over the oceans. The method to check the dependence of functions over a domain is the Gram determinant |G| where G is defined in (3.3). If |G| is zero, then not all of the given functions are linearly independent. Now let the systems $\{L_j\}$ consist of all possible surface spherical harmonics following the sequence

$$\{L_j\} = \{\overline{R}_{00}, \overline{R}_{10}, \overline{R}_{11}, \overline{S}_{11}, \overline{R}_{20}, \overline{R}_{21}, \overline{S}_{21}, \overline{R}_{22}, \overline{S}_{22}, \overline{R}_{30}, \overline{R}_{31}, \overline{S}_{31}, \overline{R}_{32}, \ldots\} \qquad (5.72)$$

where $\overline{R}_{nm} = \overline{P}_n^m(\cos\theta)\cos m\lambda$ and $\overline{S}_{nm} = \overline{P}_n^m(\cos\theta)\sin m\lambda$, are the fully normalized spherical harmonics. The Gram matrix of $\{L_j\}$ is given in (5.21). Note that we have not defined the oceans in (5.21) and the index function $w_{k\ell}$ is basically unknown. Now we would like to find the first harmonic in $\{L_j\}$ that has dependence with the previous harmonics over a certain domain. By doing this, we can determine a certain degree and order of spherical harmonic up to which no dependence of functions occurs over a specific domain. Note that we must strictly follow the sequence of the spherical harmonics shown in (5.72) in doing this analysis.

First of all, let us define the geographical boundaries of some oceanic areas (ϕ = latitude, λ = longitude):

Area 1: The Caspian Sea: $35° \le \phi \le 50°$, $45° \le \lambda \le 57°$

Area 2: The Red Sea: $12° \le \phi \le 30°$, $31° \le \lambda \le 43°$

Area 3: The Persian Gulf: $22° \le \phi \le 31°$, $46° \le \lambda \le 56°$

Area 4: The Baltic Sea: $47° \le \phi \le 60°$, $5° \le \lambda \le 30°$ and $60° \le \phi \le 67°$, $15° \le \lambda \le 30°$

Area 5: The Hudson Bay and the Hudson Straight: $50° \le \phi \le 72°$, $263° \le \lambda \le 295°$

Then we define the following domains for the dependence analysis:

Domain 1: The oceans given in Figure 5.1
Domain 2: The oceans given in Figure 5.2
Domain 3: The oceans given in Figure 5.1, excluding the 1° x 1° blocks with $\overline{H} < 0$ in areas 1, 2, 3, 4, 5
Domain 4: The oceans given in Figure 5.1, excluding the 1° x 1° blocks with $\overline{H} < 0$ in areas 1, 2, 3, 4, 5 and the area where $\phi > 72°$.
Domain 5: The ocean given in Figure 5.1, excluding the 1° x 1° blocks with $\overline{H} < 0$ in areas 1, 2, 3, 4, 5 and the area where $|\phi| > 72°$.

Five index functions $w_{k\ell}$ are created according to the definitions of the above five domains. Then the inner products using equation (5.20) are generated for the five domains up to a maximum degree of 36. For each domain and for a maximum degree of 36, it took 13 CPU seconds on the CRAY YMP/864 machine to generate the required inner products $(L_j, L_k)_\alpha$, where α = 1, 2, 3, 4, 5, indicates the domain the inner products to be formed. To find the ranks of the five Gram Matrices G_α, Linpack's routine SPPCO (Dongarra et al., 1979) is used. The maximum j value before which the elements L_j's are independent is equal to the rank of G_α (for a discussion of the independence of functions, see Section 3.2.1). The corresponding degree and order of the spherical harmonic can be calculated from this maximum j value. As a result, Table 5.3 summarizes the maximum j values and the corresponding spherical harmonics for domains 1, 2, 3, 4, and 5.

Table 5.3 Maximum j Values in System $\{L_j\}$ Before Which the Elements are Independent

Domain	j_{max}	spherical harmonic
1	>1369	$>\bar{S}_{36,36}$
2	646	$\bar{S}_{25,10}$
3	>1369	$>\bar{S}_{36,36}$
4	1321	$\bar{S}_{36,12}$
5	1321	$\bar{S}_{36,12}$

Therefore, for domain 2, we can only get a "complete" set of spherical harmonics up to a maximum degree of 24. By "complete" we mean that all the (2n+1) spherical harmonics for each degree n are present in the set of elements we choose. Apparently, at degree 25, we still can get some independent harmonics for Domain 2, according to Table 5.3. The result shown in Table 5.3 does not imply that we can only have a finite number of harmonics at a domain. To get an infinite number of harmonics, we can just exclude the one which first causes the dependence, and then introduce the next harmonic. If the dependence happens again, the corresponding harmonic is removed. Obviously, the process will lead to an "incomplete" set of spherical harmonics for a domain. Moreover, to find the harmonics that cause the dependence, we need to estimate the ranks as many times as the number of such harmonics. The process will be extremely computer-time consuming even if it is possible. For Domains 1 and 3, the spherical harmonics are independent at least up to n = m = 36. However, it is expected that the dependence will occur at some n and m beyond 36 at these two domains and the same process will be needed to get an infinite number of harmonics.

It will be shown in Chapter 6 that for the purpose of this study it is sufficient to use the spherical harmonics up to degree 24 to construct the ON functions that are needed. Thus the construction of the ON functions will basically just require this "complete" set of spherical harmonics (up to degree 24). Of course, it is without any problem if we use more harmonics. Such a choice is just for practical use and for convenience, especially in classifying the frequencies of the ON functions constructed in this way. Now, constructing the ON functions is equivalent to finding the combination coefficients c_{np} (see also Section 3.2.1) in the following equation:

$$X_n(\theta, \lambda) = \sum_{p=0}^{n} c_{np} L_p(\theta, \lambda), \quad n = 0, 1, \ldots \tag{5.73}$$

where $X_n(\theta,\lambda)$ are the desired ON functions. Writing the ON functions and the spherical harmonics with double indices, we have an alternative form of (5.73):

$$
\left.
\begin{aligned}
O_{nm}(\theta, \lambda) &= c_{kk}\bar{R}_{nm}(\theta, \lambda) + \sum_{p=0}^{k-1} c_{kp}L_p(\theta, \lambda),\ k = \begin{cases} n^2, & m = 0 \\ n^2 + 2m - 1, & m \neq 0 \end{cases} \\
Q_{nm}(\theta, \lambda) &= c_{kk}\bar{S}_{nm}(\theta, \lambda) + \sum_{p=0}^{k-1} c_{kp}L_p(\theta, \lambda),\ k = n^2 + 2m \\
Q_{nm}(\theta, \lambda) &= 0\ ,\ m = 0 \\
& n = 0, 1, \ldots ,\quad m = 0, \ldots , n
\end{aligned}
\right\} \qquad (5.74)
$$

In the above formulae, the index "p" has started from zero which will prove to be convenient in later discussions. The desired ON functions $O_{nm}(\theta,\lambda)$ and $Q_{nm}(\theta,\lambda)$ are defined by exact analogy with the spherical harmonics. Thus n is the degree and m the order of the ON functions. The definitions of degree variance, error degree variance, power spectrum, etc. for the spherical harmonics can then be applicable to the ON functions O_{nm}, Q_{nm}. It is also necessary to point out that the ON system $\{X_p\}$ or $\{O_{nm}, Q_{nm}\}$ defined in (5.73) or (5.74) is merely one of the systems that will be discussed in the following section. For the simultaneous estimation of geoid and SST, the system in (5.73) or (5.74) is in fact not the best choice. We will discuss this shortly.

We have discussed two methods of finding c_{np} in Section 3.2.2. The two methods are given in (3.18) and (3.27), respectively. The first method is based on a recursive algorithm; the second method can be characterized as a Cholesky decomposition. Both methods require the inner products that have been extensively discussed in this Chapter. Two programs have been designed separately for the computation of c_{np}. For a maximum degree of 24, the recursive algorithm consumed 3.480 CPU seconds, while the Cholesky decomposition took 1.130 CPU seconds. The computations were made on the CRAY-YMP/864 machine. In each computation, 195625 combination coefficients c_{np} were found. Note that the CPU time does not take into account the time for the inner products (see the previous section for the CPU time for the inner products).

The deviation of the two sets of combination coefficients increases as the harmonic degree increases. Table 5.4 lists the maximum difference between the two sets of combination coefficients when the computation is made up to a certain maximum degree.

Table 5.4 Maximum Difference Between Combination Coefficients From the Recursive Algorithm and Cholesky Decomposition up to a Maximum Degree

Max. degree	Max. difference
10	$3.654748 \cdot 10^{-10}$
15	$3.717914 \cdot 10^{-6}$
20	$3.623193 \cdot 10^{-2}$
24	77.977948

It is very surprising to see that the maximum difference can reach 77.977948 when a maximum degree of 24 is used. However, the magnitude of the coefficient corresponding to this maximum difference is on the order of 10^4, thus the discrepancy is about 0.8% of the magnitude. To compare further the two sets of combination coefficients, we may check

the orthonormality of the functions constructed by these two methods. The formula to be used is:

$$(X_p, X_q) = \left(\sum_{\ell=0}^{p} c_{p\ell} L_\ell, \sum_{j=0}^{q} c_{qj} L_j \right)$$

$$= \sum_{\ell=0}^{p} c_{p\ell} \sum_{j=0}^{q} c_{qj} (L_\ell, L_j) = \delta_{pq} \qquad (5.75)$$

where (L_ℓ, L_j) is the inner products of the spherical harmonics over the oceans. For a maximum degree of 24, the maximum p or q is 624. The deviation of (X_p, X_q) from 0 or 1 thus is a measure of the precision of the combination coefficients. In Table 5.5, several pairs of p, q have been chosen to check the orthonormality of the functions constructed by the two methods.

Table 5.5 Comparison of Orthonormality* of the ON Functions Constructed by the Recursive Algorithm and Cholesky Decomposition

p	q	δ_{pq} recursive	δ_{pq} Cholesky
0	1	$1.77635 \cdot 10^{-15}$	$1.77635 \cdot 10^{-15}$
4	9	$-2.44249 \cdot 10^{-15}$	$-2.12330 \cdot 10^{-15}$
5	5	1.000000000000000	1.000000000000000
14	22	$-2.99760 \cdot 10^{-15}$	$-5.08621 \cdot 10^{-15}$
16	18	$2.28983 \cdot 10^{-15}$	$1.55952 \cdot 10^{-15}$
23	23	1.000000000000000	1.000000000000000
499	524	$-9.15614 \cdot 10^{-8}$	$1.45972 \cdot 10^{-7}$
624	624	0.99999950128241	0.99999768868481

* $(X_p, X_q) = \delta_{pq}$, X_p are the ON functions on the oceans

In checking the orthonormality of the ON functions, a maximum deviation from δ_{pq} is found to be 1.25161×10^{-4} for the recursive algorithm, while the Cholesky decomposition creates a maximum deviation of 9.66554×10^{-5}. Generally speaking, both methods produce good orthonormality of the ON functions. The maximum difference of 77.977948 between the two sets of combination coefficients really has no substantial influence on the orthonormality of the constructed functions. Nevertheless, the large deviation indeed shows an instability in the computations. However, based on the computer time it is evident that the Cholesky decomposition is more efficient. In particular, the corresponding formula is easy to understand and is convenient for programming, therefore the Cholesky decomposition is the technique to be used in further computations.

To conclude this section, we list the first few orthonormal functions at Domain 2 (corresponding to the oceans in Figure 5.2) as follows (note: the area of Domain 2 is 7.420428, unitless):

$$O_{00} = 1$$

$$O_{10} = 0.338 + 1.160\ \bar{R}_{10}$$

$$O_{11} = 0.203 + 0.156\ \bar{R}_{10} + 0.949\ \bar{R}_{11}$$

$$Q_{11} = 0.179 + 0.146\ \bar{R}_{10} - 0.010\ \bar{R}_{11} + 1.011\ \bar{S}_{11}$$

$$O_{20} = 0.413 + 0.561\ \bar{R}_{10} - 0.041\ \bar{R}_{11} + 0.028\ \bar{S}_{11} + 1.316\ \bar{R}_{20}$$

$$O_{21} = 0.206 + 0.207\ \bar{R}_{10} + 0.143\ \bar{R}_{11} + 0.126\ \bar{S}_{11} + 0.064\ \bar{R}_{20} + 0.966\ \bar{R}_{21}$$

$$Q_{21} = 0.344 + 0.396\ \bar{R}_{10} + 0.151\ \bar{R}_{11} + 0.432\ \bar{S}_{11} + 0.162\ \bar{R}_{20} + 0.087\ \bar{R}_{21} + 1.183\ \bar{S}_{21}$$

5.3 Choices of Orthonormal Systems and Unitary Transformation

Our final goal of this study is the estimation of geoid undulations and SST from satellite altimetry while reducing satellite orbit error. The problem of the correlation between the estimated parameters needs to be considered and hence we must choose the functions which approximate SST in the estimation model more carefully. Before we perform the numerical experiments for such a solution, we must first theoretically investigate the possibility of closely approximating the SST and avoiding correlation when using a particular ON system. For such a purpose, we define 3 ON systems that result from 3 different kinds of combinations of the elements L_j in the sequence $\{L_j\} = \{L_0, L_1, L_2, \ldots\}$ (see (5.72)).

(1) System 1: $\{X_j\}$, or $\{O_{nm}, Q_{nm}\}$, see (5.73) or (5.74)

(2) System 2: $\{Y_j\}$, resulting from combining L_j, j = 1, 2, 3...

$$Y_n = \sum_{p=0}^{n} d_{np} L_{p+1},\ n = 0,\ 1,\ 2,\ \ldots$$

or

$$\left.\begin{aligned}
&\overline{O}_{nm} = d_{kk}\overline{R}_{nm} + \sum_{p=0}^{k-1} d_{kp} L_{p+1},\ k = \begin{cases} n^2 - 1, & m = 0 \\ n^2 + 2m - 2, & m \neq 0 \end{cases} \\
&\overline{Q}_{nm} = d_{kk}\overline{S}_{nm} + \sum_{p=0}^{k-1} d_{kp} L_{p+1},\ k = n^2 + 2m \\
&\overline{Q}_{nm} = 0,\ m = 0 \\
&n = 1,\ 2,\ \ldots,\ m = 0,\ 1,\ \ldots,\ n
\end{aligned}\right\} \quad (5.76)$$

(3) System 3: $\{Z_j\}$, resulting from combining L_j, $j = 4, 5, 6, \ldots$

$$Z_n = \sum_{p=0}^{n} e_{np} L_{p+4},\ n=0,1,2,\ldots$$

or

$$\left.\begin{aligned}
&\hat{O}_{nm} = e_{kk}\overline{R}_{nm} + \sum_{p=0}^{k-1} e_{kp} L_{p+4},\ k = \begin{cases} n^2 - 4, & m = 0 \\ n^2 + 2m - 5, & m \neq 0 \end{cases} \\
&\hat{Q}_{nm} = e_{kk}\overline{S}_{nm} + \sum_{p=0}^{k-1} e_{kp} L_{p+4},\ k = n^2 + 2m - 4 \\
&\hat{Q}_{nm} = 0,\ m = 0 \\
&n = 2,\ 3,\ \ldots,\ m = 0,\ 1,\ \ldots,\ n
\end{aligned}\right\} \quad (5.77)$$

In system 1, all the spherical harmonics are used in constructing the ON functions; in System 2, the first harmonic $\overline{R}_{00} = 1$ is not used; in System 3, the first 4 harmonics

$\bar{R}_{00}$, $\bar{R}_{10}$, $\bar{R}_{11}$, $\bar{S}_{11}$ are not used. We will prove by numerical experiments that System 2 and System 3 can approximate the Levitus SST as close as System 1 does, provided that a sufficiently high degree of the ON functions is used. By the above definitions, System 2 has no degree 0 and System 3 does not have degrees 0 and 1. System 3 has an important advantage that the correlation between its elements and the satellite orbital errors due to the initial state vector (see Chapter 7) is small, especially when a priori information is used. System 2 is virtually identical to System 1 when the signal to be approximated has zero mean value (or is centered).

We will now numerically show that the three ON systems are all complete over the oceans (see (2.20) for the definition of completeness). For system $\{X_j\}$, such a property is clear since we combine all the independent spherical harmonics into such a system; any later spherical harmonics having dependence with previous spherical harmonics are eliminated, thus the completeness of system $\{X_j\}$ is guaranteed by the fact that no other harmonics will be orthogonal to this system (see (4.26)). For system $\{Y_j\}$, we still can use the same argument. However, this time instead of eliminating "later" harmonics, we remove the first spherical harmonic L_0 before we construct system $\{Y_j\}$, since eventually there will be a spherical harmonic that is dependent on L_0 and we can keep that particular one instead of L_0. Indeed, using the Gram matrix, we can find that particular spherical harmonic as it was done numerically. Due to the complex geometry of the oceanic boundary, it is practically impossible to verify the above statement analytically. For the same reason as that given to system $\{X_j\}$, we show that system $\{Y_j\}$ is complete. For system $\{Z_j\}$, the argument is still the same, since eventually we can find the spherical harmonics that are dependent on L_0, L_1, L_2, and L_3, as verified by the author's computer work. Thus we show that the three ON systems are all complete and any signal defined on the oceans can be approximated by them to any accuracy specified. Numerical results showing the convergence of the three ON systems will be given in Chapter 6. The point of presenting three ON systems here is related to the satellite radial orbit error encountered in Chapter 7 where we try to choose an ON system that can avoid the radial orbit error's "sensitivity spectrum" (Wagner, 1986).

We now have more than one ON system on the oceans. We may change the representation of a signal from one system to another. This can be achieved by unitary transformation in a Hilbert space as follows: For the moment, we assume that we have two complex complete ON systems $\{\phi_i\}$ and $\{\psi_i\}$ on some domain. A function defined on that domain may be expanded into $\{\phi_i\}$ and $\{\psi_i\}$ as

$$f = \sum_{i=1}^{\infty} a_i\phi_i = \sum_{i=1}^{\infty} b_i\psi_i \tag{5.78}$$

where a_i and b_i are expansion coefficients with respect to system $\{\phi_i\}$ and system $\{\psi_i\}$, respectively. Now if the expansion coefficients a_i have been obtained in some way and we would like to get the coefficient b_i in order to have a representation of f in terms of ψ_j, then we can perform the transformation:

$$b_i = (f, \psi_i) = \left(\sum_{j=1}^{\infty} a_j \phi_j, \psi_i \right)$$

$$= \sum_{j=1}^{\infty} (\phi_j, \psi_i) a_j$$

$$= \sum_{j=1}^{\infty} \alpha_{ji} a_j \tag{5.79}$$

where

$$\alpha_{ji} = (\phi_j, \psi_i) \tag{5.80}$$

To show that the transformation in (5.79) is unitary, we write $\phi_i = \sum_{k=1}^{\infty} (\phi_i, \psi_k)\psi_k$ and obtain

$$\delta_{ij} = (\phi_i, \phi_j) = \sum_{k=1}^{\infty} (\phi_i, \psi_k)(\psi_k, \phi_j) = \sum_{k=1}^{\infty} (\phi_i, \psi_k)(\phi_j, \psi_k)^*$$

$$= \sum_{k=1}^{\infty} \alpha_{ik} \alpha_{jk}^* \tag{5.81}$$

Imagine that we have a vector of infinite elements formed by α_{ik}, $k = 1, \cdots, \infty$, then two such vectors are orthonormal due to (5.81). Hence (5.79) is a unitary transformation (or an orthogonal transformation if systems $\{\phi_i\}$ and $\{\psi_i\}$ are real). It is possible to find the applications of unitary transformation in our study if the signal to be represented is band limited (a signal is bandlimited with respect to a generalized Fourier series such as (5.78) if $a_i = 0$ for $i > n$ where n is the "highest" frequency). For example, the SST ($\zeta(\theta, \lambda)$) on the oceans may be represented by systems $\{Y_i\}$ and $\{Z_i\}$ as follows:

$$\zeta(\theta, \lambda) = \sum_{i=0}^{M} \beta_i Y_i(\theta, \lambda) = \sum_{i=0}^{N} \gamma_i Z_i(\theta, \lambda) \tag{5.82}$$

where M and N are two numbers corresponding to the highest harmonics for accurately representing $\zeta(\theta, \lambda)$ in systems $\{Y_i\}$ and $\{Z_i\}$, respectively. Therefore,

$$\beta_i = \sum_{j=0}^{N} (Z_j, Y_i)\gamma_j \tag{5.83}$$

where

$$(Z_j, Y_i) = \left(\sum_{p=0}^{j} e_{jp} L_{p+4}, \sum_{q=0}^{i} d_{iq} L_{q+1} \right)$$

$$= \sum_{p=0}^{j} e_{jp} \sum_{q=0}^{i} d_{iq} (L_{p+4}, L_{q+1}) \qquad (5.84)$$

Thus we obtain the representation coefficients β_i from γ_i through such a finite transformation. In the next Chapter, we will show that the β_i obtained by unitary transformations are numerically equal to those by direct expansion of $\zeta(\theta, \lambda)$ into the system $\{Y_i\}$, provided that N goes up to a number corresponding to degree 24 of the ON functions (of system $\{Z_i\}$).

The use of a unitary transformation in the estimation of geoid undulations and SST from satellite altimetry is as follows. We first choose an ON system whose representation of SST yields the least correlation with orbital errors and undulation corrections. The result is then transformed to an ON system of our choice which can then be used for comparison to some existing solutions.

5.4 Some Properties of the ON Functions Constructed by the Gram-Schmidt Process

In this section, we present two important properties of the ON functions constructed by the Gram-Schmidt process. Although the properties are common to any ON functions constructed by this process, the emphasis in the following will be on the ON functions constructed using the spherical harmonics, such as $X_i(\theta, \lambda)$, $Y_i(\theta, \lambda)$, and $Z_i(\theta, \lambda)$ presented in the previous section. Also, in the discussions that follow, system $\{X_i(\theta, \lambda)\}$ will be used. The results can be immediately applied to systems $\{Y_i(\theta, \lambda)\}$ and $\{Z_i(\theta, \lambda)\}$.

Theorem 1: $\left(-\Delta^* X_n, X_n\right) = k_n = [\sqrt{n}]([\sqrt{n}] + 1)$, (5.85)
where Δ^* is the Laplace surface operator defined in (3.98), X_n is defined in (5.73), n = 0,1,2, ... , and $[\sqrt{n}]$ is the integer part of $\sqrt{n}$.

Proof: We recall that the eigenvalue of a spherical harmonic L_p of degree ℓ is $k = \ell(\ell + 1)$. Arranging the spherical harmonics in the sequence shown in (5.72), i.e., $\{L_0, L_1, \ldots L_p\}$, it is easy to see that the eigenvalue of L_p is $k_p = [\sqrt{p}]([\sqrt{p}] + 1)$. For example, $k_0 = 0 \cdot (0 + 1) = 0$ for $\bar{R}_{00}$; $k_1 = 1 \cdot (1 + 1) = 2$ for $\bar{R}_{10}$; $k_2 = 1 \cdot (1 + 1) = 2$ for $\bar{R}_{11}$, etc. Therefore

$$\Delta^* L_p + k_p L_p = 0 \quad \text{or} \quad \Delta^* L_p = -k_p L_p \qquad (5.86)$$

Recalling the definition of X_n in (5.73), we have

$$\left(-\Delta^* X_n, X_n\right) = -\left(\Delta^* \sum_{p=0}^{n} c_{np} L_p, X_n\right)$$
$$= \left(\sum_{p=0}^{n} c_{np} k_p L_p, X_n\right)$$
$$= \left(\sum_{p=0}^{n-1} c_{np} k_p L_p + c_{nn} k_n L_n, X_n\right) \qquad (5.87)$$

From (3.11), we know that $(L_p, X_n) = 0$ if $p < n$. Thus for the inner products in (5.87), only the last one, i.e., (L_n, X_n) will have contributions to the result, the rest are all zero. Since they are zero, we may replace the eigenvalues of L_p, $p < n$ by a constant eigenvalue k_n, and the result will be unchanged. Therefore, we can show that

$$\left(-\Delta^* X_n, X_n\right) = \left(k_n \sum_{p=0}^{n-1} c_{np} L_p + c_{nn} k_n L_n, X_n\right)$$
$$= k_n \left(\sum_{p=0}^{n} c_{np} L_p, X_n\right)$$
$$= k_n (X_n, X_n) = [\sqrt{n}]\,([\sqrt{n}] + 1) \qquad (5.88)$$

where we have made use of the fact that $\{X_p\}$ is an ON system. Such a property of X_n has close resemblance to that of eigenfunctions in the Sturm-Liouville problem $L(u) + ku = 0$, where L is a linear operator, u is an eigenfunction, and k is an eigenvalue, see also Section 3.3.1. However, it will not be possible to find an eigenvalue λ_n for X_n with respect to the Laplace surface operator Δ^*. For, if such a λ_n exists, we will have

$$\Delta^* X_n + \lambda_n X_n = 0$$

Then

$$\sum_{p=0}^{n} c_{np}(k_p - \lambda_n) L_p = 0$$

Since L_p are all linearly independent, all the coefficients of L_p must vanish if the above equation is to be true. Thus

$$c_{np}(k_p - \lambda_n) = 0 \Leftrightarrow \lambda_n = k_p,\ p = 0, \ldots, n$$

This will be impossible since λ_n is a constant while k_p changes as the index p changes. Thus we conclude that X_n does not have an eigenvalue with respect to the operator Δ^*.

Theorem 2: Let $f(\theta,\lambda)$ be a bandlimited signal on the entire sphere with respect to the frequency of spherical harmonics L_p, namely,

$$f(\theta,\lambda) = \sum_{p=0}^{N} a_p L_p(\theta,\lambda) \tag{5.89}$$

where N is a number which corresponds to the maximum "frequency" of $f(\theta,\lambda)$ with respect to the system $\{L_p\}$. Then on the oceans where L_p are orthonormalized to form the ON system $\{X_p\}$, $f(\theta,\lambda)$ is exactly represented by X_p as

$$f(\theta,\lambda)\Big|_{\text{oceans}} = \sum_{p=0}^{N} \alpha_p X_p(\theta,\lambda) \tag{5.90}$$

Proof: We must state the situation further before we prove the theorem. In this case, the signal $f(\theta, \lambda)$ is only sampled on the oceans, even though it is possible to obtain it elsewhere. For example, the geoid is a global signal but only the part on the oceans can be observed by a satellite altimeter. We can further assume that the global geoid is bandlimited in the spherical harmonic expansion. Given such a signal on the oceans only, we would like to determine its property in the spectral domain. Having explained the situation, we denote the signal $f(\theta, \lambda)$ on the oceans as $f_o(\theta, \lambda)$. We assume that we have no idea about the spectral behavior of $f_o(\theta, \lambda)$ and we will expand $f_o(\theta, \lambda)$ into the ON functions $X_n(\theta, \lambda)$ to a degree as high as possible, namely,

$$f_o(\theta,\lambda) = \sum_{p=0}^{M} \alpha_p X_p(\theta,\lambda) \tag{5.91}$$

and α_p is found by

$$\alpha_p = (f_o(\theta,\lambda),\ X_p(\theta,\lambda))_{\text{oceans}} \tag{5.92}$$

where subscript "oceans" is imposed to emphasize that the inner product is carried out over the oceans. Now, one must remember the expansion in (5.89) is valid for the entire sphere, including the oceans. Thus, over the oceans the signal $f_o(\theta, \lambda)$ can be completely recovered by using (5.89) if we know a_n for n = 0, 1 ..., N. (Of course, we do not know anything about a_n). Thus α_p can be obtained by

$$\alpha_p = \left(\sum_{q=0}^{N} a_q L_q(\theta,\lambda),\ X_p \right) \tag{5.93}$$

Again we use the property that $(L_q, X_p) = 0$ if $q < p$, then we have

$$\alpha_p = 0 \text{ if } p > N \tag{5.94}$$

Thus $f_o(\theta, \lambda)$ is bandlimited with respect to the system $\{X_n\}$ and can be exactly represented by

$$f_o(\theta,\lambda) = \sum_{p=0}^{N} \alpha_p X_p(\theta,\lambda) \qquad (5.95)$$

This theorem is particularly useful in case that the sampled data are used. The beauty of this theorem and the ON functions in such a case is that we do not need to have sampled data outside the oceanic area yet we still can determine the highest frequency of the global, bandlimited signal. In the next chapter, we will create 4 sets of 1° x 1° mean geoids using OUS89B potential field (Rapp et al., 1990) up to degrees 15, 18, 20, and 24. We will then expand these sets of geoids into the ON functions using the oceanic values only. It will be seen that, after the corresponding highest degrees of the ON functions for the 4 oceanic geoids, the powers immediately drop to zero. It is not the case when we expand these oceanic geoids into spherical harmonics.

In fact, the signal that we get on the oceans, unlike the geoid, does not have to be a global one. For example, SST is not even defined on land. For the signal defined only on the oceans, the ON functions will prove to be useful for the spectral analysis. With the spherical harmonic representation, one cannot even talk about spectral analysis for the oceanic signal due to some theoretical problems and practical problems concerning some unreasonable phenomena when a relatively high degree is used. We will discuss this shortly.

6. Expansion of Sea Surface Topography and Oceanic Geoid in Orthonormal Functions

This chapter will be devoted to the expansions of SST and oceanic geoid in the orthonormal functions constructed in Chapter 5. The SST data used are the modified SST of Levitus, while the oceanic geoid used is obtained from the OSU89B potential field. The analyses of the results of the expansions will be concentrated on the spectral behavior of the signals in such expansions.

6.1 The SST Data Used in the Expansions and Some Definitions

The SST data set to be used in the following expansions is a modified SST of Levitus described as Set 3 in Engelis (1987b, p. 3). The spatial distribution of the data is given in Figure 5.2. In section 5.2, a short description of this data set can also be found. A good summary of this data set is available in (ibid., pp. 2-3). A contour plot of the SST in this data set is given in Figure 6.1.

This data set consists of 30922 1° x 1° mean SST with the mean value removed. The rms value of the data is 62.4 cm. The data in the Mediterranean Sea were replaced by the estimates from a map by Lisitzin (1974). In Figure 6.1, one can see the major signatures that are associated with oceanic currents. However, it is possible that some spurious data remained even after an editing process has been employed by Levitus (1982). This is justified by irregular contours in the Atlantic Ocean and at the west coast of the South America. According to Engelis (1987b), in this data set any signal with wavelengths less than 800 km is eliminated due to the weight function that was used in editing the data (see also Levitus, 1982, pp. 9-10). We will investigate this statement using the result of an ON function expansion later in this chapter.

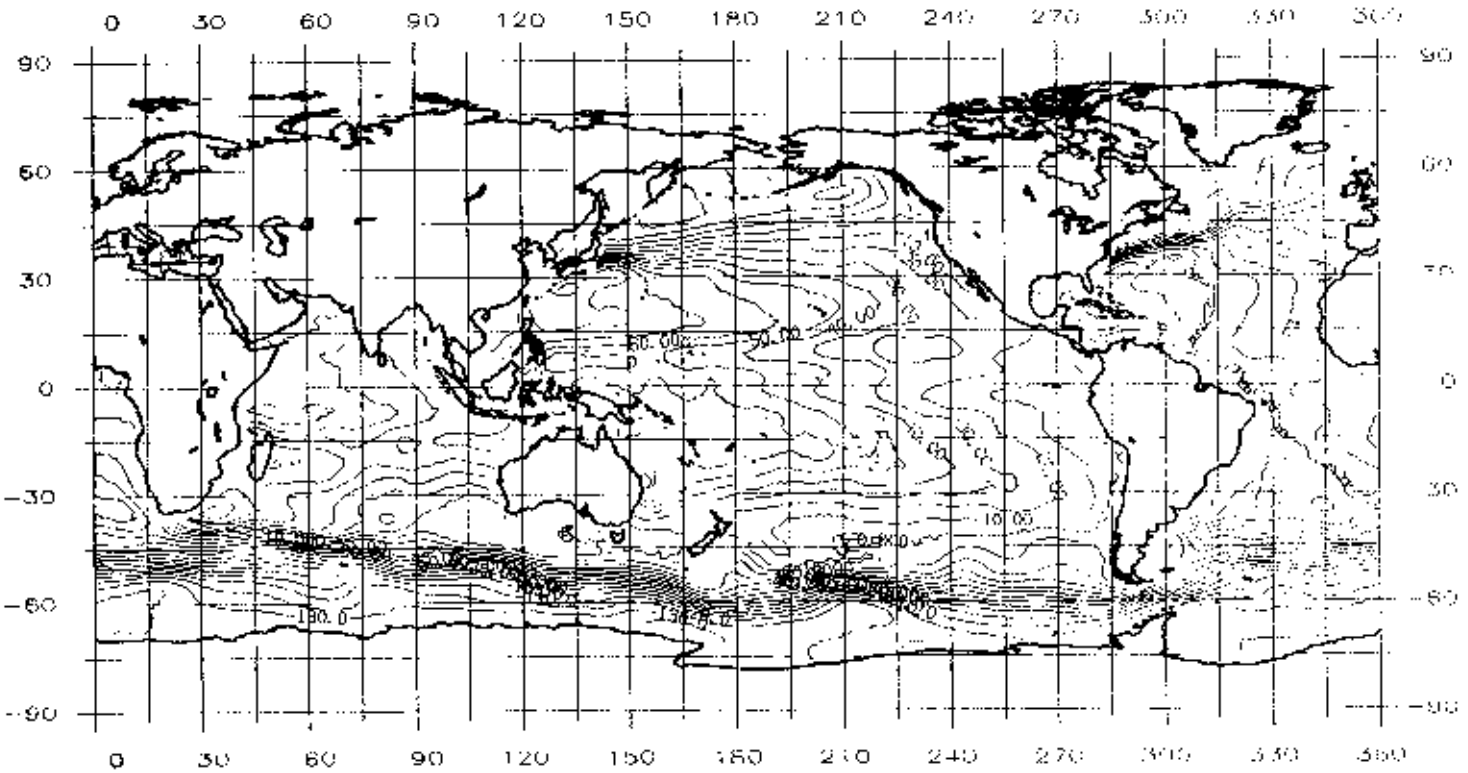


Figure 6.1 Modified Levitus SST (Set 3 in Engelis (1987b)), CI = 10 cm.

Some quantities that will help to explain the results of the ON expansions will be defined here. The total power of a signal f is

$$P_T = \iint_\sigma f^2(\sigma)d\sigma = \|f\|^2_{oceans}, \quad A \tag{6.1}$$

where σ indicates the oceans in which $f(\sigma)$ is defined. For the Levitus SST (from now on, the Levitus SST mean the SST data just discussed above), $P_T = 2.142547$ m^2 which is obtained by numerically integrating (6.1) using the 1° x 1° mean values. The average power of f is

$$P_{avg} = \frac{1}{A} P_T = \|f\|^2_{oceans} \tag{6.2}$$

where A is the area of the oceans (unitless). For the Levitus SST, $P_{avg} = 0.288736$ m^2. If a function is expanded into the ON functions in the system $\{X_j\}$, and the expansion coefficients are α_j, j = 0, 1, ... , then the quadratic content of f (Mayhan, 1984) up to j = N is

$$Q_f(N) = \sum_{j=0}^{N} \alpha_j^2 \tag{6.3}$$

Therefore by Parseval's theorem (see (2.21)) we get $Q_f(\infty) = P_{avg}$. As a measure of the approximations by the ON functions, we define the figure of merit as

$$\eta = \frac{P_{avg} - Q_f(N)}{P_{avg}} \times 100\% \tag{6.4}$$

Of course, the terms P_T, P_{avg}, $Q_f(N)$ and η are not the only measures of the quality of the expansion. Some other factors such as the fit of expansion to the data and the spectral behavior also need to be considered when investigating the expansion results. These will be discussed right after the expansions are completed.

6.2 Expansion Methods

6.2.1 Numerical Quadratures Formula

The first method to be used is the numerical quadratures. A function f defined on the oceans may be expanded into the ON functions in the system $\{X_j(\theta, \lambda)\}$ or $\{O_{nm}(\theta, \lambda), Q_{nm}(\theta, \lambda)\}$ as follows:

$$\hat{f}(\theta, \lambda) = \sum_{j=0}^{\mu} \alpha_j X_j(\theta, \lambda) \tag{6.5}$$

or

$$\hat{f}(\theta, \lambda) = \sum_{n=0}^{N_{max}} \sum_{m=0}^{n} \left(\alpha_{nm} O_{nm}(\theta, \lambda) + \beta_{nm} Q_{nm}(\theta, \lambda)\right) \tag{6.6}$$

where α_j or α_{nm}, β_{nm} are the expansion coefficients. Eq. (6.5) is a single-index form, while (6.6) is a double-index form. μ or N_{max} is the highest "degree" of the expansion and $\mu = (N + 1)^2 - 1$. To make the norm of approximation error $(f - \hat{f})$ minimum, we must have $\alpha_j = (f, X_j)$, as shown in (2.15), namely,

$$\alpha_j = \frac{1}{A}\iint_{\sigma} f(\theta, \lambda)X_j(\theta, \lambda)d\sigma\ ,\ \ \sigma = \text{oceans} \tag{6.7}$$

Using the index function $w_{k\ell}$ in (5.1), the integration becomes

$$\alpha_j = \frac{1}{A}\sum_{k=o}^{N-1}\sum_{\ell=o}^{2N-1} w_{k\ell}\iint_{\Delta\sigma_{k\ell}} f(\theta, \lambda)X_j(\theta, \lambda)\sin\theta d\theta d\lambda \tag{6.8}$$

From (6.7) to (6.8), no assumption is made if the oceans are formed by the equiangular blocks. Now we assume that the function f inside a block is a constant represented by its mean value in the block, then

$$\alpha_j = \frac{1}{A}\sum_{k=o}^{N-1}\sum_{\ell=o}^{2N-1} w_{k\ell}\bar{f}_{k\ell}\iint_{\Delta\sigma_{k\ell}} X_j(\theta, \lambda)\sin\theta d\theta d\lambda \tag{6.9}$$

where $\bar{f}_{k\ell}$ is the mean value at block k, ℓ. Recalling the relationship between the ON function $X_j(\theta, \lambda)$ and spherical harmonic $L_p = \bar{R}_{nm}$ or $\bar{S}_{nm}$ in (5.72), we obtain the desired formula:

$$\alpha_j = \frac{1}{A}\sum_{p=o}^{j} c_{jp}\left(\sum_{k=o}^{N-1}\sum_{\ell=o}^{2N-1} w_{k\ell}\bar{f}_{k\ell}\iint_{\Delta\sigma_{k\ell}} L_p(\theta, \lambda)\sin\theta d\theta d\lambda\right)$$

$$= \frac{4\pi}{A}\sum_{p=o}^{j} c_{jp}a_p \tag{6.10}$$

where c_{jp} are the combination coefficients which have been found in the previous chapter. The value a_p is precisely the spherical harmonic "expansion" coefficient of $(w_{k\ell}\,\bar{f}_{k\ell})$ on the sphere. This is not to say we need to have land value assumption for the SST or the oceanic geoid. The convenience of calculating a_p arises from the way the ON functions are constructed. a_p thus can be calculated by the formulae given by Rapp (1986, pp. 370-371), where FFT has been used. A more compact FFT form is presented as follows: we wish to compute a_p by

$$a_p = \frac{1}{4\pi}\sum_{k=o}^{N-1}\overline{IP}_{nm}^{k}\sum_{\ell=o}^{2N-1} w_{k\ell}\bar{f}_{k\ell}\begin{Bmatrix} IC_m^{\ell} \\ IS_m^{\ell} \end{Bmatrix} \tag{6.11}$$

where $\overline{IP}^k_{nm}$ is the integration of the fully normalized Legendre function $\overline{P}^m_n(\cos\theta)$ (whose degree n and order m depends on the index p) from θ_k to θ_{k+1}, and

$$\begin{Bmatrix} IC^\ell_m \\ IS^\ell_m \end{Bmatrix} = \int_{\lambda_\ell}^{\lambda_{\ell+1}} \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} d\lambda \tag{6.12}$$

Let IE^ℓ_m be defined as

$$IE^\ell_m = \int_{\lambda_\ell}^{\lambda_{\ell+1}} e^{im\lambda} d\lambda = h(m) e^{im\ell\Delta\lambda} \tag{6.13}$$

where

$$h(m) = \begin{cases} \Delta\lambda = \lambda_{\ell+1} - \lambda_\ell \ , & m = 0 \\ \dfrac{i\left(1 - e^{im\Delta\lambda}\right)}{m} \ , \ i = \sqrt{-1} \ , & m \neq 0 \end{cases} \tag{6.14}$$

Futhermore, we define

$$F^k_m = \sum_{\ell=0}^{2N-1} w_{k\ell} \bar{f}_{k\ell} IE^\ell_m = h(m) FFT\left(w_{k\ell} \bar{f}_{k\ell}\right) \tag{6.15}$$

where $FFT(w_{k\ell}\, \bar{f}_{k\ell})$ is the (fast) Fourier Transform of $(w_{k\ell}\, \bar{f}_{k\ell})$. Using Euler's formula, we have

$$\sum_{\ell=0}^{2N-1} w_{k\ell} \bar{f}_{k\ell} \begin{Bmatrix} IC^\ell_m \\ IS^\ell_m \end{Bmatrix} = \begin{Bmatrix} Re\left(F^k_m\right) \\ Im\left(F^k_m\right) \end{Bmatrix} \tag{6.16}$$

Finally we obtain

$$a_p = \frac{1}{4\pi} \sum_{k=0}^{N-1} \overline{IP}^k_{nm} \begin{Bmatrix} Re\left(F^k_m\right) \\ Im\left(F^k_m\right) \end{Bmatrix} \tag{6.17}$$

It is very easy to verify that (6.17) is a compact form of eq. (30) in (1986, p. 371), up to a constant. The program for computing a_p is obtained by modifying Colombo's subroutine "HARMIN" (Colombo, 1981, p. 107).

One concern is the mean value approximation made in (6.9). A standard technique for reducing such an approximation is the introduction of a desmoothing operator such as the one given by Rapp (1986, eq. (118)). Specifically, the desmoothing q_n operator was

used in the recovery of geopotential coefficients from mean gravity anomalies and was given by Pellinen (1966, p. 83)

$$q_n = \frac{\int_o^\pi A(\psi)P_n(\cos\psi)\sin\psi d\psi}{\int_o^\pi A(\psi)\sin\psi d\psi} \tag{6.18}$$

where

$$A(\psi) = \begin{cases} 1 & , \quad \psi \le \psi_o \\ 0 & , \quad \psi > \psi_o \end{cases} \tag{6.19}$$

ψ_o is the size of a spherical cap where a mean gravity anomaly is to be computed, n is the degree of Legendre function. In reality, we do not form the mean value in a spherical cap, rather in an equiangular block $\Delta\theta \times \Delta\lambda$. Thus for a given $\Delta\theta$ value the corresponding ψ_o value needs to be obtained. For example, Katsambalos (1979, p. 70) used

$$\sin\left(\frac{\psi_o}{2}\right) = \left[\frac{\Delta\theta\sin\Delta\theta}{4\pi}\right]^{1/2} \tag{6.20}$$

In the practical application of the desmoothing operator q_n, further considerations based on the ranges of degree can be made, as given by Colombo (1981, p. 76).

The desmoothing operator q_n is due to an important property of spherical harmonics stated by Meissl (1971, p. 57) or Miller (1962, p. 22):

$$\int_{-1}^{1}\int_o^{2\pi} k(\psi)\begin{Bmatrix} R_{nm}(\theta,\lambda) \\ S_{nm}(\theta,\lambda) \end{Bmatrix} dtd\lambda = \lambda_n \begin{Bmatrix} R_{nm}(\theta_o,\lambda_o) \\ S_{nm}(\theta_o,\lambda_o) \end{Bmatrix} \tag{6.21}$$

where $t = \cos\theta$, λ_n is the eigenvalue with respect to the kernel $K(\psi)$, ψ is the spherical distance between point (θ_o, λ_o) and an arbitrary point (θ, λ) on the sphere. The relationships between (θ_o, λ_o), (θ, λ) and ψ can be found in Figure 6.2. The value of λ_n is given by Meissl (1971, eq. (6.46a))

$$\lambda_n = 2\pi\int_o^{2\pi} K(\psi)P_n(\cos\psi)\sin\psi d\psi \tag{6.22}$$

Replacing $K(\psi)$ by $A(\psi)$ in (6.22), we get the numerator in (6.18).

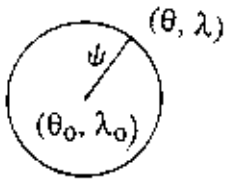


Figure 6.2 Variable ψ in Equation (6.21).

Using exactly the same argument as the one given for the eigenvalue of operator Δ^* in Section 5.4, we can show that there is no eigenvalue with respect to $K(\psi)$ for the ON functions $X_j(\theta, \lambda)$. Therefore, an analytical desmoothing operator cannot be found for $X_j(\theta, \lambda)$. Nevertheless, if the function to be expanded possesses a long wavelength feature the approximation made in (6.9) should not cause too much trouble, since the values of such a function in a block such as 1° x 1° should not vary substantially and can be assumed to be a constant. In the numerical experiments that follow, no desmoothing operator has been used.

The expansions of a function in the other two ON systems, i.e., $\{Y_j(\theta, \lambda)\}$ and $\{Z_j(\theta, \lambda)\}$ follow the same formulae given above, except that the combination coefficients need to be changed and an appropriate index for $\{L_p\}$ must be used. Further, in accordance with the definitions of degree variance and error degree variance in a spherical harmonic expansion, we define the corresponding quantities for an ON function expansion as follows:

Degree variance of ON coefficients : τ_n^2

$$\tau_n^2 = \sum_{m=o}^{n} \left(\alpha_{nm}^2 + \beta_{nm}^2\right) \tag{6.22}$$

where α_{nm} and β_{nm} are defined in (6.6).

Error degree variance of ON coefficients : ε_n^2

$$\varepsilon_n^2 = \sum_{m=o}^{n} \left(\varepsilon_{\alpha_{nm}}^2 + \varepsilon_{\beta_{nm}}^2\right) \tag{6.23}$$

where $\varepsilon_{\alpha_{nm}}$ and $\varepsilon_{\beta_{nm}}$ are the standard deviations of the expansion coefficients α_{nm} and β_{nm}.

We must emphasize that for the ON functions in the systems $\{X_j\}$, $\{Y_j\}$ and $\{Z_j\}$ introduced in Chapter 5, the terms "degree" and "order" have been used for the frequency classification in the same way as the "degree" and "order" for the spherical harmonics which are orthogonal on the entire sphere. When we analyze the expansion results, we often compare the degree variances of the ON functions and those of the spherical harmonics in the same table or figure. Thus care must be exercised in correctly interpreting the quantities of interest.

6.2.2 Least Squares Fit (ℓsf)

In the numerical quadratures formula discussed in the previous section (eq. (6.7)) the expansion coefficients are found by letting $\|f - \hat{f}\|^2$ be minimum, i.e.,

$$\|f - \hat{f}\|^2 = \frac{1}{A}\iint_{\sigma} (f - \hat{f})^2 d\sigma = \text{a minimum} \tag{6.24}$$

The total error $(f - \hat{f})$ is evaluated on the oceans by integration. Now we shall use a discrete version of (6.24) as a minimum criteria for calculating the coefficients. We still let the norm of the approximation error $(f - \hat{f})$ be minimum, but the norm of error will be defined in a different manner. To do this we calculate the mean value of $f(\theta, \lambda)$ at block ℓ by the expansion

$$\bar{f}_\ell(\theta, \lambda) = \frac{1}{\Delta\sigma_\ell}\iint_{\Delta\sigma_\ell} f(\theta, \lambda) d\sigma = \frac{1}{\Delta\sigma_\ell}\sum_{j=o}^{\infty} \alpha_j \iint_{\Delta\sigma_\ell} X_j(\theta, \lambda) d\sigma$$

$$= \frac{1}{\Delta\sigma_\ell}\sum_{j=o}^{\infty} \alpha_j I X_j^\ell \tag{6.25}$$

Now in (6.25), we let the maximum expansion "degree" be a finite number, μ. An error e_ℓ will be introduced in the approximation, namely

$$\bar{f}_\ell(\theta, \lambda) + e_\ell = \frac{1}{\Delta\sigma_\ell}\sum_{j=o}^{\mu} \alpha_j I X_j^\ell \tag{6.26}$$

Due to the use of finite expansion terms in (6.26), e_ℓ can be interpreted in two ways: If $\bar{f}_\ell$ is errorless, then e_ℓ is purely the "truncation error"; if the given data $\bar{f}_\ell$ has noise then the noise will be blended with the truncation error in the form of (6.26). In any case, we could solve for α_j by requiring

$$\phi = \sum_{\ell=1}^{M} e_\ell^2 = V_s^T V_s = \text{a minimum} \tag{6.27}$$

where $V_s = (e_1, e_2, \ldots, e_M)^T$ is a vector containing the errors with M ($M > \mu$) being the number of mean values used (or the number of the $\Delta\theta$ x $\Delta\lambda$ blocks that form the oceans). The use of (6.27) gives rise to the familiar least-squares adjustment problem in geodesy in which a solution is readily established. The squares of the error norm measured in (6.27) is a discrete version of (6.24). Recalling the discrete version of inner products defined in (2.9), we can also easily see that

$$\varphi = (e(\theta, \lambda), e(\theta, \lambda))_{oceans} = \|error\|^2 \tag{6.28}$$

where $e(\theta, \lambda)$ is the pointwise error function on the oceans. Now we can treat (6.26) as "observation equations" at "points" ℓ, $\ell = 1, 2, ..., M$. In accordance with the classical least-squares adjustment problem (Uotila, 1986), we define the following matrices:

$$A_s = \text{design matrix} = \begin{bmatrix} IX_1^1/\Delta\sigma_1 & IX_2^1/\Delta\sigma_1 & \cdots & IX_\mu^1/\Delta\sigma_1 \\ \vdots & & & \vdots \\ IX_1^M/\Delta\sigma_M & IX_2^M/\Delta\sigma_M & \cdots & IX_\mu^M/\Delta\sigma_M \end{bmatrix}_{M\times\mu} \tag{6.29}$$

$$X_s = \text{unknown vector} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_\mu \end{bmatrix}_{\mu\times 1} \tag{6.30}$$

$$L_s = \text{observation vector} = \begin{bmatrix} \bar{f}_1 \\ \vdots \\ \bar{f}_M \end{bmatrix}_{M\times 1} \tag{6.31}$$

Thus (6.26) can be written in a matrix form:

$$V_s = A_s X_s - L_s \tag{6.32}$$

To get a minimum error norm according to (6.27), we must have

$$\frac{\partial\varphi}{\partial\alpha_j} = 0\ ,\ \ j = 1, ..., \mu \tag{6.33}$$

The solution for the unknown vector X_s is then

$$X_s = (A_s^T A_s)^{-1}(A_s^T L_s) \tag{6.34}$$

Eq. (6.34) can be reduced to a form which can take advantage of the least squares fit problem using the spherical harmonics. First of all, we recall the relationship between the ON functions and the spherical harmonics in (5.73):

$$\begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_\mu \end{bmatrix} = \begin{bmatrix} c_{00} & & & \\ c_{10} & c_{11} & & 0 \\ \vdots & & \ddots & \\ c_{\mu 0} & c_{\mu 1} & \cdots & c_{\mu\mu} \end{bmatrix} \begin{bmatrix} L_0 \\ L_2 \\ \vdots \\ L_\mu \end{bmatrix} \tag{6.35}$$

where c_{ij} are the combination coefficients. Let C denote the matrix of the combination coefficients in (6.35). Now we would like to find the relationship between matrix A_s in

(6.29) and a matrix that contains the integrations of spherical harmonics. To this end, we transpose and integrate both sides of (6.35), then divide the result by the area of a block, -

$$A_s = \begin{bmatrix} \mathrm{IL}_1^1/\Delta\sigma_1 & \mathrm{IL}_2^1/\Delta\sigma_1 \cdots & \mathrm{IL}_\mu^1/\Delta\sigma_1 \\ \vdots & & \vdots \\ \mathrm{IL}_1^M/\Delta\sigma_M & \mathrm{IL}_2^M/\Delta\sigma_M \cdots & \mathrm{IL}_\mu^M/\Delta\sigma_M \end{bmatrix} C^T$$

$$= BC^T \tag{6.36}$$

where matrix B is easily identified by the relationship in the equation, and

$$\mathrm{IL}_p^\ell = \iint_{\Delta\sigma_\ell} L_p d\sigma \tag{6.37}$$

is the integration of surface spherical harmonic at block ℓ. Substituting (6.36) into (6.34), we have

$$X_S = (C^T)^{-1}(B^TB)^{-1}(B^TL_s) \tag{6.38}$$

$$= (C^T)^{-1}Y_S$$

where

$$Y_S = (B^TB)^{-1}(B^TL_S) \tag{6.39}$$

is regarded as an intermediate vector for solution X_S. Therefore, we can split the solution into two steps. The first step is to get Y_s by (6.39), then X_s is found by (6.38). To obtain Y_S, we first interpret the meaning of B^TB and B^TL_S: The elements K_{pq}^j in matrix B^TB are

$$K_{pq}^j = \sum_{\ell=1}^{M} \frac{1}{\Delta\sigma_\ell^2} \mathrm{IL}_p^\ell \mathrm{IL}_q^\ell \quad , \quad j = 1, 2, 3, 4 \tag{6.40}$$

Recalling that IL_p^ℓ is the integration of spherical harmonic $\overline{R}_{nm} = \overline{P}_n^m(\cos\theta)\cos m\lambda$ or $\overline{S}_{nm} = \overline{P}_n^m(\cos\theta)\sin m\lambda$, and using the index function $w_{k\ell}$ in (5.1), we can write K_{pq}^j explicitly as

$$\begin{Bmatrix} K_{nmrs}^1 \\ K_{nmrs}^2 \\ K_{nmrs}^3 \\ K_{nmrs}^4 \end{Bmatrix} = \sum_{k=0}^{N-1} \frac{1}{\Delta\sigma_k^2} \overline{\mathrm{IP}}_{nm}^k \overline{\mathrm{IP}}_{rs}^k \sum_{\ell=0}^{2N-1} w_{k\ell} \begin{Bmatrix} \mathrm{IC}_m^\ell \mathrm{IC}_s^\ell \\ \mathrm{IS}_m^\ell \mathrm{IS}_s^\ell \\ \mathrm{IC}_m^\ell \mathrm{IS}_s^\ell \\ \mathrm{IS}_m^\ell \mathrm{IC}_s^\ell \end{Bmatrix} \tag{6.41}$$

where n, m correspond to p, and r, s correspond to q in (6.40). The area of a block will depend on latitude only, thus the notation $\Delta\sigma_k$ is used. Note that not every element in (6.41) will enter matrix B^TB. Similarly, the element T_{nm}^j in matrix B^TL_S are

$$\begin{Bmatrix} T^1_{nm} \\ T^2_{nm} \end{Bmatrix} = \sum_{k=0}^{N-1} \frac{1}{\Delta\sigma_k} \overline{IP}^k_{nm} \sum_{\ell=0}^{2N-1} w_{k\ell} \bar{f}_{k\ell} \begin{Bmatrix} IC^\ell_m \\ IS^\ell_m \end{Bmatrix} \tag{6.42}$$

Again, we know that elements T^2_{n0} will not enter $B^T L_s$. The elements K^j_{numrs} in (6.41) and T^j_{num} in (6.42) have close resemblances with the elements shown in equation (4.27) and equation (4.29) in (Pavlis, 1988, pp. 72-73). To evaluate these elements, one can just accumulate the contributions at each latitude belt k in a straight forward manner, and take advantage of the property of the associated Legendre function $\bar{P}^m_n(-t) = (-1)^{n+m}\bar{P}^m_n$. (See Pavlis, 1988, Chapter 4, for a nice discussion). Nevertheless, the regular forms in (6.41) and (6.42) should enable a more efficient method to be developed. We now exploit these forms by FFT and later we shall arrive at an astonishing result as compared to the result obtained in Pavlis (1988, Table 5) in terms of computational efficiency.

We shall basically follow the principle used in Section 5.1.1. It is clear that

$$IE^\ell_m = h(m)e^{im\ell\Delta\lambda} = IC^\ell_m + iIS^\ell_m \tag{6.43}$$

where all the needed definitions can be found in (6.12), (6.13) and (6.14). Let us define

$$\begin{aligned} \alpha &= IC^\ell_m\, IC^\ell_s \\ \beta &= IS^\ell_m\, IS^\ell_s \\ \gamma &= IC^\ell_m\, IS^\ell_s \\ \delta &= IS^\ell_m\, IC^\ell_s \end{aligned} \tag{6.44}$$

and

$$\begin{aligned} a &= IE^\ell_m\, IE^\ell_{-s} = h(m)h(-s)e^{i(m-s)\ell\Delta\lambda} \\ b &= IE^\ell_m\, IE^\ell_s = h(m)h(s)e^{i(m+s)\ell\Delta\lambda} \\ c &= IE^\ell_{-m}\, IE^\ell_{-s} = h(-m)h(-s)e^{-i(m+s)\ell\Delta\lambda} \\ d &= IE^\ell_{-m}\, IE^\ell_s = h(-m)h(s)e^{-i(m-s)\ell\Delta\lambda} \end{aligned} \tag{6.45}$$

If we expand a, b, c and d into real parts and imaginary parts using the relationship given in (6.43), we get

$$\begin{aligned} a &= \alpha + \beta - i(\gamma - \delta) \\ b &= \alpha - \beta + i(\gamma + \delta) \\ c &= \alpha - \beta - i(\gamma + \delta) \end{aligned} \tag{6.46}$$

$$d = \alpha + \beta + i(\gamma - \delta)$$

Thus

$$\begin{aligned} \alpha &= \frac{1}{4}\ (a + b + c + d) \\ \beta &= \frac{1}{4}\ (a - b - c + d) \\ \gamma &= \frac{1}{4i}\ (-a + b - c + d) \\ \delta &= \frac{1}{4i}\ (a + b - c - d) \end{aligned} \tag{6.47}$$

From (6.46), it is easy to see that

$$a = d^* \ ,\ b = c^* \tag{6.48}$$

or

$$a + b = (c + d)^* \ ,\ a - b = (d - c)^* \tag{6.49}$$

Furthermore, we define the FFT of $w_{k\ell}$ at frequency (m + s) as

$$F^k(m, s) = \sum_{\ell=0}^{2N-1} w_{k\ell}\, e^{i(m+s)\ell\Delta\lambda} \ ,\ \Delta\lambda = \frac{2\pi}{2N} \tag{6.50}$$

where k indicates the k^{th} latitude belt. Moreover, we define

$$U^k(m, s) = h(m)h(-s)F^k(m, -s) + h(m)h(s)F^k(m, s) \tag{6.51}$$

$$V^k(m, s) = h(m)h(-s)F^k(m, -s) - h(m)h(s)F^k(m, s) \tag{6.52}$$

Using (6.44) to (6.52), it is not difficult to see that

$$\begin{Bmatrix} K^1_{nmrs} \\ K^2_{nmrs} \\ K^3_{nmrs} \\ K^4_{nmrs} \end{Bmatrix} = \frac{1}{4} \sum_{k=0}^{N-1} \frac{1}{\Delta\sigma_k^2}\, \overline{IP}^k_{nm}\, \overline{IP}^k_{rs} \begin{Bmatrix} [U^k(m, s) + (U^k(m, s))^*] \\ [V^k(m, s) + (V^k(m, s))^*] \\ [-V^k(m, s) + (V^k(m, s))^*]/i \\ [U^k(m, s) - (U^k(m, s))^*]/i \end{Bmatrix}$$

$$= \frac{1}{2} \sum_{k=0}^{N-1} \frac{1}{\Delta\sigma_k^2} \overline{IP}^k_{nm} \overline{IP}^k_{rs} \begin{Bmatrix} Re(U^k(m, s)) \\ Re(V^k(m, s)) \\ Im(-V^k(m, s)) \\ Im(U^k(m, s)) \end{Bmatrix} \tag{6.53}$$

which is the desired FFT form for evaluating the elements in matrix B^TB. To get the required elements, one can follow exactly the same computational procedure listed at the end of Section 5.1.1. The process will be first calculating K^j_{nmrs}, $j = 1, 2, 3, 4$, then selecting the needed elements and finally identifying the right positions of the selected elements in matrix B^TB.

It is much easier to deal with elements T^j_{nm} required in matrix B^TL_s. In fact, the needed FFT algorithm has been developed in (6.17). One only needs to remove the factor 4π and consider the area element $\Delta\sigma_k$ at each latitude belt in (6.17). Specifically, the computational formula for T^j_{nm} is

$$\begin{Bmatrix} T^1_{nm} \\ T^2_{nm} \end{Bmatrix} = \sum_{k=0}^{N-1} \frac{1}{\Delta\sigma_k} \overline{IP}^k_{nm} \begin{Bmatrix} Re(F^k_m) \\ Im(F^k_m) \end{Bmatrix} \tag{6.54}$$

where the definition of F^k_m is given in (6.15).

A program called FFTSOL has been developed for computing the elements needed in matrices B^TB and B^TL_s by the FFT approach and solving for the Y_s vector. A program which solves for Y_s based on the straight forward accumulation method was also available in Professor Rapp's program library (it is called ADJSST, the program sequence number is to be determined). Now we shall call the process of obtaining elements K^j_{nmrs} and T^j_{nm} as the formation of the "normal matrix". According to Pavlis (1988, p. 88), the most expensive part in solving for Y_s is the formation of the normal matrix. Let us now look at how much improvement the FFT method can achieve as opposed to the conventional method (in program ADJSST). In Table 6.1, we list the CPU times on CRAY Y-MP/864 needed for the formations of normal matrices by programs FFTSOL and ADJSST. The CPU times for inversions of (B^TB), which are the same for both programs, are also listed. Nmax is the maximum degree of the spherical harmonics. The solutions Y_s by the two programs are exactly the same within the accuracy of the computer used. However, as we can see from Table 6.1, enormous saving of computer time in the formations of normals can be achieved by the FFT method. One can also observe that the saving factor (CPU ratio in Table 6.1) grows with the Nmax used. This is due to the fact that a larger Nmax will enable a more efficient use of the FFT method.

Table 6.1 CPU Times Comparison using FFTSOL and ADJSST on CRAY Y-MP/864.

Nmax	Normals (seconds) FFTSOL	Normals (seconds) ADJSST	CPU ratios	Inversions† (seconds)
10	0.409	7.015	17	0.042
15	0.882	41.961	48	0.237
24	3.856	186.120	65	2.160
36	9.273	890.410	96	17.285
50	27.319	~3272	121	103.740
70	86.930	~12986	149	~750

~ estimated

† same for both programs. Linpack's routines SPPCO and SPPDI (Dongarra et al., 1979) are used.

The FFT method will achieve the maximum efficiency in the case that Nmax is equal to the Nyguist frequency (2N/2) due to the fact that no waste of computation in (6.50) will be made in such a case. Roughly after Nmax = 30, the CPU time needed for inversion will exceed the CPU time needed for the formulation of the normal equations in the FFT method. In the conventional method, however, the former is always less than the latter.

The success of FFT method suggests that the formation of normal matrix for the geopotential coefficients from the surface gravity anomalies can be made much more efficient. To see this, we list the alternative forms of equations (4.27) and (4.29) in (Pavlis, 1988, pp. 73-74):

$$[N]\,C^{\alpha}_{nm}C^{\beta}_{rs} = GM^2(n-1)(r-1)\sum_{k=0}^{N-1}\frac{1}{\Delta\sigma_k^2}\overline{IP}^k_{nm}\,\overline{IP}^k_{rs}\sum_{\ell=0}^{2N-1}\frac{1}{\bar{r}^4_{k\ell}}\left[\frac{a}{\bar{r}_{k\ell}}\right]^{n+r}P_{k\ell}\begin{Bmatrix} IC^{\ell}_m\,IC^{\ell}_s \\ IS^{\ell}_m\,IS^{\ell}_s \\ IC^{\ell}_m\,IS^{\ell}_s \\ IS^{\ell}_m\,IC^{\ell}_s \end{Bmatrix} \qquad (6.55)$$

$$[U]C^{\alpha}_{nm} = GM(n-1)\sum_{k=0}^{N-1}\frac{1}{\Delta\sigma_k}\overline{IP}^k_{nm}\sum_{\ell=0}^{2N-1}\frac{1}{\bar{r}^2_{k\ell}}\left[\frac{a}{\bar{r}_{k\ell}}\right]^{n}P_{k\ell}\,\overline{\Delta g}_{k\ell}\begin{Bmatrix} IC^{\ell}_m \\ IS^{\ell}_m \end{Bmatrix} \qquad (6.56)$$

where $[N]C^{\alpha}_{nm}C^{\beta}_{nm}$ are the elements of the normal matrix, $[U]C^{\alpha}_{nm}$ are elements of the "U vector" (the vector on the right-hand side of the normal equations), and C^{α}_{nm} are geopotential coefficients. To avoid unnecessary descriptions, the reader is referred to Pavlis (1988, Chapter 2) for the precise definitions of GM, $\bar{r}_{k\ell}$, a and $\overline{\Delta g}_{k\ell}$. Here the data $\Delta g_{k\ell}$ are the mean gravity anomalies. The weight function $P_{k\ell}$ is different from the index function $w_{k\ell}$ in (6.41) in that

$$P_{k\ell} = \begin{cases} \dfrac{1}{\sigma^2_{k\ell}}\ , & \text{mean gravity anomaly exists at block } k\ell \\ 0\ , & \text{mean gravity anomaly does not exist at block } k\ell \end{cases} \qquad (6.57)$$

where $\sigma_{k\ell}$ is the standard deviation of the mean gravity anomaly $\overline{\Delta g}_{k\ell}$. More computer time is expected in this case due to the factors

$$\left[\frac{a}{\bar{r}_{k\ell}}\right]^{n+r} , \left[\frac{a}{\bar{r}_{k\ell}}\right]^{n}$$

that involve degrees n, r at each latitude belt. Unlike the forms in (6.41) and (6.42) where only one FFT process is needed at one latitude belt for all degrees and orders, the forms in (6.55) and (6.56) also require the FFT processes for degree n and r at each latitude belt. To eliminate the degree dependence (or the ratio $a/\bar{r}_{k\ell}$ in (6.58)) of the FFT process, one can reduce the mean gravity anomalies to a sphere with radius = a using downward continuation and ellipsoidal correction (see Rapp, 1986). By doing such a reduction, we can achieve exactly the same efficiency as we have in (6.53) and (6.54).

If we insist on the rigorous forms in (6.55) and (6.56) and consider the degree dependence for the FFT process, then the desired formulae for a FFT approach are

$$[N]_{C_{nm}^{\alpha} C_{rs}^{\beta}} = \frac{1}{2} GM^2(n-1)(r-1) \sum_{k=0}^{N-1} \frac{1}{\Delta\sigma_k^2} \overline{IP}_{nm}^{k} \overline{IP}_{rs}^{k} \begin{Bmatrix} \mathrm{Re}\left(\overline{U}^{k}(n+r, m, s)\right) \\ \mathrm{Re}\left(\overline{V}^{k}(n+r, m, s)\right) \\ \mathrm{Im}\left(-\overline{V}^{k}(n+r, m, s)\right) \\ \mathrm{Im}\left(\overline{U}^{k}(n+r, m, s)\right) \end{Bmatrix} \qquad (6.58)$$

$$[U]_{C_{nm}^{\alpha}} = GM(n-1) \sum_{k=0}^{N-1} \frac{1}{\Delta\sigma_k} \overline{IP}_{nm}^{k} \begin{Bmatrix} \mathrm{Re}\left(\overline{E}^{k}(n, m)\right) \\ \mathrm{Im}\left(\overline{E}^{k}(n, m)\right) \end{Bmatrix} \qquad (6.59)$$

where

$$\overline{U}^{k}(n+r, m, s) = h(m)h(-s)\overline{F}^{k}(n+r, m, -s) + h(m)h(s)\overline{F}^{k}(n+r, m, s) \qquad (6.60)$$

$$\overline{V}^{k}(n+r, m, s) = h(m)h(-s)\overline{F}^{k}(n+r, m, -s) - h(m)h(s)\overline{F}^{k}(n+r, m, s) \qquad (6.61)$$

and

$$\overline{F}^{k}(n+r, m, s) = \sum_{\ell=0}^{2N-1} \frac{1}{\bar{r}_{k\ell}^{4}} \left[\frac{a}{\bar{r}_{k\ell}}\right]^{n+r} P_{k\ell}\, e^{i(m+s)\ell\Delta\lambda} \qquad (6.62)$$

$$\overline{E}^{k}(n, m) = h(m) \sum_{\ell=0}^{2N-1} \frac{1}{\bar{r}_{k\ell}^{2}} \left[\frac{a}{\bar{r}_{k\ell}}\right]^{n} P_{k\ell}\, \overline{\Delta g}_{k\ell}\, e^{im\ell\Delta\lambda} \qquad (6.63)$$

These formulae can be obtained by using exactly the same derivations for elements K_{nmrs}^{j} and T_{nm}^{j}. A program called FFTSOLA was also developed for computing the elements in (6.58) and (6.59) by FFT. Having these elements, we can calculate the disturbing geopotential coefficients $\overline{C}_{nm}$, $\overline{S}_{nm}$ in the formula (Pavlis, 1988, eq. (4.11))

$$\overline{\Delta g_{k\ell}} = \frac{1}{\Delta\sigma_k}\frac{GM}{\bar{r}_{k\ell}^2}\sum_{\substack{n=0\\ n\neq 1}}^{Nmax}(n-1)\left[\frac{a}{\bar{r}_{k\ell}}\right]^n\sum_{m=0}^{n}\left(\bar{C}_{nm}\,IC_m^{\ell}+\bar{S}_{nm}IS_m^{\ell}\right)\overline{IP}_{nm}^{k} \tag{6.64}$$

A simulated data set of mean gravity anomalies was generated from the OSU89B geopotential model (Rapp et al., 1990) to Nmax = 50 at 1° x 1° blocks (the ratio $a/\bar{r}_{k\ell}$ is considered). Then the data set was served as the input file for program FFTSOLA to compute the geopotential coefficients $\bar{C}_{nm}$, $\bar{S}_{nm}$ to Nmax = 50. Comparing the computed coefficients with the original ones, some numerical differences (about 1% of the signal after degree 24) were found. Since the recovery of geopotential coefficients from Δg is beyond the scope of this study, no further effort was made to find out the reason of numerical differences. The discussion here is merely to emphasize the importance of the FFT technique. Further, to show the substantial reduction of computer times in forming the elements in (6.58) and (6.59) by FFT, we list the CPU times on CRAY Y-MP/864 needed for these calculations in Table 6.2.

Table 6.2 CPU Times on CRAY Y-MP/864 for Forming Normals and U Vectors by FFT

Nmax	CPU times† (seconds)
10	3.661
15	6.144
24	11.636(432)*
36	24.617(2119)
50	53.206(8644)
70	146.154

† The ratio $a/\bar{r}_{k\ell}$ is considered
* CPU times in parenthesis are from Rapp (1989b, p. 278, Table 1), who used CRAY X-MP/24.

The CPU times in Table 6.2 are much less than those in Table 1 of Rapp (1989b). One reason is the different computers used; another is of course the methods used. To do an expansion to Nmax = 100 from the surface gravity anomalies, it is estimated that 1300 seconds is needed for forming the normal by FFT, while 6400 seconds is needed for inverting the normal. The memory required will be about 66 Megawords (for FFTSOLA). So now one can see that the major difficulty in high degree expansion using "least squares adjustment" method is the CPU time of inversion and computer's storage space, even if a FFT approach as illustrated above is used. To overcome this difficulty, one may just consider the near diagonal terms of the normal as suggested by Rapp (1989b, p. 276), provided that the parameters (i.e., the coefficients) are organized in an optimum way so that the terms not considered can be safely assumed to be zero.

Returning to (6.34), we have accomplished our goal of finding the coefficients of ON expansion by solving Y_s by the FFT method. In the experiments shown later, we will find that the numerical quadratures method and the method of least squares error fit produce essentially the same expansion coefficients. Such an outcome is expected since the two methods are closely related.

6.2.3 Correlation Analysis of Two Signal Components in the ON Function Expansion

Spectral analysis of a function using certain basis functions lies on an important property, namely, independence of two signal components. Without independence we essentially cannot isolate one signal component from the other, thus the term "spectral analysis" does not make too much sense in such a case. In the statistical sense, independence implies null correlation between two signal components. However, the term "correlation" has somewhat different definitions in the geophysical literature and the tatistical literature. Geophysicists (e.g., Bath, 1974, Chapter 3) define correlation in connection with signal analysis of analytically defined functions, although in most cases empirically observed (or discrete) data are used; statisticians define correlation in the context of estimation theory. Nevertheless, certain relationships between these two definitions exist. We will discuss their relationships below by gradually stepping from the geophysicist's point of view to the statistician's point of view. Such a discussion will help to interpret the results from the ON function expansions or the spherical harmonic expansions.

As we have shown in (2.24), the two signal components $\alpha_p X_p(\theta, \lambda)$ and $\alpha_q X_q(\theta, \lambda)$ in the ON expansion (6.5) are statistically independent, since (assuming ergodicity)

$$\overline{M}(\alpha_p\alpha_q X_p X_q) = \alpha_p\alpha_q \frac{1}{A}\iint_\sigma X_p X_q d\sigma = \alpha_p^2\delta_{pq} \tag{6.65}$$

where $\overline{M}$ is the averaging operator over the oceans. In a more general discussion, Let us define the correlation coefficient between two ergodic signals $f_1(\sigma)$, $f_2(\sigma)$ on σ as (cf. Bath, 1974, p. 90 and Papoulis, 1977, p. 359)

$$\bar{c}_{12} = \frac{\iint_\sigma f_1(\sigma)f_2(\sigma)d\sigma}{\left\{\iint_\sigma f_1^2(\sigma)d\sigma \cdot \iint_\sigma f_2^2(\sigma)d\sigma\right\}^{1/2}} \tag{6.66}$$

where α_{1j} and α_{2j} are the ON expansion coefficients of $f_1(\sigma)$ and $f_2(\sigma)$, respectively. Clearly in the ON expansions, we should ideally have for two signal components

$$\bar{c}_{pq} = \delta_{pq} \tag{6.67}$$

In the numerical quadratures, the condition (6.67) is automatically satisfied otherwise eq. (6.7) will not be true. In the method of ℓsf, one can calculate a "modified" correlation coefficient $\bar{c}_{pq}$ between two signal components from the elements of matrix $A_s^T A_s$. Now, the elements of $A_s^T A_s$ are

$$J_{pq} = \sum_{\ell=1}^{M} \frac{1}{\Delta\sigma_\ell^2} IX_p IX_q \tag{6.68}$$

Eq. (6.68) is another kind of "inner product" which is different from either the one in (2.9) or (6.28) due to the presence of the area element $\Delta\sigma_\ell$. To calculate $\bar{c}_{pq}$ according to the definition of inner product in (6.68), we follow (6.66) and define a "correlation coefficient-like" quantity

$$\bar{c}'_{pq} = \frac{J_{pq}}{(J_{pp}J_{qq})^{1/2}} \tag{6.69}$$

For two X_p and X_q, there is no guarantee that $\bar{c}'_{pq} = \delta_{pq}$. But the departure of $\bar{c}'_{pq}$ from δ_{pq} will be small if the blocksize $\Delta\sigma_\ell$ is small as can be verified by performing some limiting process in (6.68) and taking into account the orthonormality of X_p. Therefore by using the method of ℓsf, we still can expect statistically independent signal components $\alpha_p X_p(\theta, \lambda)$ if $\Delta\sigma_\ell$ is small.

A different interpretation of correlation coefficients arises if we can remove the signal contribution from "degrees" $\mu + 1$ to ∞ and hence e_ℓ in (6.26) is regarded as purely "noise". In this case, the problem becomes a "minimum variance" estimation problem if we still impose the condition in (6.27) and possibly consider the "weights" of the observations. The estimation formula for X_s will remain the same except that a weight matrix could be introduced. In general, the estimation formula in this case is

$$X_s = (A_s^T P A_s)^{-1}(A_s^T P L_s) \tag{6.70}$$

where P is a weight matrix associated with the data. Let $N_s = A_s^T P A_s$ and $Q_s = N_s^{-1}$. Mikhail (1976, p. 301) recommended two quantities for the definition of correlation:

(1) $$\rho_{pq} = \frac{-n_{pq}}{(n_{pp}n_{qq})^{1/2}} \quad , \; n_{pq} \text{ is an element of } N_s \, . \tag{6.71}$$

(2) $$\hat{\rho}_{pq} = \frac{\omega_{pq}}{(\omega_{pp}\omega_{qq})^{1/2}} \quad , \; \omega_{pq} \text{ is an element of } Q_s \, . \tag{6.72}$$

As noted by Mikhail (ibid.), $\hat{\rho}_{pq}$ is generally larger than ρ_{pq}. In the usual least-squares adjustment problem, the definition of the correlation coefficient according to (6.72) is widely adopted since ω_{pp}'s are closely related to the accuracy estimates of parameters. Nevertheless, the partial "correlation coefficient" ρ_{pq} (defined by Mikhail, 1976) has exactly the same definition as $\bar{c}'_{pq}$ in (6.69) except the sign convention. Therefore, the partial "correlation coefficient" ρ_{pq} is applicable to either the case where e_ℓ in (6.26) is regarded as "truncation error + noise" or to the case where e_ℓ is purely noise. However, in accordance with the classical definition of correlation coefficients of least-squares adjustment problems in geodesy, the definition (6.72) will be the only acceptable choice for the definition of correlation coefficient as we present the numerical results later in this chapter.

Based on the above discussion, we can see that the connection between the correlation coefficient defined in (6.66) and the correlation coefficient defined in the parameter estimation theory occurs as soon as discrete data are used. The discussion here is mainly to distinguish the two kinds of correlation coefficients arising from two different disciplines (the key is the assumption of "ergodicity" implicit in formula (6.66)). One thing we have to emphasize is that in the approximation theories such as those in Davis (1975), Rivlin (1981), Gaier (1987), Sansone (1959), Tolstov (1976), etc., the approximation accuracy is a major issue, the statistical properties such as the accuracy of the coefficients are not even mentioned. However, by proper manipulations of the given function such as removing the "truncation error" or "high frequency content", we may treat the approximation problem as a parameter estimation problem provided that discrete data are given. In such a case the coefficients of expansion become "parameters" and all the statistical concepts such as correlations, accuracies, etc. can be applicable to the estimated coefficients. Using such a statistical approach, obviously the ON coefficients are still nearly uncorrelated due to the way we construct the orthonormal functions (see Section 5.1, 5.2, 5.3) and the forms in (6.68) and (6.69). On the other hand, the spherical harmonic coefficients are expected to be highly correlated due to the non-orthonormality of spherical harmonic functions provided that an incomplete set of data is given on a sphere.

Although a spherical harmonic expansion for oceanic data (no land value is involved) cannot provide independent signal components, Engelis (1987b, p. 10) found that up to a sufficiently high degree the corresponding expansion can reproduce the original data. The good fit in case of using high degree expansion is attributable to the completeness and closedness of non-orthonormal systems of functions which build a so-called "Schauder basis" for the space of continuous functions (Schaffrin et al., 1977, p. 151). (For a proof of the completeness and closedness of non-orthonormal systems, see Smirnov and Lebedev, 1968, pp. 235-237). A further discussion on the relationship between the spherical harmonic expansion and the ON function expansion will be made in Section 6.3.

6.2.4 Harmonic Synthesis

Given the expansion coefficients α_j or α_{nm}, β_{nm} of the ON functions for a signal $f(\theta, \lambda)$ on the oceans, we now seek a way to perform the harmonic synthesis to a "degree" below the maximum "degree" available. We shall still take advantage of the relationship between the ON functions and the spherical harmonics. Now the problem is:

Given: α_j, $j = 0, 1, \ldots, \mu$

or

α_{nm}, β_{nm}, $n = 0, 1, \ldots, Nmax$, $m = 0, \ldots, n$

Find:

$$f_\eta(\theta, \lambda) = \sum_{j=0}^{\eta} \alpha_j X_j(\theta, \lambda) \qquad , \eta \le \mu$$

or

$$f_K(\theta, \lambda) = \sum_{n=0}^{K} \sum_{m=0}^{n} \left(\alpha_{nm} O_{nm}(\theta, \lambda) + \beta_{nm} Q_{nm}(\theta, \lambda) \right) \quad , \quad K \le Nmax$$

For system $\{X_j(\theta, \lambda)\}$, the relationship between μ and Nmax is $\mu = (Nmax + 1)^2 - 1$ (note that j starts from 0). It turns out that the discussion using the one-index form of ON functions is easier to handle. The transform between the one-index form and the double-index form for system $\{X_j(\theta, \lambda)\}$ can be found in (5.73); for system $\{Y_j(\theta, \lambda)\}$, the transform is found in (5.75); for system $\{Z_j(\theta, \lambda)\}$, the transform is found in (5.76).

We express the expansion in a vector-product form as:

$$f_\eta(\theta, \lambda) = [X_0 \quad X_1 \cdots X_\eta] \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_\eta \end{bmatrix} \tag{6.73}$$

Recalling the relationship in (6.35), we have

$$f_\eta(\theta, \lambda) = [L_0 \quad L_1 \cdots L_\eta] \begin{bmatrix} c_{00} & c_{10} & c_{20} & \cdots & c_{\eta 0} \\ & c_{11} & c_{21} & \cdots & c_{\eta 1} \\ & & \ddots & & \vdots \\ 0 & & & & c_{\eta\eta} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_\eta \end{bmatrix}$$

$$= [L_0 \quad L_1 \cdots L_\eta] \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_\eta \end{bmatrix}$$

$$= \sum_{j=0}^{\eta} a_j L_j \tag{6.74}$$

where a_j are defined as "pseudo coefficients" and can be found by

$$a_j = \sum_{p=j}^{\eta} c_{jp} \alpha_p \tag{6.75}$$

The a_j values can be interpreted as intermediate coefficients for the synthesis process. An interpretation of a_j is that they are just the spherical harmonic coefficients due to the form in (6.74). (Remember that L_j are surface spherical harmonics). The second interpretation is true from computational point of view. We must emphasize that the signal $f(\theta, \lambda)$ is assumed to exist only over the oceans. The expansion in (6.74) will basically allow us to compute $f(\theta, \lambda)$ anywhere on a sphere. However, any signals outside the oceans computed by (6.74) will be meaningless since they do not exist and the ON functions are defined only over the oceans.

The form in (6.75) thus provides a convenient way to perform harmonic synthesis using existing package such as SSYNTH in (Colombo, 1981). In all the discussions made so far in this chapter, one can see the importance of the combination coefficients c_{jp}, without which the ON function expansions are impossible. However, to get these combination coefficients, the complex computations discussed in Chapter 5 must be made. In later discussions on the numerical results, we will see how much we can benefit from the analysis and use of the ON function expansion.

It is necessary to point out that (6.75) is equivalent to (6.38) where we also interpret Y_S as an intermediate vector containing the spherical harmonic coefficients. In the following section we will further address the relationship between the ON function expansion and the spherical harmonic expansion using a minimum error norm principle. Important distinctions between them, especially in the context of spectral analysis, will be also made.

6.3 Spherical Harmonic Expansion and ON Function Expansion

Since the ON functions are constructed from spherical harmonic functions, the expansions by using these two kinds of basis functions possess certain relationships. First of all, we still can approximate the signal on the oceans by the spherical harmonics, although they do not form an orthonormal system on the oceans. Such an approximation, or expansion can be written as (note: the domain is always oceans)

$$\begin{aligned}\tilde{f}(\theta, \lambda) &= \sum_{j=0}^{\mu} a_j L_j(\theta, \lambda) \\ &= \sum_{n=0}^{Nmax} \sum_{m=0}^{n} \left(a_{nm}\bar{R}_{nm}(\theta, \lambda) + b_{nm}\bar{S}_{nm}(\theta, \lambda)\right)\end{aligned} \tag{6.76}$$

where $\hat{f}(\theta, \lambda)$ is an approximation to $f(\theta, \lambda)$. Now we still can demand that the error norm be minimum, namely

$$\begin{aligned}\varphi &= (f - \hat{f}, f - \hat{f}) \\ &= \frac{1}{A}\iint_{\sigma} (f - \hat{f})^2 d\sigma \\ &= \frac{1}{A}\iint_{\sigma} \left[f - \sum_{j=0}^{\mu} a_j L_j(\theta, \lambda)\right]^2 d\sigma = \text{a minimum}\end{aligned} \tag{6.77}$$

where σ is the oceans, and A is the area of the oceans. A necessary condition for (6.77) is that

$$\frac{\partial \varphi}{\partial a_j} = 0 \quad , \quad \text{for } j = 0, \ldots, \mu \tag{6.78}$$

By (6.78) we obtain a set of equations:

$$\begin{aligned}
&(L_0, L_0)a_0 + (L_0, L_1)a_1 + \ldots + (L_0, L_\mu)a_\mu = (f, L_0) \\
&(L_1, L_0)a_0 + (L_1, L_1)a_1 + \ldots + (L_1, L_\mu)a_\mu = (f, L_1) \\
&\qquad\vdots \\
&(L_\mu, L_0)a_0 + (L_\mu, L_1)a_1 + \ldots + (L_\mu, L_\mu)a_\mu = (f, L_\mu)
\end{aligned} \tag{6.79}$$

Written in a matrix form, (6.79) becomes the normal equations:

$$\begin{bmatrix} (L_0, L_0)(L_0, L_1)\cdots(L_0, L_\mu) \\ (L_1, L_0)(L_1, L_1)\cdots(L_1, L_\mu) \\ \vdots \\ (L_\mu, L_0)(L_\mu, L_1)\cdots(L_\mu, L_\mu) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_\mu \end{bmatrix} = \begin{bmatrix} (f, L_0) \\ (f, L_1) \\ \vdots \\ (f, L_\mu) \end{bmatrix} \tag{6.80}$$

from which the spherical harmonic coefficients a_j can be solved. Recalling the definition of Gram matrix of spherical harmonics in (5.21) and denoting

$$\overline{Y}_s = (a_0\ a_1 \cdots a_\mu)^T \tag{6.81}$$

$$\overline{L}_s = \big((f, L_0)(f, L_1) \cdots (f, L_\mu)\big)^T \tag{6.82}$$

we can write (6.80) as

$$G\overline{Y}_s = \overline{L}_s \tag{6.83}$$

Now, according to (3.27), we have

$$G = C^{-1}(C^{-1})^T = C^{-1}(C^T)^{-1} \tag{6.84}$$

Therefore, by substituting (6.84) into (6.83), we get

$$(C^{-1})^T \overline{Y}_s = C\overline{L}_s \tag{6.85}$$

It is not difficult to see that (cf. (5.72)):

$$C\,\overline{L}_s = \begin{bmatrix} c_{00} & & \\ c_{10} & c_{10} & 0 \\ \vdots & & \ddots \\ c_{\mu 0} & c_{\mu 1} & \cdots\ c_{\mu\mu} \end{bmatrix} \begin{bmatrix} (f, L_0) \\ (f, L_1) \\ \vdots \\ (f, L_\mu) \end{bmatrix}$$

$$= \begin{bmatrix} (f, X_0) \\ (f, X_1) \\ \vdots \\ (f, X_\mu) \end{bmatrix} \tag{6.86}$$

Denoting $\overline{X}_s = (C^{-1})^T \overline{Y}_s = (\alpha_0\ \alpha_1 \cdots \alpha_\mu)^T$, we have

$$\overline{X}_s = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_\mu \end{bmatrix} = \begin{bmatrix} (f, X_0) \\ (f, X_1) \\ \vdots \\ (f, X_\mu) \end{bmatrix} \tag{6.87}$$

The matrix form in (6.87) is precisely the ON expansion expressed in (6.7) which has been approximated by the numerical quadratures formula. Therefore, by the minimum error norm principle, the spherical harmonic expansion and the ON function expansion are equivalent in the sense that both functions can approximate a function on the oceans equally well. However, there are some important distinctions between them in both theory and application:

(1) It is not justified to perform spectral analyses using the surface spherical harmonics on the oceans, since under ergodicity they are not statistically independent there (see (6.65)). In reality, such a practice has been done by numerous researchers (e.g., Engelis, 1987b, Nerem, 1989). As shown later, the spectral components (i.e., coefficients) will look "reasonable" under either of the following conditions:

a. The maximum degree of expansion is below approximately 12 or 13 (using the ℓsf method and the oceanic data only). For, in these low degree expansions, "reasonable" amplitudes of coefficients can be obtained. Normally we claim that the coefficients are reasonable only by comparing the amplitude of the expanded function and the amplitudes of the coefficients, since there is no theoretical justification for performing spectral analysis for ocean data using functions which are not orthogonal over the oceans. These coefficients are just the result of minimum error fit, as was done in (6.80). Now, if the maximum degree of expansion exceeds a certain limit and hence the amplitudes of coefficients become unreasonably large, then comes the second condition:

b. An a-priori power rule of spherical harmonic coefficients is used for estimating the spherical harmonic coefficients. Such a power rule implies that the amplitudes of coefficients diminish as the degree goes higher. An example is given by Nerem et al. (1990, eq. (20)):

$$v_\ell = 0.19\ell^{-1.27} \quad \text{meters} \tag{6.88}$$

which they derived from the Levitus SST with zero values on land (obviously they treated the Levitus SST as a global signal). In (6.88), ℓ is the degree of the spherical harmonic and v_ℓ is defined as (ibid., eqn. (19)):

$$v_\ell^2 = \sum_{m=0}^{\ell} \frac{a_{\ell m}^2 + b_{\ell m}^2}{2\ell + 1} \tag{6.89}$$

To use the power rule, we can change the minimum error criterion from (6.77) to (now the degree "ℓ" is associated with "j")

$$\bar{\phi} = (f - \hat{f}, f - \hat{f}) + \sum_{j=0}^{\mu} a_j^2 / v_j^2 = \text{a minimum} \tag{6.90}$$

Then by (6.78) and following the derivation for (6.83) we get

$$\bar{Y}_s' = (G + P)^{-1} \bar{L}_s \tag{6.91}$$

where P is a diagonal matrix containing $1/v_j^2$, namely,

$$P = \begin{bmatrix} \frac{1}{v_1^2} & & & 0 \\ & \frac{1}{v_2^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{v_\mu^2} \end{bmatrix} \tag{6.92}$$

The new solution $\bar{Y}_s'$ in (6.91) is thus different from $\bar{Y}_s$ in (6.83) due to the presence of P matrix.

In practice, we may only have mean values for the signal f on the oceans, then (6.76) becomes

$$\bar{f}_\ell(\theta, \lambda) + e_\ell = \frac{1}{\Delta\sigma_\ell} \sum_{j=0}^{\mu} a_j \iint_{\Delta\sigma_\ell} L_j d\sigma = \frac{1}{\Delta\sigma_\ell} \sum_{j=0}^{\mu} a_j \mathbb{L}_j^\ell \tag{6.93}$$

where $\bar{f}_\ell$ is the mean value at block ℓ, and e_ℓ is the error due to the finite terms used. If we change the definition of inner product to

$$\left.\begin{aligned}(L_p, L_q) &= \sum_{\ell=1}^{M} \frac{1}{\Delta\sigma_\ell^2} \mathrm{I}L_p^\ell \mathrm{I}L_q^\ell \\ (f, L_p) &= \sum_{\ell=1}^{M} \frac{1}{\Delta\sigma_\ell} \mathrm{I}L_p^\ell \bar{f}_\ell\end{aligned}\right\} \qquad (6.94)$$

then all the above derivations and formulae are valid for the solutions of the spherical harmonic coefficients using mean values.

Now the problem of using a power rule in the estimation of the coefficients of a spherical harmonic expansion is as follows: a power rule with diminishing amplitude such as the one in (6.88) is normally valid for global signals. When it is applied to an oceanic signal such as SST, the convergence of a spherical harmonic expansion is slow due to the fact that the oceanic signal is artificially forced to follow the spectral behavior of a global signal. This means that to achieve the same approximation accuracy the expansion with a power rule will require more harmonic terms than the expansion without a power rule. As demonstrated later, the expansion to degree 10 without a power rule will approximate the Levitus SST as well as the expansion to degree 24 with a power rule. The power rule method will also create Gibbs' phenomenon (the poor approximation near the discontinuities of the approximated function) at the continental boundary for SST if the maximum degree is not sufficiently high.

In the author's opinion, the power rule, which is used to stabilize the solution for certain problems, should be used in a compatible domain. For example, Kaula's rule (or a modified version) may be used to constrain the behavior of geopotential coefficients since the earth's gravity field is a global signal, see, for example, Nerem et al. (1990, eq. (22)), Marsh et al. (1989), etc.

Even if a power rule is used for a spherical harmonic expansion and "reasonable" coefficients are obtained, misleading results could be obtained from the spectral analyses (for example, the study of degree variances, error degree variances at various degrees). Indeed, it is possible that, due to the use of power rule, the correlation coefficients from the inverse of normal matrix may be small and the spherical harmonic coefficients are "almost" statistically independent.

(2) The ON function expansion, of course, allows us to perform spectral analysis on the oceans. The expansion will always result in coefficients with decreasing magnitude. If the ON coefficients are to be estimated in a model where other parameters are also present, then we may use a power rule of ON coefficients to stabilize the solution over the oceans. When using a power rule of ON coefficients, we are basically dealing with a compatible domain and hence a faster convergence of the expansion is expected.

(3) The ON function expansion possesses the property of permanence (Davis, 1975, p. 173), while the spherical harmonic expansion does not. By permanence we mean that the expansion coefficients that have been determined will not be affected by adding more terms in the expansion. For the ON function expansion, this is clear from the use of the expansion formula $\alpha_j = (f, X_j)$, which shows that α_j is determined independently. However, by adding one more element in the spherical harmonic expansion, we need to

add one more row and column in the normal matrix in (6.80), thus by inverting such a new normal matrix the whole solution will be changed.

6.4 Numerical Experiments and Results Using the Levitus SST in the Expansions

In this section, we shall describe some experiments and results related to the ON function expansions of the Levitus SST data specified in Section 6.1. We shall carry out the experiments using the three ON systems that are described in Chapter 5 and this current chapter, namely, $\{X_j\}$, $\{Y_j\}$ and $\{Z_j\}$. Although most of the formulae in this chapter are developed for system $\{X_j\}$, it is rather simple to extend these formulae to systems $\{Y_j\}$ and $\{Z_j\}$, since all that is required is the change of combination coefficients in the existing formulae bearing in mind that for the latter two systems the lowest j values of system $\{L_j\}$ from which the two systems are constructed are 1 and 4, respectively. If the data used are centered, namely, the mean value over the oceans is zero, then system $\{Y_j\}$ is equivalent to system $\{X_j\}$. As stated in Section 6.1, the Levitus SST has been centered by Engelis (1987b) so that we shall obtain the same results from using systems $\{X_j\}$ and $\{Y_j\}$. For comparisons, the spherical harmonic expansions will be made in some cases.

Due to a large number of experiments that have been done, only selected results are presented. The important implications of the results will be stressed in Section 6.4.2.

6.4.1 Some Definitions

In order to avoid the use of lengthy names for the solutions, we introduce the following abbreviations:

- Onnqtoxx :
The ON function expansion using the numerical quadratures. "xx" is the maximum degree of the ON functions in a expansion. The formula used is (6.10).

- Onℓsftoxx :
The ON function expansion using the least squares fit. "xx" is defined as above. The formula used is (6.34).

- Shnqtoxx :
The spherical harmonic expansion using the numerical quadratures. This is only possible when the land value of SST are assumed to be zero (see also Engelis, 1987b). The formula used is (6.17).

- Shℓsftoxx:
The spherical harmonic expansion using the least squares fit. The accuracy of approximation to the data should be as good as that of onℓsftoxx, as proved in the early sections. The "spectral content" in this expansion will be different from the one in onℓsftoxx. The formula used is (6.39). Many comments will be made on the spectral content (e.g. degree variance, etc.).

- Shℓsfptoxx :
The spherical harmonic expansion using the least squares fit with a power rule obtained by using the degree variances of the coefficients in solution shnqtoxx. The formula used is (6.91) with the inner products defined by (6.94).

In addition, the following definitions are introduced for later references:

- Degree variance of the spherical harmonic coefficients:

$$\sigma_n^2 = \sum_{m=0}^{n} (a_{nm}^2 + b_{nm}^2) \tag{6.95}$$

where a_{nm} and b_{nm} are defined in (6.76). Attention must be paid to the distinction between σ_n^2 in (6.95) and τ_n^2 in (6.22).

- Percentage of energy up to degree N: (for the ON functions only!)

$$E_N = Q_f(N)/P_{avg} \tag{6.96}$$

where P_{avg} and $Q_f(N)$ are defined in (6.2) and (6.3), respectively. $Q_f(N)$ can be regarded as the cumulative average power (it is "average" due to the area of the oceans A in (6.2)). The E_N is "absolute" not relative percentage because P_{avg} is derived from the original signal.

- Geostrophic currents: (Officer, 1974, eqs. (4.70) - (4.73))

$$v_x = \frac{-g}{2R\omega\sin\phi} \frac{\partial\zeta(\phi, \lambda)}{\partial\phi} \tag{6.97}$$

$$v_y = \frac{g}{R\omega\sin 2\phi} \frac{\partial\zeta(\phi, \lambda)}{\partial\lambda}, \quad \phi = 90^\circ - \theta \tag{6.98}$$

where v_x and v_y are the velocity components of the geostrophic currents along east and north directions, g the mean gravity, R the mean radius of the earth, and ω the rotational velocity of the earth. The quantity $\zeta(\phi, \lambda)$ is the SST from any of the expansions. The physics behind the current (v_x, v_y) is that, when water particles are in constant motion, the Coriolis force (or the geostrophic acceleration) arises. In such a case, the sea surface is not a level surface so that a force due to the horizontal pressure gradient is created to balance the Coriolis force, resulting in eqs. (6.97) and (6.98). To calculate the geostrophic currents from the ON function expansions, we first transform the ON coefficients to the spherical harmonic coefficients using (6.75). Then the computation can be made using the existing package that deals with the spherical harmonics.

6.4.2 Results

In Table 6.3, we have summarized the experiments (for the Levitus SST) that have been done and the fits of the expansions to the original data. The abbreviations in Table 6.3 are defined in the previous section. The selected contour plots of SST from the expansions are given at the end of this chapter.

From Table 6.3, we see that based on the RMS fit to the original data the ON function expansions and the spherical harmonic expansions have the same degree of approximation to the data. The fit of the expansions to the original data improves as the

maximum degree (Nmax) of the ON functions or the spherical harmonic function increases. Above degree 15 of the ON functions, systems $\{X_j\}$, $\{Y_j\}$ and $\{Z_j\}$ produce essentially the same approximation accuracy if the Nmax used is the same.

For all the systems (including the spherical harmonics), the expansions with Nmax $\leq$ 15 yield unacceptable approximation accuracies. In particular, the Gibbs' phenomena prevail at the continental boundaries. For example, in Figure 6.8 and Figure 6.10 the discrepancies are fairly large in the areas of Gulf Stream (especially at the east coast of USA), the Antarctic circumpolar current and the Kuroshio current. Using the method of least squares fit without the use of a power rule (note: the power rule is applied only in solution shℓsfpto24), the agreement between the original SST and the SST from the expansions is on the order of 2 cm when Nmax = 24. For all the expansions, the maximum discrepancies occurred in the Mediterranean Sea and the Black Sea, which are basically two "isolated" regions of the domain of the oceans. If we exclude these two regions, we can reduce the discrepancy to about 1 cm when Nmax = 24 and the ℓsf method without the use of the power rule.

When using a power rule in the ℓsf method for a spherical harmonic expansion, as we expected the approximation accuracy has been degraded. This case is shown in Figure 6.22. As compared to Figure 6.12 where the discrepancies correspond to a spherical harmonic expansion to Nmax = 24, Figure 6.22 shows very large discrepancies at the continental boundaries even if the Nmax is still 24. Note that in Figure 6.12 we have used the ON function expansion which is equivalent to the spherical harmonic expansion if the ℓsf method is used. Using the RMS difference criterion, we may say that solution shℓsfpto24 is equivalent to solution onℓsfto10 which is shown in Figure 6.7. Using the numerical quadratures for the spherical expansion with SST = 0 on land, the approximation accuracy is even worse than that obtained from the ℓsf method with a power rule, as shown in Figure 6.24 (assuming that Nmax is the same in both expansions). As we explained earlier in this chapter, the use of a power rule in the spherical harmonic expansion is equivalent to forcing the expansion coefficients in the ℓsf method to follow the behavior of the expansion coefficients as obtained by assuming SST = 0 on land so that solution shℓsfpto24 yields poorer results than solution shℓsfto24, but better results than solution shnqto24.

So why do we use a power rule for the spherical harmonic expansion in the ℓsf method if it yields degraded results? The most important reason is to get "reasonable" degree variances, or, amplitudes of the spherical harmonic coefficients. This can be explained by the use of Figures 6.4 and 6.5. In Figure 6.4, the maximum square root of degree variance of the spherical harmonic coefficient is about 2.90 meters (at degree 13) which is impossible for a low-amplitude signal such as SST. Now, in Figure 6.5 we have successfully "reduced" the amplitudes of the spherical harmonic coefficients. Of course, the accuracy of approximation has been sacrificed in such a practice. Engelis (1987b, p. 7) has attributed the cause of the large amplitudes of spherical harmonic coefficients to the folding of frequencies from high degree terms to low frequency terms when an incomplete data set is used.

Table 6.3
Various Expansions of the Levitus SST and the Comparisons with the Original Data

Expansions	Mean diff.	RMS diff.	Max. diff.	Figures
onℓsfto10, $\{X_j\}$	-0.001	0.067	0.393	6.7, 6.8, 6.13
onℓsfto15, $\{X_j\}$	-0.001	0.049	0.281	6.9, 6.10
onℓsfto24, $\{X_j\}$	-0.001	0.021	0.217	6.11, 6.12, 6.14
onnqto10, $\{X_j\}$	0.001	0.066	0.406	
onnqto24, $\{X_j\}$	-0.001	0.025	0.440	
onℓsfto10, $\{Y_j\}$	-0.001	0.067	0.394	
onℓsfto15, $\{Y_j\}$	-0.001	0.045	0.280	
onℓsfto24, $\{Y_j\}$	-0.001	0.021	0.216	
onℓsfto10, $\{Z_j\}$	-0.013	0.083	1.080	6.15, 6.16, 6.19
onℓsfto15, $\{Z_j\}$	-0.003	0.049	0.328	
onℓsfto24, $\{Z_j\}^{\dagger}$	-0.001	0.021	0.216	6.17, 6.18, 6.20
shℓsfto10	-0.001	0.067	0.393	
shℓsfto15	-0.001	0.049	0.281	
shℓsfto24	-0.001	0.021	0.217	
shℓsfpto24$^{\Delta}$	-0.015	0.069	0.723	6.21, 6.22
shnqto24*	-0.001	0.110	0.718	6.23, 6.24

unit = meters
† : Best; * : worst; $^{\Delta}$: power rule

On the other hand, the ON functions provide a way to approximate accurately the SST without worrying about the excessively large amplitudes of the expansion coefficients. This is shown through the degree variances in Figure 6.4 where the property of permanence of the ON function expansion (see Section 6.3) can be seen, since the expansion coefficients are not affected by the use of different Nmax values. For example, for solution onℓsfto24, Nmax = 24; for solution onℓsfto10, Nmax = 10, but the coefficients before degree = 10 are the same for both solutions. The property of permanence (over the oceans) does not hold for the spherical harmonics since the coefficients vary as the Nmax values are changed. One can also see that the term a_{oo} of the spherical expansions is not zero for all the expansions, which contradicts the fact that the SST data have been centered. On the contrary, the ON function expansions using system $\{X_j\}$ always yield a zero α_{oo}, where by definition α_{oo} is the mean of the SST over the oceans.

The ON function expansion, like the Fourier series expansion, enable us to study the energy distribution of SST. Table 6.4 shows the cumulative average powers and percentages of energy from the expansions using system $\{X_j\}$. Again, we confirm from Table 6.4 that there is no energy at degree 0 (of the ON functions). Up to degree 10, the percentage of energy has achieved 98.52 (by the ℓsf method); to degree 24, the corresponding value is 99.90. For system $\{Y_j\}$, the result is the same. For system $\{Z_j\}$, the energy distribution is shown in Table 6.5 where only selected degrees are presented. The figure of merit (see Section 6.1) is 0.1 up to degree 24, for all ON function expansions using the ℓsf method. Therefore, by using Nmax = 24 in the ON function expansions, we can recover 99.90% of the energy of SST with a RMS fit of 2 cm (including the Mediterranean Sea and the Black Sea). Based on the energy distribution using the ON

function expansion and Theorem 2 in Section 5.4, we conclude that the Levitus SST has a resolution of 750 km (= 180/24 x 100 km).

It is clear that we cannot perform a spectral analysis over the oceans using the spherical harmonics. However, if we assume that SST = 0 on land, the SST can be artificially regarded as a global signal. By the land value assumption, we have expanded the SST into the spherical harmonics for Nmax = 36 by the numerical quadratures. To investigate the energy distribution, it will be more appropriate to use the cumulative total power:

$$P_{T_N} = 4\pi \sum_{n=0}^{N} \sigma_n^2 \tag{6.99}$$

Then the cumulative percentage of energy is P_{T_N}/P_T, where P_T is the total power of SST defined in (6.1). It was found that the cumulative percentages of energy are 91.73, 94.26 up to N = 24 and N = 36, respectively. This indicates a relatively slow convergence in such an expansion. The land value assumption also leads to a RMS fit of 11 cm when Nmax = 24—an approximation accuracy poorer than that obtained from the ON expansion with Nmax = 10.

It must be pointed out that there is no need for a land value assumption in the ON function expansion (This is also true for the spherical harmonic expansion if the ℓsf method is used). In addition, a power rule is not required to control the behavior of the ON coefficients and yet the expansion still yields a rapid convergence and excellent spectral behavior of the expansion coefficients. We can say that the spherical harmonics can have fairly good performance in accurately approximating an oceanic signal but fail to yield good spectral behavior due to the obvious theoretical problem pointed out in Section 6.2.3.

The next issue to be discussed is the correlation between expansion coefficients and the accuracy estimates of the coefficients. In the usual least-squares adjustment in geodetic problems, the mathematical model is exact. For example, if (x_1, y_1, z_1) and (x_2, y_2, z_2) are rectangular coordinates of two points on the earth, then the distance d between the two points is $d = ((x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2)^{1/2}$ which describes an "exact" relationship between d and the coordinates. Then, we may use the measurement d, together with some other measurements such as angles, to determine the coordinates. The measurement d cannot be perfect so we have a noise v due to factors such as instruments, etc. In such a case, the observation equation is

$$d + v = f(x_1, y_1, z_1, x_2, y_2, z_2) \tag{6.100}$$

Now to approximate a function by a linear combination of other functions, the concept is somewhat different. In (6.77), we have talked about the approximation "error", but not the noise which is basically random. The "error" is caused by the insufficient terms of the approximating functions used, but not the noise. If the data such as SST really have noises, we should re-write (6.76) as (in case of approximation by the ON functions)

$$\zeta_{ob} + e + v = \sum_{n=0}^{Nmax} \sum_{m=0}^{n} \left(\alpha_{nm} O_{nm}(\theta, \lambda) + \beta_{nm} Q_{nm}(\theta, \lambda)\right) \tag{6.101}$$

Table 6.4
Cumulative Average Powers in Meter**2 and Percentages of Energy for Solutions
(1) onℓsfto24, $\{X_j\}$ (2) onnqto24, $\{X_j\}$

degree	cumulative ave. power (1)	(2)	percentage (1)	(2)
0	0.0000	0.0000	0.00	0.00
1	0.0922	0.0925	31.95	32.05
2	0.2305	0.2309	79.83	79.98
3	0.2533	0.2534	87.73	87.74
4	0.2706	0.2695	93.71	93.33
5	0.2729	0.2717	94.51	94.11
6	0.2761	0.2753	95.63	95.36
7	0.2780	0.2773	96.27	96.06
8	0.2821	0.2816	97.70	97.52
9	0.2837	0.2833	98.27	98.12
10	0.2845	0.2841	98.52	98.39
11	0.2849	0.2845	98.67	98.54
12	0.2855	0.2851	98.87	98.73
13	0.2861	0.2857	99.10	98.94
14	0.2867	0.2861	99.28	99.10
15	0.2869	0.2864	99.35	99.19
16	0.2872	0.2867	99.47	99.29
17	0.2874	0.2869	99.54	99.38
18	0.2877	0.2872	99.63	99.46
19	0.2879	0.2874	99.71	99.53
20	0.2881	0.2876	99.77	99.59
21	0.2882	0.2877	99.80	99.64
22	0.2883	0.2878	99.84	99.67
23	0.2884	0.2879	99.87	99.72
24	0.2885	0.2881	99.90	99.76

Table 6.5
Cumulative Average Powers, Percentage of Energy and Figure of Merit for Solution onℓsfto24, $\{Z_j\}$

Degree	Cum. ave. power	Percent. energy	Figure of merit*
10	0.2856	98.92	1.08%
15	0.2870	99.40	0.60%
24	0.2884	99.90	0.10%

* see eq. (6.4)

where ζ_{ob} is the observation of the signal, e the approximation error or the truncated signal and v is the noise of the observation. Unless Nmax goes to infinity or ζ is a bandlimited signal, the approximation error will never disappear from (6.101). When e exists, we are minimizing the norm of (e + v), instead of v alone, to get a solution for the coefficients α_{nm} and β_{nm}. In such a procedure, we cannot claim that the values from the diagonal elements of the inverted normal matrix be the "standard deviations" of the coefficients, due to the mixed effect of e and v (as one could find from Figure 6.8 that e can be very large when Nmax = 10). A trick to avoid this dilemma is made by Rapp (1989b) who removed the error e by assuming that the truncated signal from Nmax + 1 to ∞ is available from some model and the observation equation is reformulated by using the "modified" signal (ζ - e). (Note that Rapp (1989b) dealt with gravity anomalies and geopotential coefficients, but his principle is applicable in this discussion).

In short, the problem in estimating the accuracies of the expansion coefficients is that the truncated signal exists and we cannot get an exact model such as (6.100). Therefore, we may just talk about the correlation coefficients according to the discussions made in Section 6.2.3. Engelis (1987b) has found the phenomenon that the correlations between the spherical harmonic coefficients are high if Nmax is greater than 10 and only oceanic data are used. The results from the above expansions agree with his findings. However, we found that by the *ℓ*sf method the correlations between the ON coefficients are negligible regardless of the Nmax values used, and these low correlations also explain why in Figure 6.4 the two sets of ON coefficients agree very well.

Next we shall discuss the unitary transformation between system $\{Y_j\}$ and $\{Z_j\}$ (see also Section 5.3). Although system $\{Z_j\}$ is constructed from $\bar{R}_{20}, \bar{R}_{21}, \bar{S}_{21}, \bar{R}_{22}, \bar{S}_{22}, \cdots$ (without $\bar{R}_{00}, \bar{R}_{10}, \bar{R}_{11}$ and $\bar{S}_{11}$ terms), it approximates the SST as well as system $\{Y_j\}$ does provided that Nmax > 15 (below 15 $\{Z_j\}$ has poorer performance than $\{Y_j\}$, see Figure 6.16). If we prefer a SST expansion using system $\{Y_j\}$, we can perform a unitary transformation from the coefficients of system $\{Z_j\}$ to the coefficients of system $\{Y_j\}$ using (5.82). As we mentioned in Section 5.3, the maximum expansion degree (the N value in (5.82)) must theoretically be infinite in order to make the unitary transformation possible. Since with Nmax = 24, the expansion using system $\{Z_j\}$ can recover 99.9% of the SST energy, we have tried this limited expansion for our unitary transformation. Let us denote the transformed coefficients as α'_{nm} and β'_{nm} (clearly now they become the expansion coefficients of system $\{Y_j\}$), and the coefficients from the direct SST expansion using system $\{Y_j\}$ as $\bar{\alpha}_{nm}$ and $\bar{\beta}_{nm}$, the RMS difference by degree is

$$D_n = \left[\sum_{m=0}^{n} \left(\bar{\alpha}_{nm} - \bar{\alpha}'_{nm} \right)^2 + \left(\bar{\beta}_{nm} - \bar{\beta}'_{nm} \right)^2 \right]^{1/2} \qquad (6.102)$$

Up to degree 15, the D_n values are shown in Table 6.6. Clearly, the difference between the two sets of coefficients is negligible. Thus for a signal such as the SST used in this study, it is possible to perform a unitary transformation between two ON systems using a finite terms in (5.78). (Note that (5.78) is a "theoretical" one and (5.82) is the "practical" one used in this discussion).

As stated at the end of Section 5.3, we again emphasize the importance of the unitary transformation in the joint estimation of geoid and SST using satellite altimetry: we try to find an ON system for the SST in the joint model that can avoid the "sensitivity

spectrum" of the geoid and the strong 1 cycle/rev radial orbit error (cf. Wagner, 1986). After recovering the coefficients of this system from the joint solution, we then transform these coefficients to the coefficients corresponding to another system.

Table 6.6
RMS Differences D_n Between the Unitarily Transformed Coefficients and the Coefficients From the Direct SST Expansion Using System $\{Y_j\}$
Unit : meters

Degree (n)	D_n	Degree	D_n
1	0.0001	9	0.0002
2	0.0002	10	0.0002
3	0.0002	11	0.0003
4	0.0001	12	0.0003
5	0.0002	13	0.0002
6	0.0002	14	0.0003
7	0.0002	15	0.0003
8	0.0002		

6.4.3 Geostrophic Currents at the Continental Boundaries

Finally we have to settle the problem of the directions of current flow at the continental boundaries. The directions of geostrophic current are computed by (6.97) and (6.98). Apparently according to those equations the only required quantity is the approximated SST, assuming that everything else can be found in some standard formulae. Based on oceanographers' idea, "near continental coasts, the currents are forced to follow the coastlines, forming eastern and western boundary currents" (Neumann, 1968, p. 75). However, depending on the character of forces that act upon a water body, and on the state of balance of these forces, the resulting motions of water particles can be different (ibid., p. 127). In addition to geostrophic currents, other major types of currents are inertia currents, gradient currents, drift currents, etc. (ibid., Chapter 4). Thus the boundary currents could be a combination of various currents. So, using the geostrophic currents alone, can we demand that the current flows must follow the coastlines? In addition, Wunsch and Gaposchkin (1980, p. 729) pointed out that, due to the existence of alongshore pressure gradients on the coasts, the current flows must become significantly nongeostrophic in the immediate vicinity of the boundaries. They also quantitatively showed such a deviation on that page. Furthermore, even if the flows computed by (6.97) and (6.98) must follow the coastlines, the observations that can be used for developing a SST model may not be close enough (depending on the depths) to the coastlines (a typical case in satellite altimetry due to data editing) and hence the computed flows at the place where no observation is made really do not make sense.

Unfortunately, as shown in Figures 6.13, 6.14, 6.19 and 6.20, none of the expansions described in the previous section yield the current flows that everywhere follow the coastlines. To get a better insight into the condition of the current flows at the oceanic boundaries and possibly to make improvement of the representation, we first present the geometry of some needed unit vectors at a point of the oceanic boundaries in Figure 6.3.

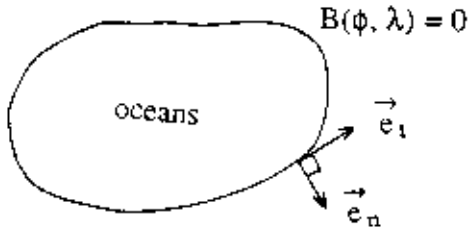


Figure 6.3 Unit Vectors at the Oceanic Boundaries

In Figure 6.3, $B(\phi, \lambda) = 0$ is an equation describing the oceanic boundaries, $\vec{e}_t$ is a tangent vector, $\vec{e}_n$ is an outer normal obtained by the cross-product $\vec{e}_t \times k$, where k is the unit vector along the radial direction. Applying differential geometry, we have

$$\vec{e}_t = (1, y') / \sqrt{1 + (y')^2}$$

$$\vec{e}_n = (y', -1) / \sqrt{1 + (y')^2} \tag{6.103}$$

where

$$y' = -\frac{1}{\cos\phi} \frac{\partial B / \partial \lambda}{\partial B / \partial \phi} \tag{6.104}$$

Now, if the flow vector (v_x, v_y) is to be parallel to the tangent vector, then we should have

$$\vec{e}_n \cdot (v_x, v_y) = 0 \tag{6.105}$$

Substituting v_x and v_y in (6.97) and (6.98) and $\vec{e}_n$ in (6.103) into (6.105), we get

$$\frac{\partial \zeta}{\partial \phi}\frac{\partial B}{\partial \lambda} - \frac{\partial \zeta}{\partial \lambda}\frac{\partial B}{\partial \phi} = 0 \tag{6.106}$$

which is the desired constraint for the current flow. To incorporate this constraint to our least squares error approximation, we write (6.106) as

$$KX_S = 0 \tag{6.107}$$

where X_S contains the expansion coefficients. Assuming we use the ℓsf method and the ON functions, then the new approximation model is (in case that mean values are given, see also (6.26))

$$\left.\begin{aligned} & V_s^T V_s = (A_s X_s - L_s)^T (A_s X_s - L_s) = \text{minimum} \\ & \text{subject to} \\ & \quad K X_s = 0 \end{aligned}\right\} \tag{6.108}$$

where A_S, L_S are defined in (6.29) and (6.31), respectively. The solution of (6.108) can be found in (Uotila, 1986, p. 105):

$$\overline{X}_s = N_s^{-1} U_s - N_s^{-1} K^T (K N_s^{-1} K^T)^{-1} K N_s^{-1} U_s \tag{6.109}$$

where $N_S = A_s^T A_s$, $U_S = A_s^T L_s$. In (6.109), we have denoted the new solution vector as $\overline{X}_s$ to distinguish this new one from the solution vector without constraint, i.e., X_S. As in (Uotila), we can write (6.109) as

$$\begin{aligned} \overline{X}_s &= X_s - N_s^{-1} K^T (K N_s^{-1} K^T)^{-1} K X_s \\ &= X_s + \Delta X_s \end{aligned} \tag{6.110}$$

where ΔX_S denotes the change of X_S due to the constraint given in (6.107).

There are many problems in using the constraint in (6.107) to get the new solution, i.e., $\overline{X}_s$. First of all, since the boundaries of the oceans are not simply connected polygons (see Section 3.4.1), the flow vector constraints have to be specified locally, namely, we would need several equations $B_j(\phi, \lambda) = 0$, $j = 1, 2, \ldots$, constructed in local coastal areas by fitting polynomials to the (ϕ, λ) coordinates of shorelines, to accomplish the constraint equation (6.107). The construction of B_j will be a formidable job even if it is possible. Secondly, from (6.109), it is clear that the number of rows in matrix K (i.e., the number of constraints) cannot exceed the number of columns in matrix K (i.e., the number of expansion coefficients) or the matrix $(K N_s^{-1} K^T)$ will be singular. Therefore, only limited constraint equations are allowed to be established. If the total length of the world's shorelines is 10 times of the equator, then for Nmax = 10, we can at most establish a constraint equation every 3300 km along the shorelines. This sparse distribution of constraints, of course, cannot prevent the flow vectors from pointing to the continents.

Based on the above discussion, the constraint equation (6.107) could be hard to realize in reality. The second possibility of achieving the parallelism of flow vectors along the coastlines in the expansions of SST is to find a system of functions (not necessarily orthonormal) whose elements all satisfy the boundary constraint in (6.106) so that the resulting expansion gives a SST model from which the flow vectors follow the coastlines automatically. From the experience of eigenfunction and eigenvalue problems in Chapter 3, such a system probably does not exist.

6.5 Expansions of Oceanic Geoid in ON Functions

In this section, we shall perform expansions for the oceanic geoid obtained from the OSU89B potential coefficients (Rapp and Pavlis, 1990). We can do so if we regard the geoid undulation as a geometric quantity. Later on we shall also introduce an important technique to assess the error spectrum of the geoid undulation correction (to the satellite-

field implied undulation) from satellite altimetry based on this concept. By Bruns' formula, the geoid undulation N at a point (r, θ, λ) may be calculated by a set of disturbing potential coefficients $\bar{C}_{nm}$, $\bar{S}_{nm}$ as follows (Rapp, 1986):

$$N(r, \theta, \lambda) = \frac{GM}{\gamma r} \sum_{n=2}^{Nmax} \left(\frac{a}{r}\right)^n \sum_{m=0}^{n} \left(\bar{C}_{nm}\cos m\lambda + \bar{S}_{nm}\sin m\lambda\right) \bar{P}_n^m(\cos\theta) \tag{6.111}$$

where γ is the normal gravity and a is a scaling factor which is normally the equatorial radius. By using $r = a = R_e$ and assuming $\gamma = GM/R_e^2$, we get

$$N(\theta, \lambda) = R_e \sum_{n=2}^{Nmax} \sum_{m=0}^{n} \left(\bar{C}_{nm}\cos m\lambda + \bar{S}_{nm}\sin m\lambda\right) \bar{P}_n^m(\cos\theta) \tag{6.112}$$

By treating N as quantities like the SST, we can perform exactly the same experiments as before. Since most of the important points of the ON function expansions and the spherical harmonic expansions have been discussed in the previous experiments, we shall focus on a different issue here.

For the simulated data, we generate 1° x 1° mean undulations by Colombo's SSYNTH (Colombo, 1981) and use a modified formula of (6.112) for the mean value:

$$\bar{N}_{k\ell} = \frac{R_e}{\Delta\sigma_k} \sum_{n=2}^{Nmax} \sum_{m=0}^{n} \left(\bar{C}_{nm} IC_m^\ell + \bar{S}_{nm} IS_m^\ell\right) \overline{IP}_{nm} \tag{6.113}$$

where $\Delta\sigma_k$, IC_m^ℓ, etc. can be found in (6.64). Four sets of 1° x 1° mean undulations with Nmax = 15, 18, 20, 24 were generated. To perform the experiments, we use the oceanic mean undulations only, although (6.113) can give a global coverage. The issue to be discussed here is related to Theorem 2 in Section 5.4. That theorem implies that if now we expand the above oceanic geoids into the ON functions up to a maximum degree, then all the expansion coefficients with degree greater than Nmax will be zero. This is, however, not the case for the spherical harmonic expansion for the oceanic geoid.

Since it is clear that the three ON systems will yield identical approximation accuracy if the Nmax value exceeds a certain limit, we shall only use one system here, namely, system $\{X_j\}$. For all the ON function expansions and the spherical harmonic expansions, the Nmax value is 24 and the ℓsf method is used. The four sets of 1° x 1° oceanic geoids have the names as follows:

Set 1: Nmax = 15
Set 2: Nmax = 18
Set 3: Nmax = 20
Set 4: Nmax = 24

To show the validity of Theorem 2 in Section 5.4, we list the degree variances of the ON functions (τ_n^2 in (6.22)) and the degree variances of the spherical harmonics (σ_n^2 in (6.95)) from the expansions in Table 6.7.

Based on the results in Table 6.7, clearly Theorem 2 is numerically proved. For example, for Set 1, the Nyquist frequency of the oceanic geoid is 15; thus after degree 15 of the ON functions, the square root of the degree variances (of the ON function) immediately drop to zero. On the other hand, for Set 1, the spherical harmonic expansion still possesses some energy after degree 15 (of the spherical harmonic) . For the other sets, the same phenomenon can be found. Therefore by using the ON functions we can detect the maximum spatial resolution of an oceanic signal if it is bandlimited.

Table 6.7
Square Roots of Degree Variances of the ON Functions and the Spherical Harmonics
Using 4 Sets of 1° x 1° Mean Oceanic Geoids

Degree	Set 1		Set 2		Set 3		Set 4	
	ON	SH	ON	SH	ON	SH	ON	SH
0	0.181	0.735	0.280	0.722	0.303	0.707	0.349	0.976
10	1.382	2.240	1.373	2.233	1.345	2.249	1.316	2.345
14	0.642	1.076	0.660	1.081	0.651	1.090	0.648	1.960
15	0.404*	0.949	0.606	0.951	0.666	0.955	0.674	1.840
16	0.010	0.332	0.517	0.920	0.530	0.915	0.546	1.648
17	0.000	0.249	0.482	0.728	0.494	0.733	0.426	1.303
18	0.000	0.179	0.356*	0.764	0.352	0.755	0.287	1.232
19	0.000	0.123	0.011	0.117	0.358	0.648	0.416	1.074
20	0.000	0.077	0.006	0.071	0.247*	0.589	0.325	0.826
21	0.000	0.044	0.000	0.043	0.010	0.041	0.322	0.744
22	0.000	0.021	0.000	0.021	0.005	0.022	0.312	0.592
23	0.000	0.008	0.000	0.008	0.000	0.008	0.225	0.486
24	0.000	0.002	0.000	0.002	0.000	0.002	0.144*	0.425

unit : meters
ON : orthonormal function expansion
SH : spherical harmonic expansion
* : No energy beyond this degree of ON function (theoretically)

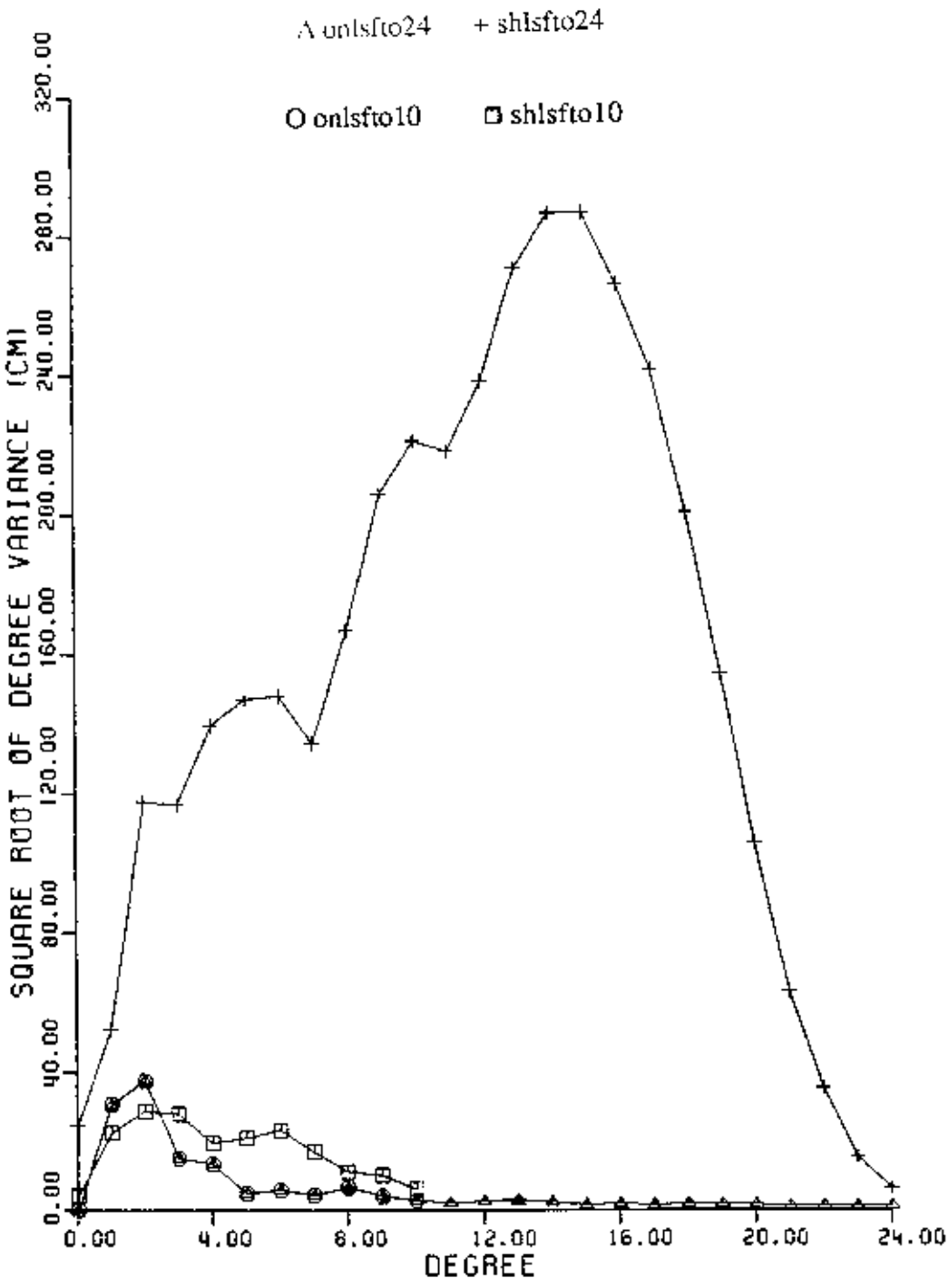


Figure 6.4 Comparisons of the Degree Variances of the ON Function Expansion and the Degree Variances of the Spherical Harmonic Expansion from Various Solutions.

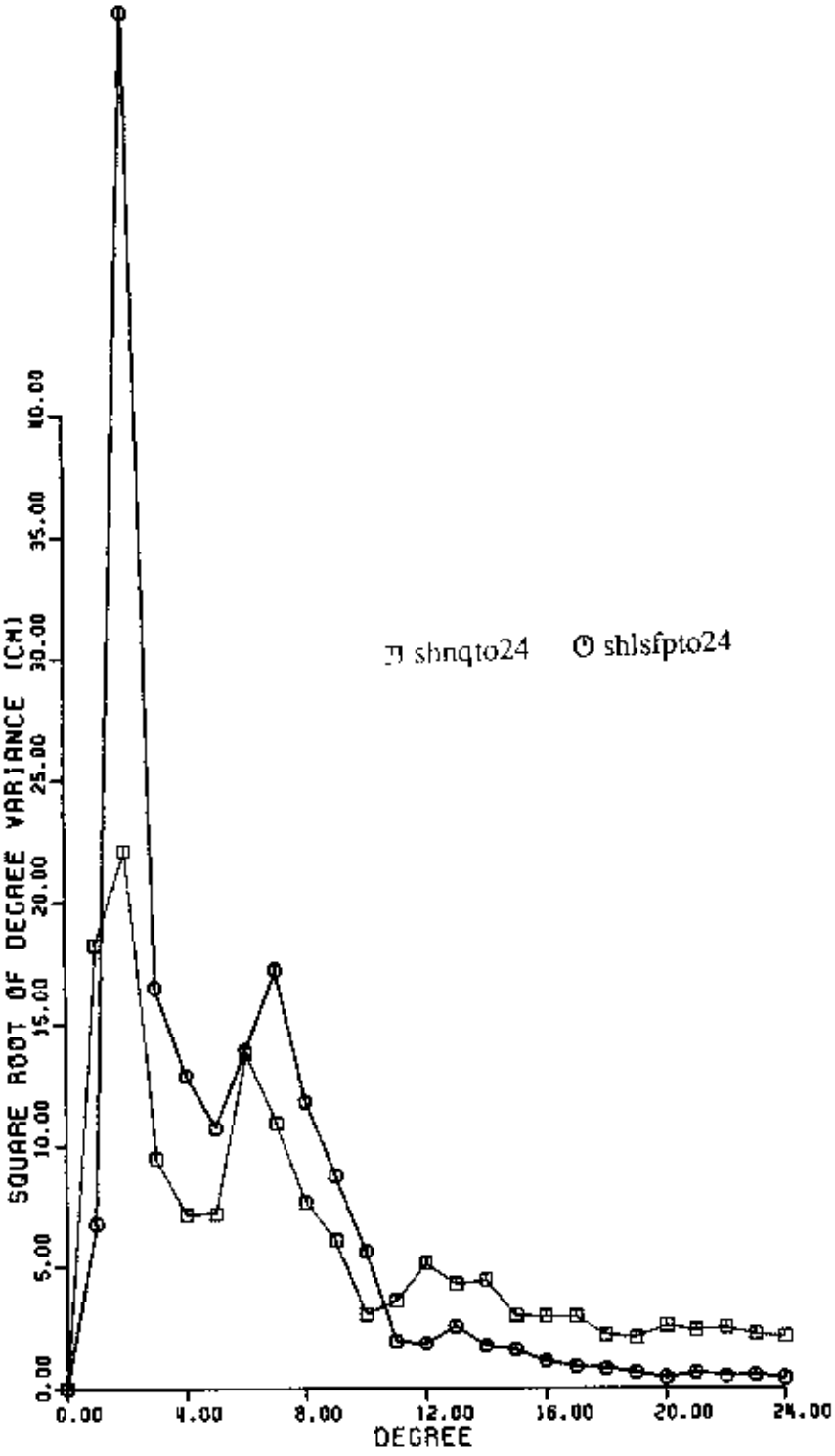


Figure 6.5 Degree Variances From Solutions shnqto24 and shlsfpto24.

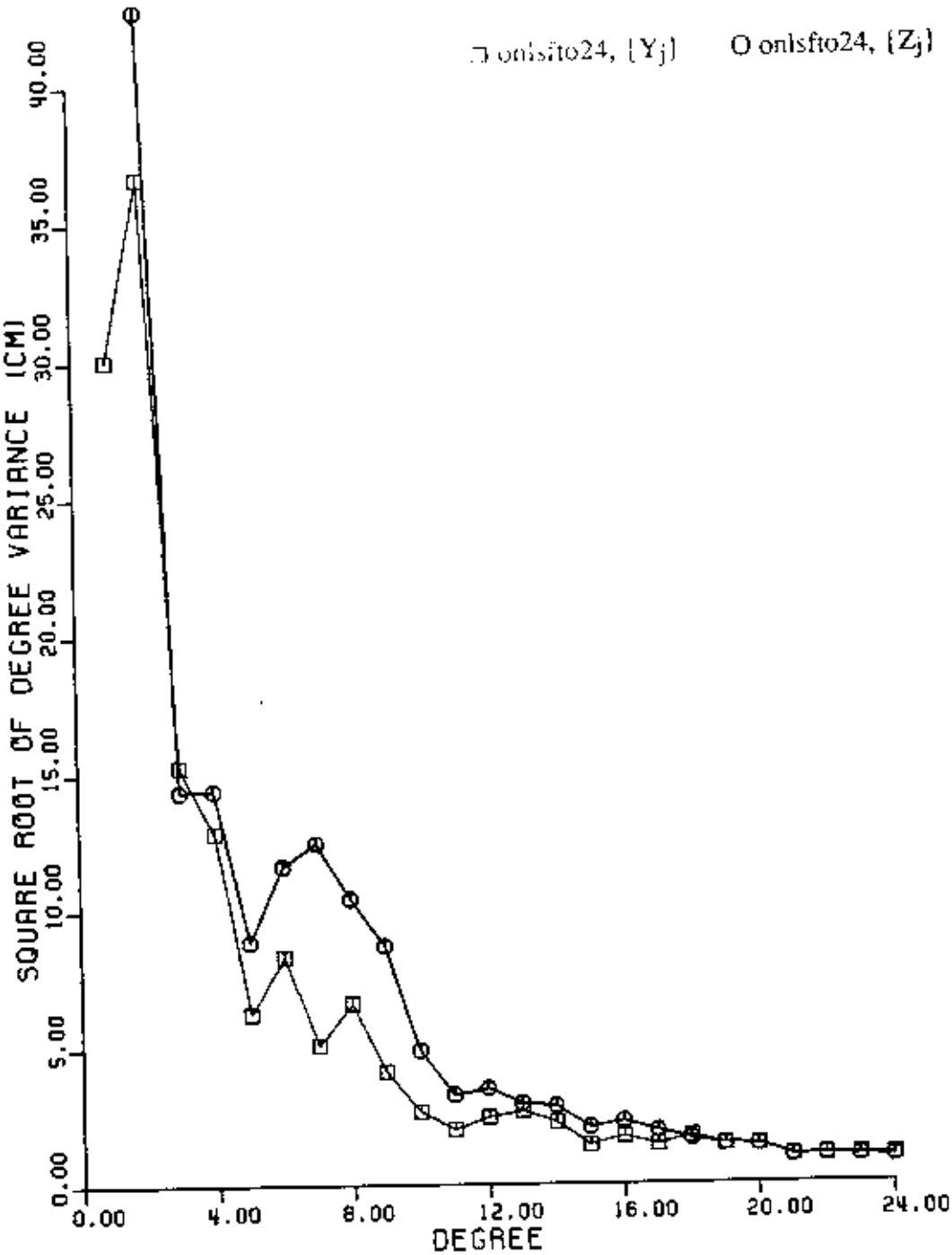


Figure 6.6 Degree Variances from Solutions onlsfto24, {Yj} and onlsfto24, {Zj}.

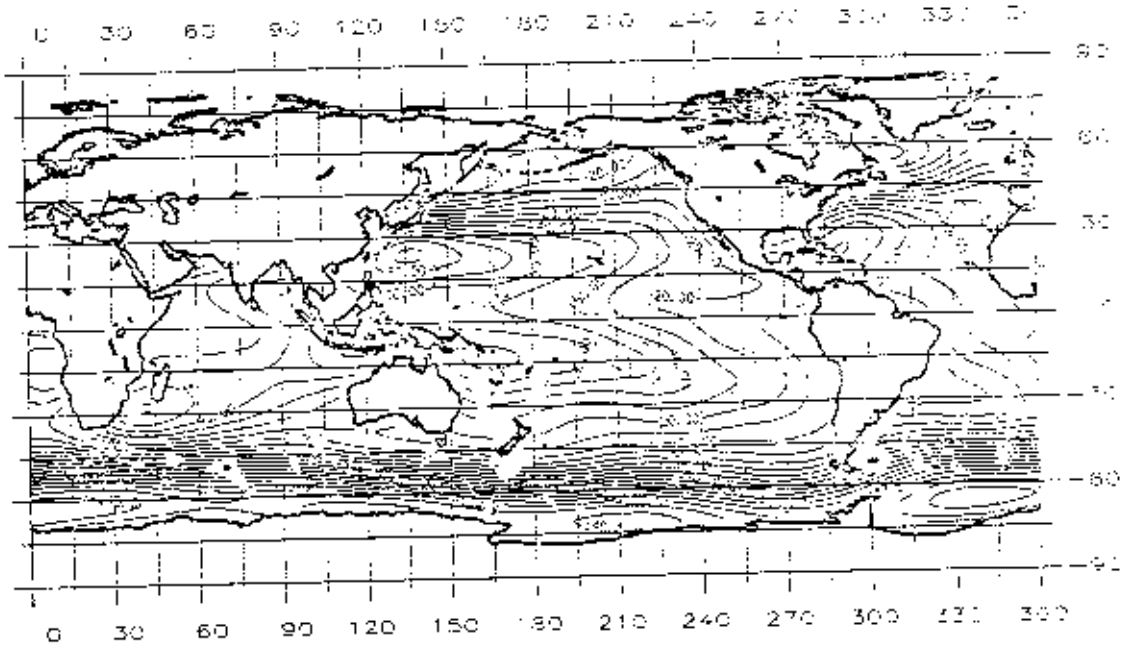


Figure 6.7 SST From Solution onlsfto10, $\{X_j\}$, CI = 10 cm.

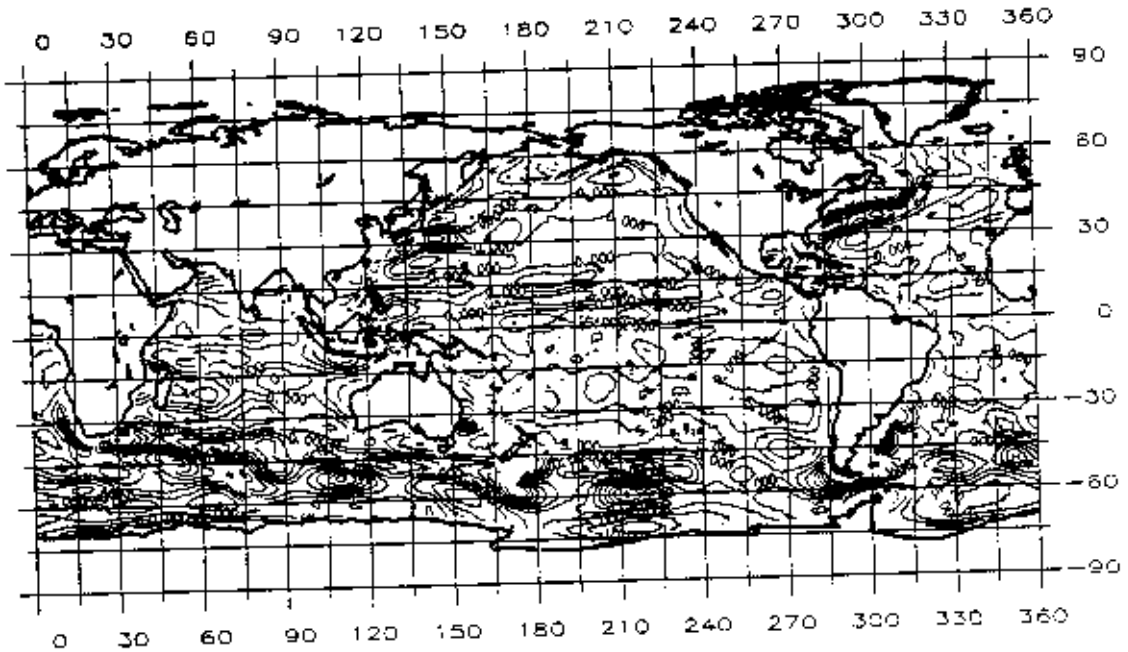


Figure 6.8 Difference Between the Levitus SST and SST From Solution onlsfto10, $\{X_j\}$, CI = 5 cm.

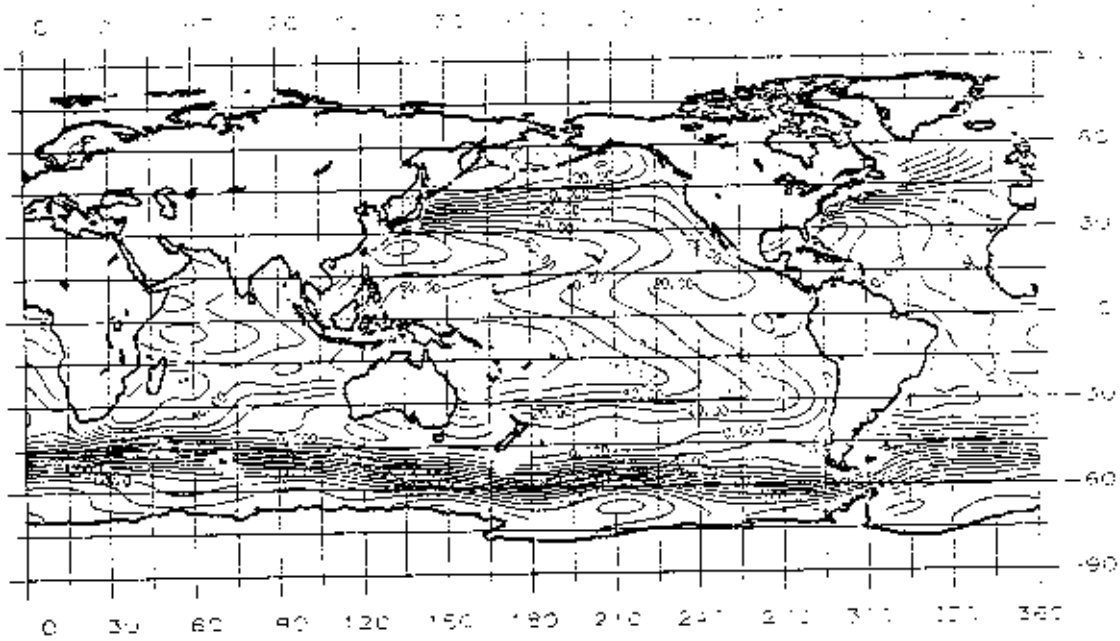


Figure 6.9 SST From Solution onlsfto15, $\{X_j\}$, CI = 10 cm.

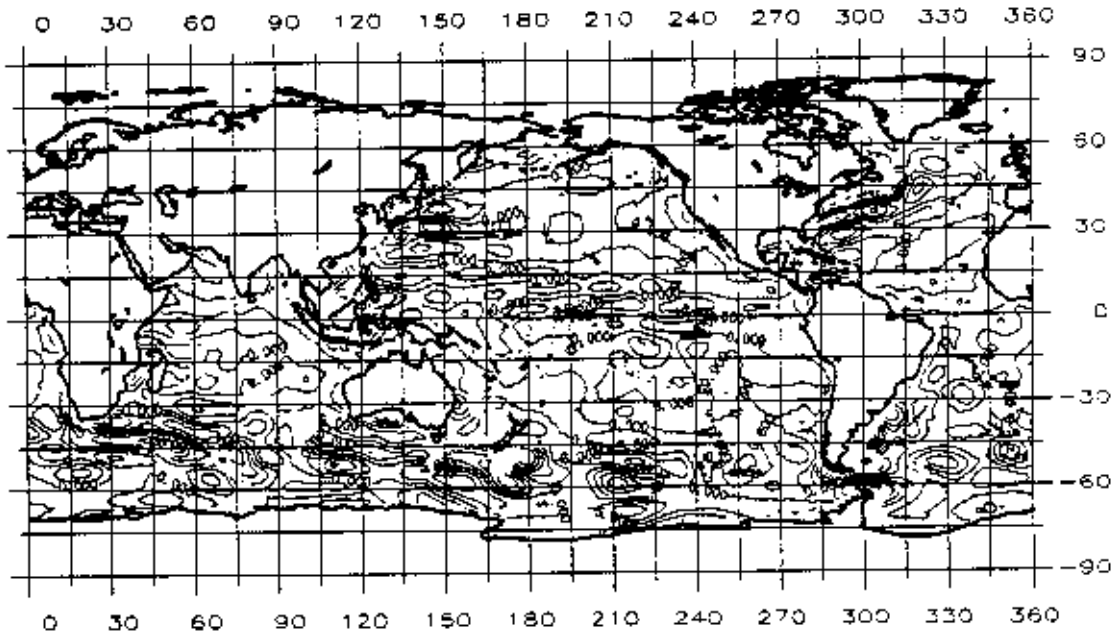


Figure 6.10 Difference Between the Levitus SST and SST From Solution onlsfto15, $\{X_j\}$, CI = 5 cm.

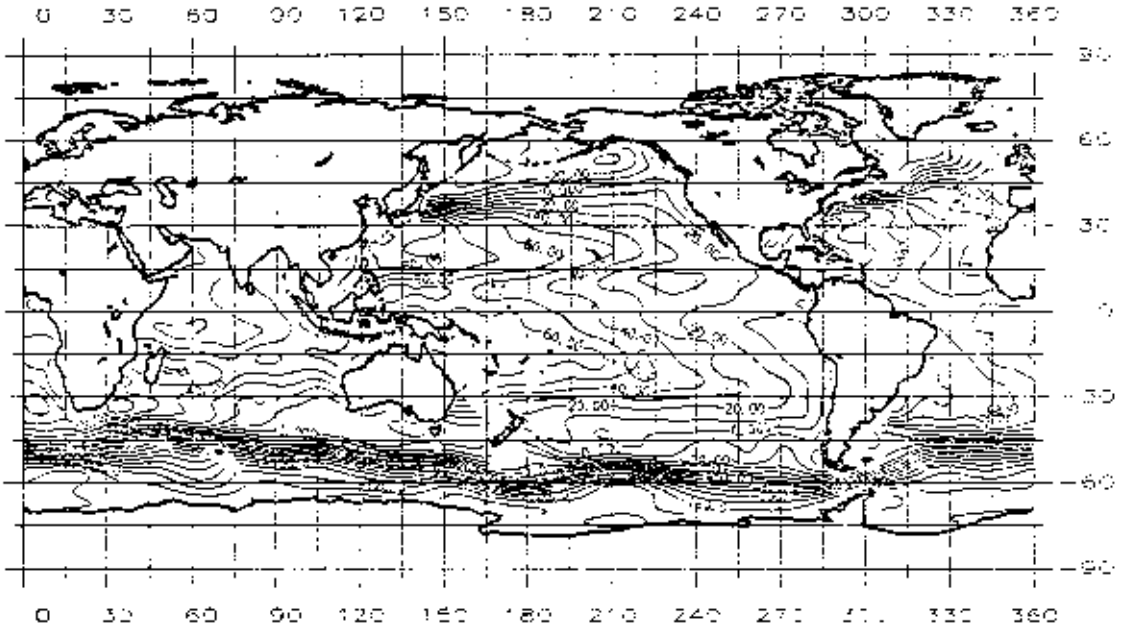


Figure 6.11 SST From Solution onlsfto24, $\{X_j\}$, CI = 10 cm.

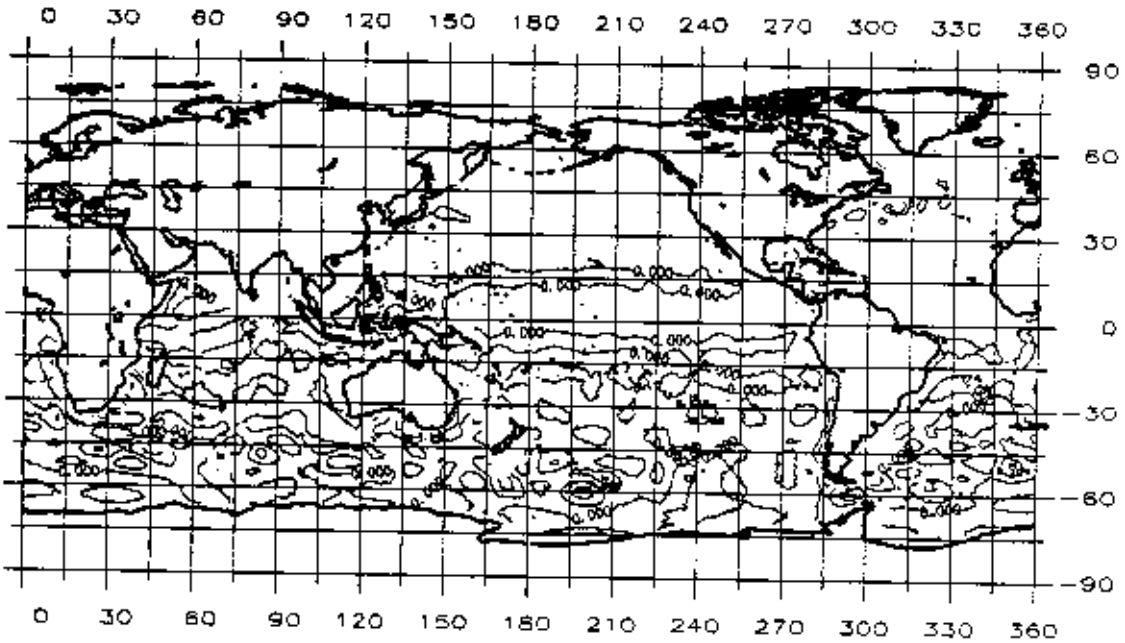


Figure 6.12 Difference Between the Levitus SST and SST From Solution onlsfto24, $\{X_j\}$, CI = 5 cm.

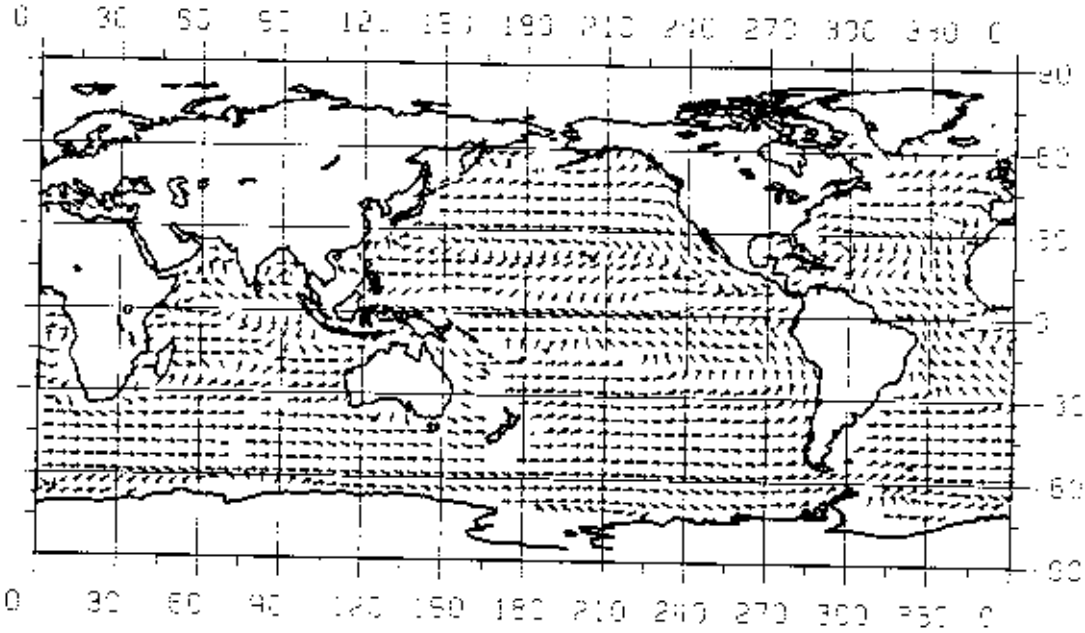


Figure 6.13 Geostrophic Currents From Solution onlsfto10, $\{X_j\}$.

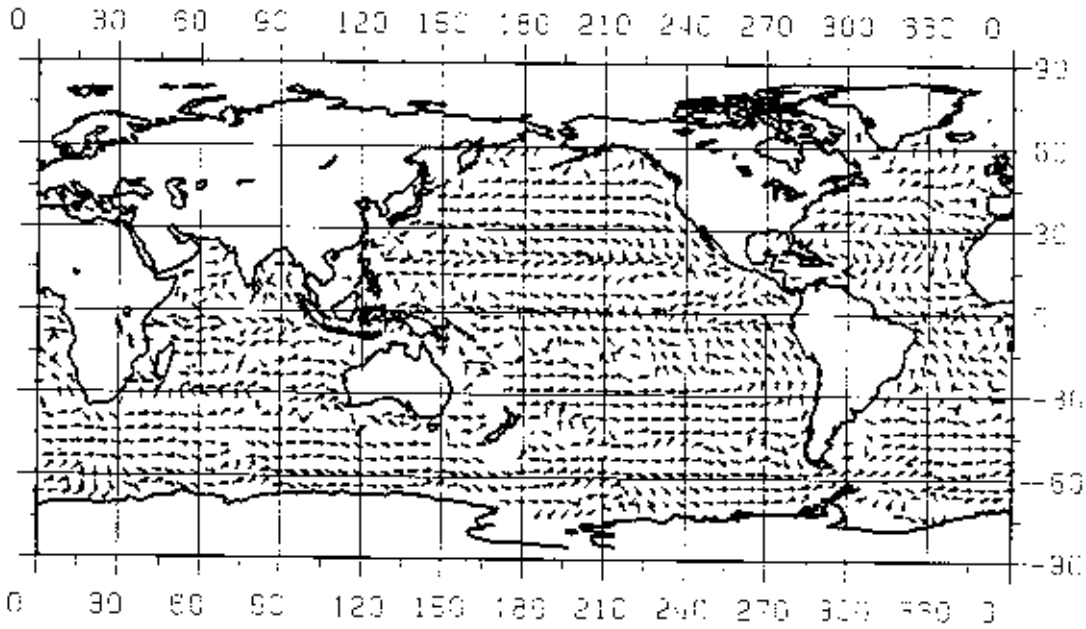


Figure 6.14 Geostrophic Currents From Solution onlsfto24.

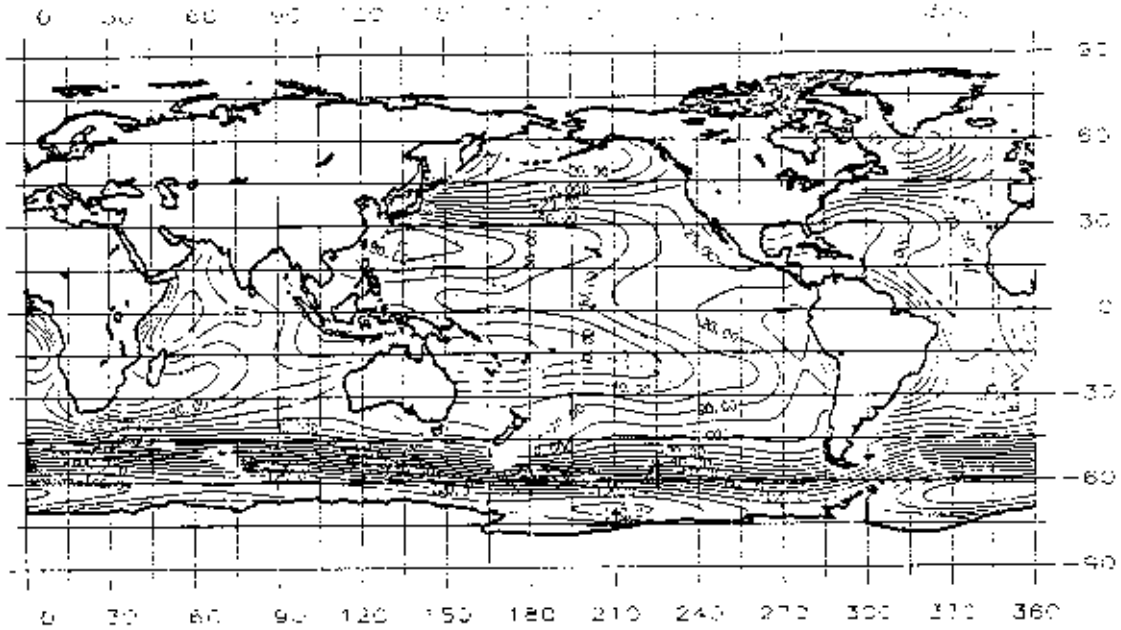


Figure 6.15 SST From Solution onlsfto10, {Z_j}, CI = 10 cm.

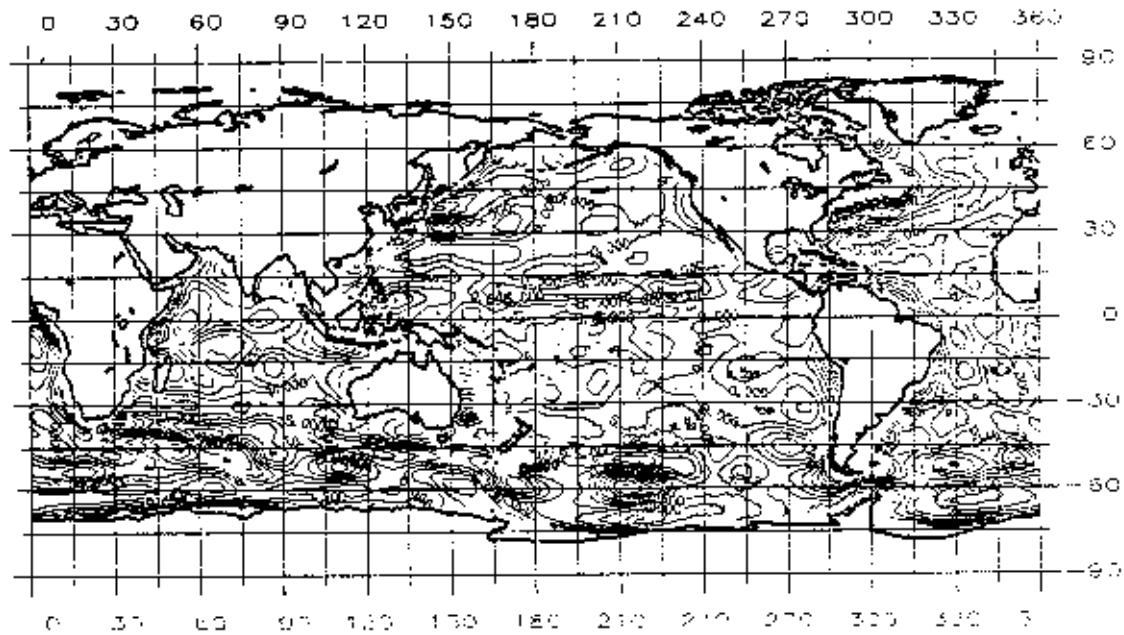


Figure 6.16 Difference Between the Levitus SST and SST From Solution onlsfto10, {Z_j}, CI = 5 cm.

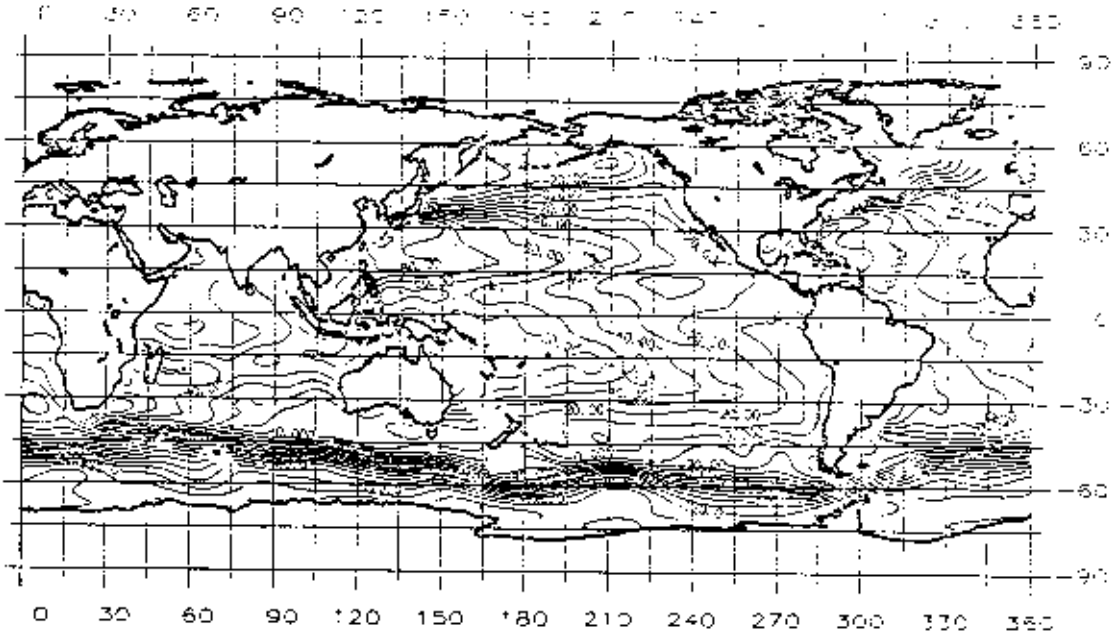


Figure 6.17 SST From Solution onlsfto24, $\{Z_j\}$, CI = 10 cm.

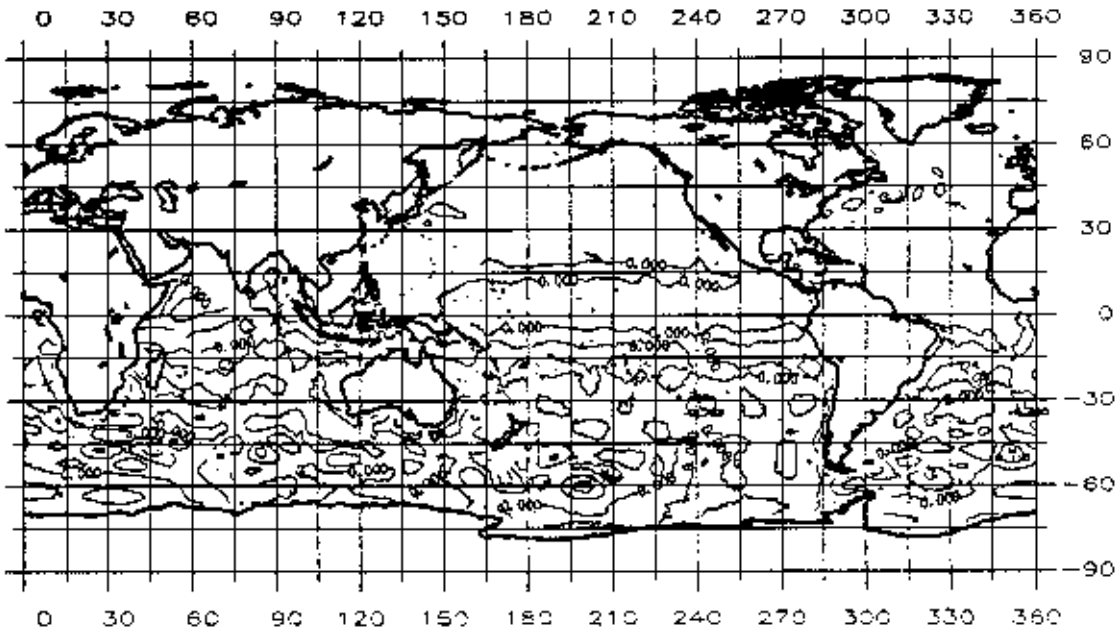


Figure 6.18 Difference Between the Levitus SST and SST From Solution onlsfto24, $\{Z_j\}$, CI = 5 cm.

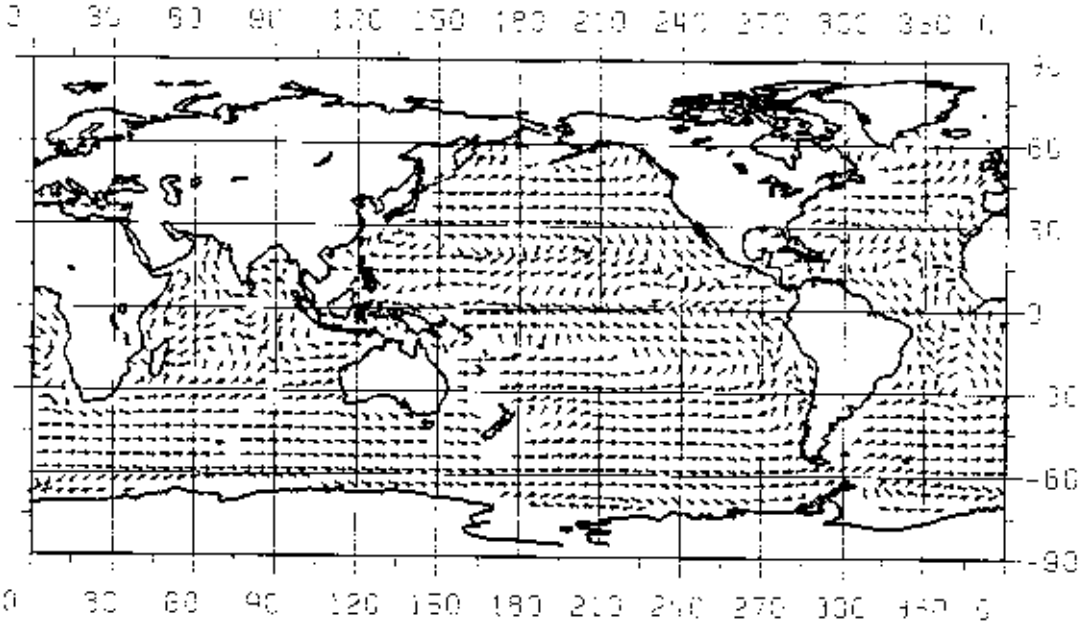


Figure 6.19 Geostrophic Currents From Solution onlsfto10, $\{Z_j\}$.

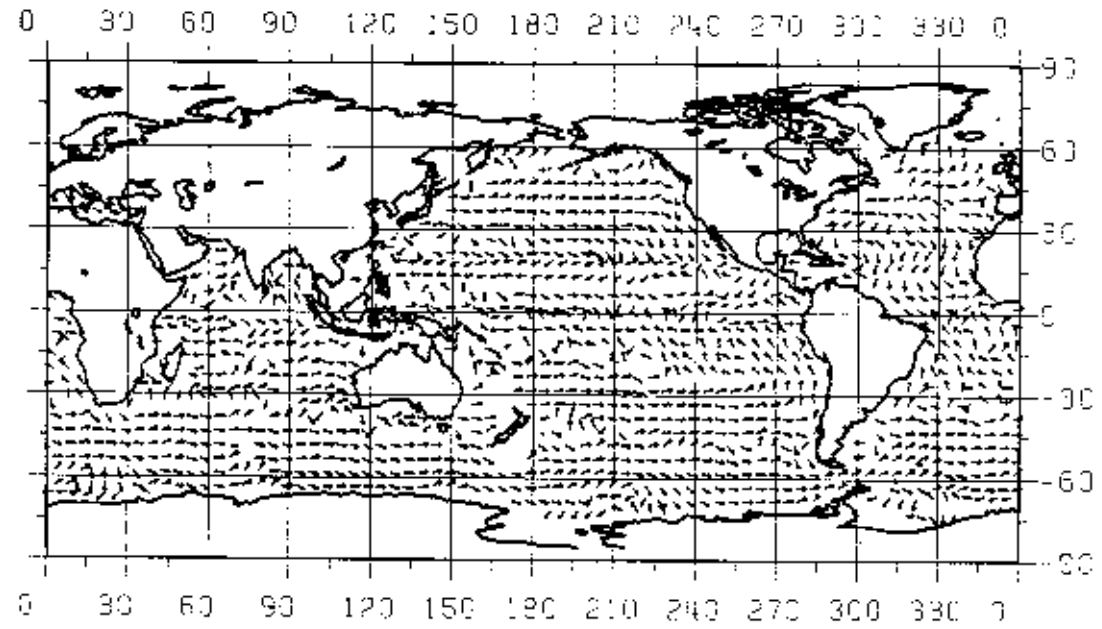


Figure 6.20 Geostrophic Currents From Solution onlsfto24, $\{Z_j\}$.

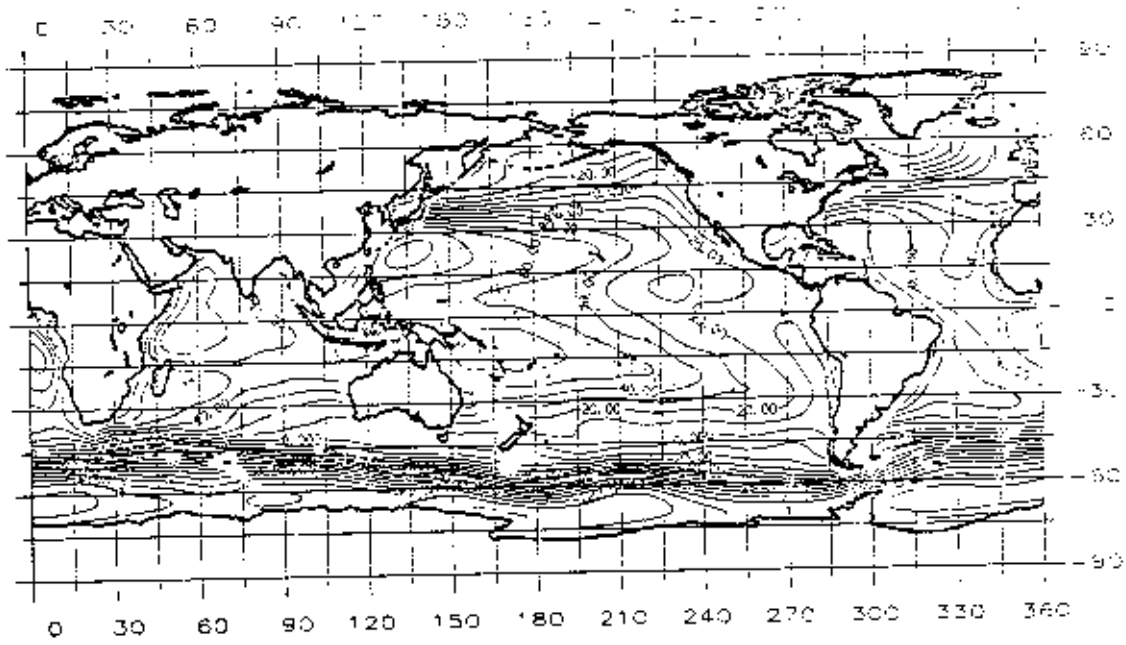


Figure 6.21 SST From Solution shlsfpto24, CI = 10 cm.

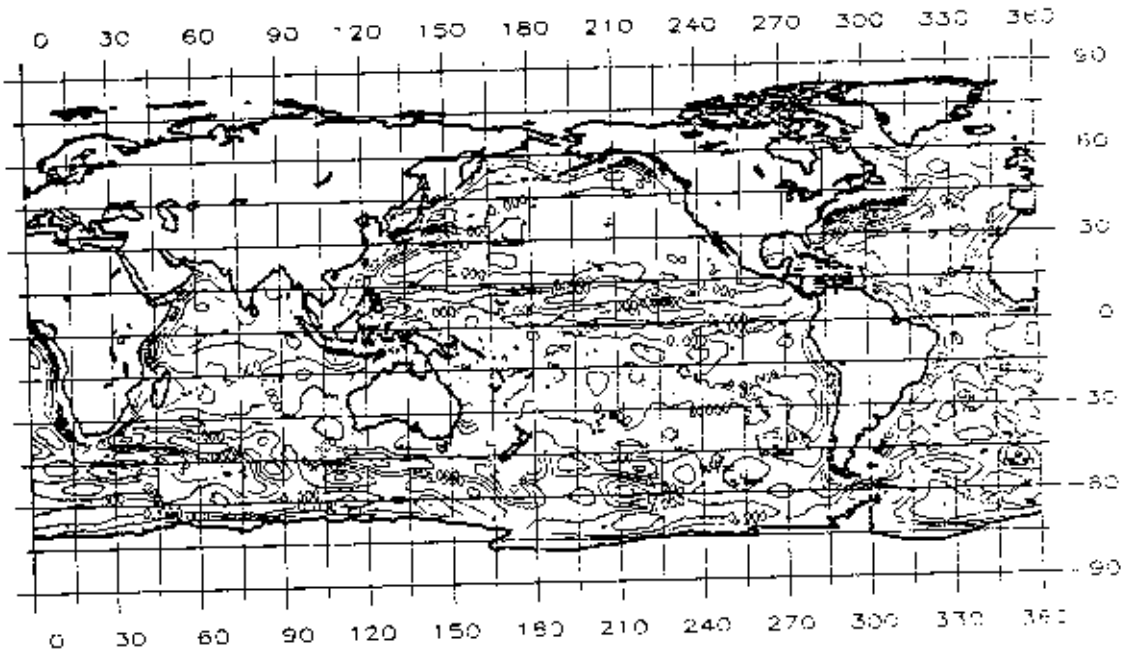


Figure 6.22 Difference Between the Levitus SST and SST From Solution shlsfpto24, CI = 5 cm.

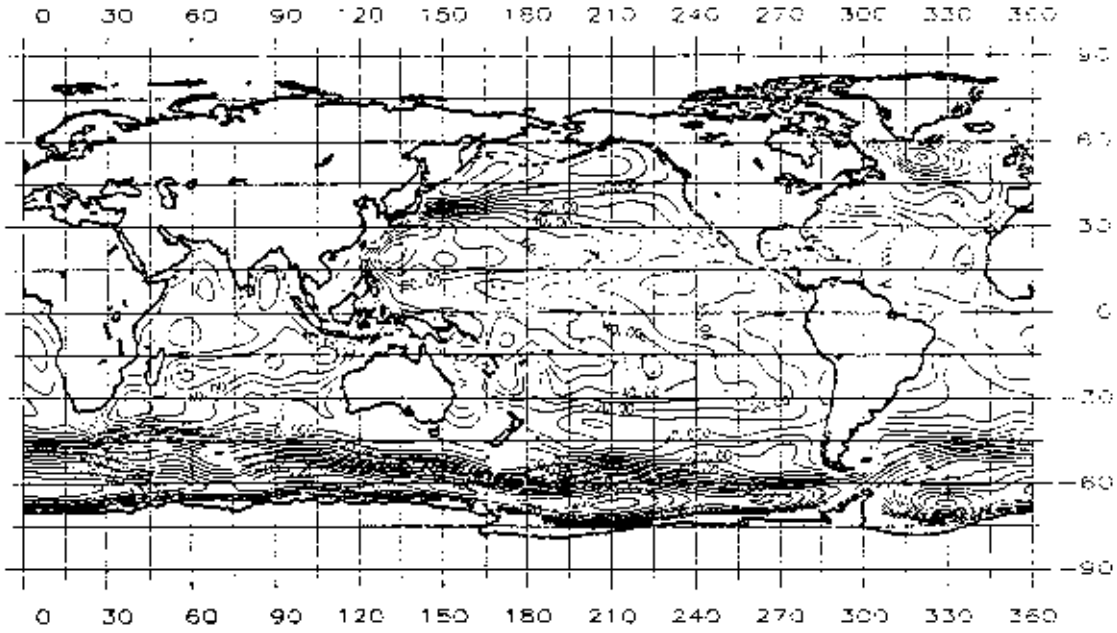


Figure 6.23 SST From Solution shnqto24, CI = 10 cm.

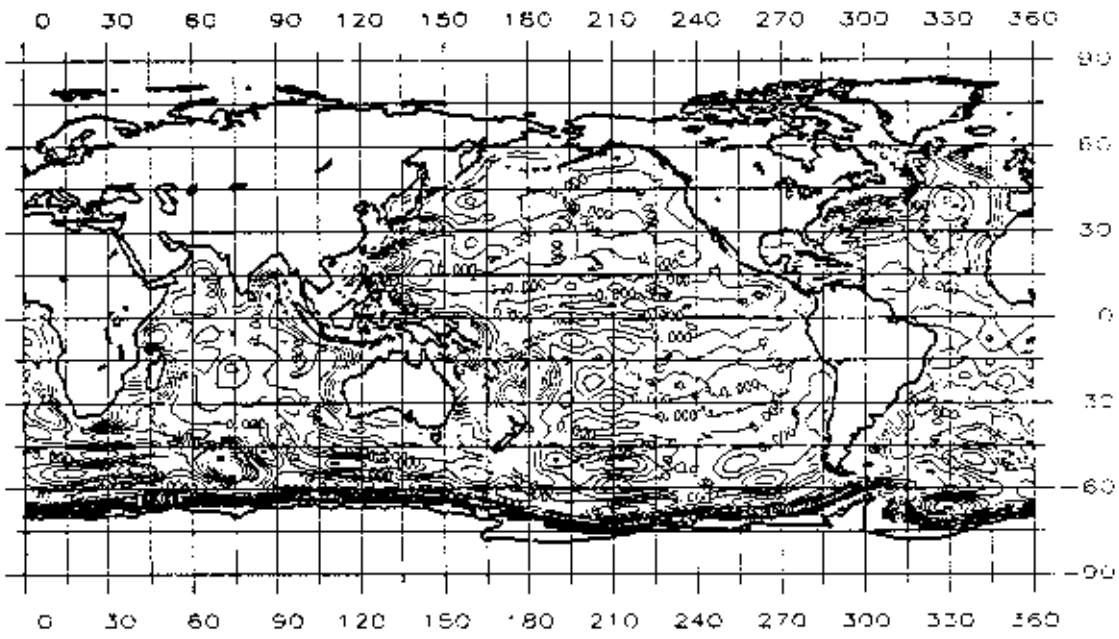


Figure 6.24 Difference Between the Levitus SST and SST From Solution shnqto24, CI = 5 cm.

7. Application of Orthonormal Functions to the Simultaneous Radial Orbit Error Reduction and Geoid-SST Estimation

In this chapter, we shall apply the orthonormal functions developed in Chapters 5 and 6 to the joint model of radial orbit error reduction and SST estimation. A short review of analytical satellite theory related to the radial orbit perturbation will be made first. Then, some practical problems in the existing joint model used by previous investigators will be pointed out and some new considerations are made. Finally, the experiments will be conducted with emphasis on the use of the ON functions.

7.1 Linearized Lagrange's Equations of Motion and Radial Orbit Error

Tracing back to the late 50's and early 60's, one could find that the starting works on analytical satellite theory are due to Brouwer (1959), Garfinkel (1959), and Kozai (1959), which all appeared in the journal "Celestial mechanics", Vol. 64. Following these pioneering works, numerous analytical satellite theories have emerged (even up to date).

For this study, it is felt that a brief review of analytical satellite theory, especially the one related to radial perturbation is needed, since later on some additional considerations for our estimation model will be made. Let us start with the motion of an artificial satellite in the earth's gravitational field (the non-gravitation origin force such as air drag, solar radiation pressure and the luni-solar attraction are not considered for the time being):

$$\ddot{\bar{r}} = \nabla V \tag{7.1}$$

where V is the earth's gravitational field and $\bar{r} = (x, y, z)$ contains the rectangular coordinates in either an earth-fixed system or the inertial system. From classical mechanics, such as Goldstein (1980), it is known that in some cases a set of generalized coordinates, will be a better choice to describe the motion of a system either for the purpose of interpretation or for the purpose of calculation. One particular choice involves the canonical variables (coordinates) described as follows. Let T and V be the kinematical energy and potential energy respectively of a system and L = T - V be the Lagrangian of that system. Furthermore, let the generalized coordinate be q_α, $\alpha = 1, 2, \ldots$, so that the generalized momenta of the system are:

$$p_\alpha = \frac{\partial L}{\partial \dot{q}_\alpha}, \alpha = 1, 2, \ldots \tag{7.2}$$

Defining the Hamiltonian of a system as H = T + V for a conservative system (such as the earth's gravitational field, excluding all the non-gravitational effects described above), one can find the equations of motion of a system as (Spiegel, 1967):

$$\dot{p}_\alpha = \frac{\partial H}{\partial q_\alpha}, \dot{q}_\alpha = \frac{\partial H}{\partial p_\alpha} \tag{7.3}$$

Such a formulation of equations of motion have been used to derive the analytical satellite theories in some cases. The representation (p_α, q_α) for a system is called a phase space

(Spiegel, 1967). The quantities p_α, q_α are known as the canonical variables (Goldstein, 1980). One well-known example of such a formulation is the use of Delaunay's variables L, G, H and l, g, h in the analytical satellite theory.

Despite the usefulness of canonical variables, in some cases a set of non-canonical variables can provide a better insight into the geometry and characteristics of a system. It turns out that for the motion of a satellite in the earth's gravitational field, the use of the Keplerian elements {a, e, I, M, ω, Ω} lead to a special system suitable for many analyses. The definitions of the six Keplerian elements are: a - the semi-major axis, e - the eccentricity of the osculating ellipse which osculates with the actual orbit at the point of tangency, I - the inclination, M - the mean anomaly, ω - the argument of perigee and Ω - the right ascension of the satellite ascending node with respect to the vernal equinox. The use of the Keplerian elements for the motion of a satellite leads to Lagrange's equations of motion (Kaula, 1966):

$$\frac{da}{dt} = \frac{2}{na}\frac{\partial R}{\partial M}$$

$$\frac{de}{dt} = \frac{1-e^2}{na^2e}\frac{\partial R}{\partial M} - \frac{(1-e^2)^{1/2}}{na^2e}\frac{\partial R}{\partial \omega}$$

$$\frac{dw}{dt} = -\frac{\cos I}{na^2(1-e^2)^{1/2}\sin I}\frac{\partial R}{\partial I} + \frac{(1-e)^{1/2}}{na^2e}\frac{\partial R}{\partial e}$$

$$\frac{dI}{dt} = \frac{\cos I}{na^2(1-e^2)^{1/2}\sin I}\frac{\partial R}{\partial \omega} - \frac{1}{na^2(1-e^2)^{1/2}\sin I}\frac{\partial R}{\partial \Omega}$$

$$\frac{d\Omega}{dt} = \frac{1}{na^2(1-e^2)^{1/2}\sin I}\frac{\partial R}{\partial I}$$

$$\frac{dM}{dt} = n - \frac{1-e^2}{na^2e}\frac{\partial R}{\partial e} - \frac{2}{na}\frac{\partial R}{\partial a} \qquad (7.4)$$

where R is the earth's gravitational field, excluding the central term GM/r (GM is gravitational constant, r is the geocentric distance to the satellite) and n is the mean motion defined as

$$n = \sqrt{GM/a^3} \qquad (7.5)$$

Apparently (7.4) is a system of non-linear first-order differential equations. A solution for (7.4) is possible only in some neighborhood of reference points (or coordinates) through some linearization. To this end, we first transform the function R from the spherical coordinates to the Keplerian elements, as done in Kaula (1966):

$$R = \sum_{\ell m} \frac{GMa_e^{\ell}}{a^{\ell+1}} \sum_{p=0}^{\ell} F_{\ell mp}(I) \sum_{q=-\infty}^{\infty} G_{\ell pq}(e) S_{\ell mpq}(\omega, M, \Omega, \theta) \tag{7.6}$$

where ℓ and m are degree and order of spherical harmonics respectively, θ is the Greenwich sidereal angle, a_e is the earth's equatorial radius and $S_{\ell mpq}$ is

$$S_{\ell mpq} = \begin{bmatrix} C_{\ell m} \\ -S_{\ell m} \end{bmatrix}_{\ell-m \text{ odd}}^{\ell-m \text{ even}} \cos \psi_{\ell mpq} + \begin{bmatrix} S_{\ell m} \\ C_{\ell m} \end{bmatrix}_{\ell-m \text{ odd}}^{\ell-m \text{ even}} \sin \psi_{\ell mpq} \tag{7.7}$$

and where $C_{\ell m}$ and $S_{\ell m}$ are the geopotential coefficients. The argument $\psi_{\ell mpq}$ is

$$\psi_{\ell mpq} = (\ell - 2p)\omega + (\ell - 2p + q)M + m(\Omega - \theta) \tag{7.8}$$

The function $F_{\ell mp}(I)$ is known as inclination function for which many new algorithms for evaluation have been derived, see e.g., Goad (1987), Kostelecky et al. (1986) and Sneeuw (1991). The function $G_{\ell pq}(e)$ is the eccentricity function which is basically the coefficient in the expansion (Gooding and King-Hele, 1988):

$$\left(\frac{a}{r}\right)^{\ell+1} e^{i(\ell-2p)f} = \sum_{q=-\infty}^{\infty} G_{\ell pq}(e) e^{i(\ell-2p+q)M} \tag{7.9}$$

or equivalently (multiplying $e^{-i(\ell-2p)M}$ on both sides of (7.9)):

$$\left(\frac{a}{r}\right)^{\ell+1} e^{i(\ell-2p)(f-M)} = \sum_{q=-\infty}^{\infty} G_{\ell pq}(e) e^{iqM} \tag{7.10}$$

which enables a FFT evaluation of $G_{lpq}(e)$, as described in Goad (1987). In (7.9), f is the true anomaly (the angle between line of apsides and the line from geocenter to satellite). Based on the observed fact that the dominant perturbations of geodetic satellite are the secular motions in ω, Ω, M due to the earth's oblateness, we may linearize (7.4) in the neighborhood of an initial system whose motion has the form

$$\frac{da}{dt} = \frac{de}{dt} = \frac{dI}{dt} = 0, \frac{d\omega}{dt} = \dot{\omega}_o, \frac{d\Omega}{dt} = \dot{\Omega}_o, \frac{dM}{dt} = \dot{M}_o \tag{7.11}$$

where $\dot{\omega}_o$, $\dot{\Omega}_o$ and $\dot{M}_o$ are the secular rates of ω_o, Ω_o, M_o to be described later. Considering that the only time dependent variables on the right hand sides of (7.4) are ω, Ω, and M, we can write the arguments of the cosine-sine functions in R as:

$$\psi_{\ell mpq} = \dot{\psi}_{\ell mpq} t + \psi_o \tag{7.12}$$

where

$$\dot{\psi}_{\ell mpq} = (\ell - 2p)\dot{\omega}_o + (\ell - 2p + q)\dot{M}_o + m(\dot{\Omega}_o - \dot{\theta}_o) \tag{7.13}$$

and ψ_o is a constant that will become clear in the coming discussions and $\dot{\theta}_o$ is the mean rotational rate of the earth. For each ℓmpq component of R we may integrate both sides of (7.4) to get the first order correction (or perturbation) to the initial system (Kaula, 1966):

$$\Delta a_{\ell mpq} = \frac{2\mu a_e^{\ell}}{n a^{\ell+2}\dot{\psi}} (k+q)\, F_{\ell mp} G_{\ell pq} S_{\ell mpq}$$

$$\Delta e_{\ell mpq} = \frac{\mu a_e^{\ell}}{n a^{\ell+3} e\dot{\psi}} \eta[\eta(k+q)-k]\, F_{\ell mp} G_{\ell pq} S_{\ell mpq}$$

$$\Delta \omega_{\ell mpq} = \frac{\mu a_e^{\ell}}{n a^{\ell+3}\dot{\psi}} \left[\eta e^{-1} F_{\ell mp} G'_{\ell pq} - \cot I\, \eta^{-1} F'_{\ell mp} G_{\ell pq}\right] \overline{S}_{\ell mpq}$$

$$\Delta I_{\ell mpq} = \frac{\mu a_e^{\ell}}{n a^{\ell+3}\eta \sin I\dot{\psi}} (k \cos I - m)\, F_{\ell mp} G_{\ell pq} S_{\ell mpq}$$

$$\Delta \Omega_{\ell mpq} = \frac{\mu a_e^{\ell}}{n a^{\ell+3}\eta \sin I\dot{\psi}} F'_{\ell mp} G_{\ell pq} \overline{S}_{\ell mpq}$$

$$\Delta M_{\ell mpq} = \frac{\mu a_e^{\ell}}{n a^{\ell+3}\dot{\psi}} \left[-\eta^2 e^{-1} G'_{\ell pq} + 2(\ell+1)\, G_{\ell pq}\right] F_{\ell mp} \overline{S}_{\ell mpq} \tag{7.14}$$

where $\mu = GM$, $k = \ell - 2p$, $\eta = (1 - e^2)^{1/2}$, $F'_{\ell mp} = dF_{\ell mp}(I)/dI$, $G'_{\ell pq} = dG_{\ell pq}(e)/de$, $\dot{\psi} = \dot{\psi}_{\ell mpq}$ and $\overline{S}_{\ell mpq}$ is the integral of $S_{\ell mpq}$ with respect to its argument. By (7.11) and (7.14), we can obtain the approximate orbit at some epoch:

$$a(t) \approx a_o + \Delta a$$

$$e(t) \approx e_o + \Delta e$$

$$I(t) \approx I_o + \Delta I$$

$$\omega(t) \approx \omega_o + \dot{\omega}_o t + \Delta\omega = \omega_o(t) + \Delta\omega$$

$$\Omega(t) \approx \Omega_o + \dot{\Omega}_o t + \Delta\Omega = \Omega_o(t) + \Delta\Omega$$

$$M(t) \approx M_o + \dot{M}_o t + \Delta M = M_o(t) + \Delta M \tag{7.15}$$

where Δa, ..., ΔM are the sums over individual terms in (7.14) up to a maximum harmonic degree, a_o, ..., M_o are mean Keplerian elements at the beginning of the orbit, and t is the

time elapsed since the beginning of the orbit. The quantities a_o, e_o, I_o, ω_o, Ω_o, M_o and $\dot{\omega}_o$, $\dot{\Omega}_o$, $\dot{M}_o$ define a reference orbit which, together with Δa, ..., ΔM, provides an approximate solution to Lagrange's equations of motion (7.4) in the form of (7.15).

At least two methods can be used to determine a reference orbit. In one method (e.g., Engelis, 1987a, p. 32), I_o, ω_o, Ω_o, M_o are obtained from a satellite's initial state vector; a_o and e_o are found by subtracting the periodic effect on a and e due to J_2 from the corresponding values given in the initial state vector; $\dot{\Omega}_o$ and $\dot{M}_o$ are computed by (Kaula, 1966)

$$\dot{\Omega}_o = \frac{-3nJ_2a_e^2}{2(1-e^2)^2a^2}\cos I, \quad \dot{M}_o = n + \frac{3nJ_2a_e^2}{4(1-e^2)^{3/2}a^2}(3\cos^2 I - 1) \tag{7.16}$$

and finally $\dot{\omega}_o$ is calculated by

$$\dot{\omega}_o = \frac{-3nJ_2a_e^2}{4(1-e^2)^2a^2}(1-5\cos^2 I) + \dot{\omega}_o' \tag{7.17}$$

where $\dot{\omega}_o'$ is the effect on ω due to all odd zonal harmonics, as described by Cook (1966). In a second method (e.g., Colombo, 1984), the reference orbit is obtained from a least-squares fit of the precise ephemeris by a model $\bar{S}_o + \dot{\bar{S}}_o\Delta t$ where $\bar{S}_o$ indicates the mean elements a_o, ..., M_o and $\dot{\bar{S}}$ the secular rates $\dot{\omega}_o$, $\dot{\Omega}_o$, $\dot{M}_o$. In this study, we will use Engelis' method to determine a reference orbit. However, the secular rates $\dot{\Omega}_o$ and $\dot{M}_o$ are computed by (7.16), plus the second order effect due to J_2^2. The second order effect used here is from Kaula (1966, p. 48, eq. (3.113)).

In altimetry, the radial position of the satellite is of greatest interest. For a small e (on the order of 10^{-3}), we have

$$r = a(1-e\cos E) = a(1-e\cos M) + O(e^2) \tag{7.18}$$

The approximate solution for r at epoch t (relative to the beginning of the orbit) is then

$$r(t) = r_o(t) + \Delta r(t) \tag{7.19}$$

where

$$\Delta r(t) = \Delta a(1-e_o\cos M_o(t)) - a_o\Delta e\cos M_o(t) + a_oe_o\Delta M\sin M_o(t) \tag{7.20}$$

and $r_o(t) = a_o(1-e_o\cos M_o(t))$ is the reference radial position. In deriving (7.19), the second order terms such as $\Delta a\Delta e$ are neglected and the assumption that $\sin\Delta M = \Delta M$ is made. The beauty of (7.20) is that it is already a linear form in terms of the geopotential coefficients C_{nm}, S_{nm} which appear in Δa, Δe and ΔM. Thus, the errors in the coefficients can be linearly propagated to the radial position through (7.20). On the other hand, when r(t) is the observable, as in altimetry, the linear form of (7.20) has automatically provided the

elements of the design matrix in the parameter estimation problem involving the geopotential coefficients.

Consequently, using the linear form (7.20) one is able to investigate a satellite radial orbit error if the errors in the geopotential coefficients are known. In such a case, we may regard $\Delta r(t)$ in (7.20) as the radial orbit error and express Δr as:

$$\Delta r = \Delta r(\Delta C_{nm}, \Delta S_{nm}, t) \tag{7.21}$$

where ΔC_{nm} and ΔS_{nm} are coefficient errors (so one can merely replace C_{nm} and S_{nm} in Δa, Δe and ΔM by ΔC_{nm} and ΔS_{nm} to make Δr become "error" or "increment" to the initial value). Another advantage of such a linear form is offered by the fact that the argument $\psi_{\ell mpq}$ as expressed in (7.12) allows the radial orbit error to be classified into different frequencies and thus provides a possibility of spectral analysis for the radial orbit error. To get a complete form of the frequency representation for Δr in (7.21), we have to substitute the errors Δa, Δe, ΔM (now they are "errors" due to ΔC_{nm}, ΔS_{nm}) from (7.14) into (7.20). Such a substitution requires considerable algebra and will not be shown here. A detailed exposition of the substitution can be found in Colombo (1984) or Engelis (1987a). Engelis' (1987a) derivations for Δr showed almost every possible step needed for an actual calculation and the software for the simultaneous solutions conducted in this chapter will be based on his results (for the orbit error part only). Furthermore, for a small e, it is sufficient to use q values only up to $q = \pm 1$ for the components in Δa, Δe and ΔM (Engelis, 1987a).

An approximate and compact form of Δr using the Fourier representation approach is provided by Wagner (1985). In his approach, the following approximate formulae for the eccentricity function $G_{\ell pq}$ are used:

$$G_{\ell p0}(e) = 1,\ G_{\ell p\pm1}(e) = \frac{e}{2}(\ell \pm 2k + 1) \tag{7.22}$$

where $k = \ell - 2p$. Then, Δr can be expressed as (Wagner, 1985, eq. (13); also Schrama, 1989; Engelis, 1987a, p. 72):

$$\Delta r(t) = \sum_{k=-\ell_{max}}^{\ell_{max}} \sum_{m=0}^{\ell_{max}} \sum_{\ell=\ell_{min},\ \ell,\ k\ \text{parity}}^{\ell_{max}} H_{\ell km}\left[C_{\ell km}\cos\dot{\psi}_{km}t + S_{\ell km}\sin\dot{\psi}_{km}t\right] \tag{7.23}$$

where

$$(C_{\ell km}, S_{\ell km}) = \begin{bmatrix} \Delta C_{\ell m} & \Delta S_{\ell m} \\ -\Delta S_{\ell m} & \Delta C_{\ell m} \end{bmatrix}_{\ell-m\ \text{odd}}^{\ell-m\ \text{even}} \cos\psi_{kmo} + \begin{bmatrix} \Delta S_{\ell m} & -\Delta C_{\ell m} \\ \Delta C_{\ell m} & \Delta S_{\ell m} \end{bmatrix} \sin\psi_{kmo} \tag{7.24}$$

$$H_{\ell km} = a\left(\frac{a_e}{a}\right)^{\ell} F_{\ell m, \frac{\ell-k}{2}} \frac{\beta(\ell+1) - 2k}{\beta(\beta^2 - 1)} \tag{7.25}$$

$$\dot{\psi}_{km} = k(\dot{\omega}_o + \dot{M}_o) + m(\dot{\Omega}_o - \dot{\theta}_o),\ \dot{\psi}_{km0} = k(\omega_o + M_o) + m(\Omega_o - \theta) \tag{7.26}$$

$$\beta = \frac{\dot{\psi}_{km}}{\dot{M}_o} \tag{7.27}$$

$$\ell_{min} = \max(|k|, 2, m + \mathrm{mod}(|k| - m, 2)),\ k = \ell - 2p \tag{7.28}$$

The summation over ℓ is evaluated for the ℓ values having the same parity with k (k and ℓ are simultaneously even or odd). ℓ_{max} is the maximum harmonic degree sensitive to the earth's gravitational field (depending mainly on the satellite's height). The β value shows the satellite orbit frequency in terms of cyc/rev. Of high importance is $H_{\ell km}$ known as "sensitivity spectrum" (Wagner, 1985). $H_{\ell km}$ becomes singular when (1) $\beta = 0$, the case when $\dot{\psi}_{km}=0$, or the resonance case; (2) $\beta = \pm 1$, this will require that m = 0 and k = 1, hence $\ell = 2p + 1$. Thus the second case is due to the odd zonal terms. Further, in the second case since k = 1, m = 0, the corresponding orbit frequency is 1 (assuming $\dot{\omega}_o$ is small enough to be neglected, see (7.26)). Due to the singularity of $H_{\ell mk}$ in these two cases, it turns out that in the linear orbit theory the resonance effect and the 1 cyc/rev frequency cause most trouble in representing the satellite radial orbit errors.

Although Wagner's (1985) approximate formula (7.23) provides a good insight into the radial orbit frequency, a more rigorous classification of frequency will be based on the $\dot{\psi}_{\ell mpq}$ value in (7.13). Re-writing (7.13) as

$$\begin{aligned}\dot{\psi}_{\ell mpq} &= (\ell - 2p + q)(\dot{\omega}_o + \dot{M}_o) + m(\dot{\Omega}_o - \dot{\theta}_o) - q\dot{\omega}_o \\ &= k(\dot{\omega}_o + \dot{M}_o) + m(\dot{\Omega}_o - \dot{\theta}_o) - q\dot{\omega}_o\end{aligned} \tag{7.29}$$

we may use the conditions of k, m, ℓ to classify the orbit frequency, as done in Table 7.1. Table 7.1 has been based on Colombo (1984), Schrama (1986), Engelis (1987a) and Reigber (1989). In (7.29) we have defined that $k = \ell - 2p + q$ and the q values are restricted to 0, ± 1.

Table 7.1 Frequency Classification of Satellite Orbit Using Keplerian Elements

Designation	k	m	ℓ	$\dot{\psi}_{\ell mpq}$	Typical period (days)
perfect resonance*	0	0	even	0	∞
deep resonance	0	0	odd	$-\dot{\omega}_o$	100 - 200
shallow resonance†	> 0	> 9		≈ 0	5
m-daily	0	$1 \le m \le 9$		$m(\dot{\Omega}_o - \dot{\theta}_o)$	1/m
long period	0	0	odd	$(\ell - 2p)\dot{\omega}_o$	
short period	$\neq 0$	> 0			1/20

*Excluding the commensurate orbit: $m = \beta\gamma$, $k = \alpha\gamma$, $\gamma = 1, 2, ..., q = 0$

† $k(\dot{\omega}_o + \dot{M}_o) \approx m(\dot{\theta}_o - \dot{\Omega}_o)$, q = 0

In the case of perfect resonance or even deep/shallow resonance, the formulae such as (7.14) and (7.23) become invalid and hence a different formula for the radial orbit error must be derived. Such a derivation may be found in Engelis (1987a) or Colombo (1984). In addition to the radial orbit error arising from the linearized solution of Lagrange's equations of motion, the radial orbit error can also be caused by the second order effect not accounted for by the linear theory discussed above. Furthermore, the errors of non-gravitational origin, such as initial state vector error, air drag, solar radiation pressure, etc., will also give rise to radial orbit error. The following is a summary of radial orbit errors considered in this study:

(1) First order radial orbit error

The error $\Delta r(t)$ in (7.23), but with Engelis' (1987a) rigorous formulations (see Section 5.2, ibid.). We now denote such an error as $\Delta_1 r(t)$.

(2) Resonance effect (perfect, deep and shallow)

As mentioned before, the solution such as (7.23) fails if the argument $\dot{\psi}_{\ell mpq} \approx 0$. In such a case, we will still assume that a, e, I on the righthand side of (7.4) are time-invariant and now the rates of change of a, e M become secular or nearly secular (depending on whether $\dot{\psi}_{\ell mpq}$ is exactly zero or very "close" to zero, see also Table 7.1). Colombo (1984) showed that, in such a case, the perturbation Δa^r, Δe^r and ΔM^r (the counterparts of those in (7.14) for a resonance case) can be modeled as

$$\Delta S_i^r = \sum_{j=0}^{2} c_j^i (\Delta t)^j \qquad (7.30)$$

where ΔS_i^r are Δa^r, Δe^r and ΔM^r, c_j^i are constants and Δt is the time relative to the middle of an orbit arc. Substituting (7.30) into (7.20), the radial error due to resonance effect is

$$\Delta_r(t) = c_0 + c_1 \Delta t \cos M_o(t) + c_2 \Delta t \sin M_o(t) + c_3 \Delta t^2 \cos M_o(t) + c_4 \Delta t^2 \sin M_o(t) \quad (7.31)$$

where c_j are coefficients to be determined. In the experiments conducted later, we will show that c_3 and c_4 are very small for the Geosat arcs from GEM-T2 orbit.

(3) Second order radial orbit error

The linear theory provided by Kaula (1966) only gives an approximate orbit and hence an approximate radial orbit error model such as (7.23). Improvement to the linear theory can be made by considering the interaction between the Δa, ... , ΔM values in (7.14) and rates of change da/dt, ... , dM/dt in (7.4). An elegant approach for such an improvement is Von Zeipel's method which may be found in Kaula (1966) or Brouwer and Clemence (1961). Another approach is to use the Lie-series. A detailed exposition of the application of the Lie-series for the first order and second order solution can be found in Cui (1990) who used Hill's variables instead of Keplerian elements.

Engelis (1987a) carried out direct integrations for the second order effect by considering the interaction between $\Delta_1 S_i$ and $\dot{S}_i$ in such a way:

$$\Delta_2 S_i = \sum_{j=1}^{6} \int \frac{\partial \dot{S}_i}{\partial S_j} \Delta_1 S_j dt, \; i = 1, \ldots, 6 \tag{7.32}$$

where $\dot{S}_j$ are the rates of change of Keplerian elements in (7.4), $\Delta_1 S_j$ are the first order perturbations in (7.14), and $\Delta_2 S_i$ are the second order perturbation of Keplerian elements. Neglecting small quantities, Engelis (1987a) obtained the final result for the second order radial orbit error as:

$$\Delta_2 r = b_1 \Delta t \sin M_o(t) + b_2 \Delta t \sin 2M_o(t) \tag{7.33}$$

where b_1 and b_2 are coefficients.

(4) Radial orbit error due to initial state vector

The initial state vectors are necessary quantities for orbit calculations. If errors exist in the initial state vectors, the first order radial orbit error due to such errors can be approximated by

$$\begin{aligned} \Delta_I r &= \frac{\partial r(t)}{\partial a} \Delta a_I + \frac{\partial r(t)}{\partial e} \Delta e_I + \frac{\partial r(t)}{\partial M} \Delta M_I \\ &= \Delta a_I - \left(\Delta a_I + a \Delta e_I\right) \cos M_o(t) + a e_I \Delta M_I \sin M_o(t) \end{aligned} \tag{7.34}$$

where Δa_I, Δe_I and ΔM_I are the initial state vector errors in the components a, e and M respectively. The expression for r(t) is taken from (7.18). If we further consider the interaction between the initial state vector error and the first order radial error, we will get time-dependent coefficients for the sine-cosine functions, as shown by Engelis (1987a). Therefore, the total error due to the initial state vector error is (cf. Engelis, 1988, eqns (33) and (34)):

$$\Delta_I r = \alpha_o + \alpha_1 \cos M_o(t) + \alpha_2 \sin M_o(t) + \alpha_3 \Delta t \sin M_o(t) + \alpha_4 \Delta t \sin 2 M_o(t) \tag{7.35}$$

(5) Other radial orbit errors

Other errors due to factors such as air drag, solar radiation pressure, etc. are mostly composed of a constant term and 1 cyc/rev terms with both time-dependent and time-dependent amplitudes (Engelis and Knudsen, 1989). Therefore, they can be modeled by the sine-cosine functions in a similar form as those in (2)-(4).

(6) Total radial orbit error considered in this study

From (1)-(5), the total radial orbit error considered in this study is:

$$\begin{aligned} \Delta_T r(t) = \Delta_1 r(t) &+ a_o + a_1 \cos M_o(t) + a_2 \sin M_o(t) + a_3 \Delta t \cos M_o(t) \\ &+ a_4 \Delta t \sin M_o(t) + a_5 \Delta t \sin 2M_o(t) + a_6 \Delta t^2 \cos M_o(t) + a_7 \Delta t^2 \sin M_o(t) \end{aligned} \tag{7.36}$$

Therefore, the first order radial orbit error $\Delta_1 r(t)$ is the "centerpiece" of the errors, the rest of the errors possess simple mathematical forms. In some cases, the coefficients $a_0,...,a_7$ are called "empirical" coefficients, due to the mixed effect arising from various error sources. Model (7.36) presents the radial orbit errors that are both of gravitational origin and non-gravitational origin.

It turns out that for a near-circular orbit, the position of the perigee cannot be easily defined (see Cook, 1966, or Engelis, 1987a, p. 31), therefore, the argument $M_0(t)$ of the sine-cosine series in (7.36) is hard to determine. The argument $\omega_0(t) + M_0(t)$, however, is well defined, as pointed out in Engelis (1987a, p. 31). If we treat the coefficients as truly "empirical" coefficients, then for a near-circular orbit it will be more stable to use the argument $\omega_0(t) + M_0(t)$, instead of $M_0(t)$. Using such a concept, we replace $M_0(t)$ in (7.36) by $\psi_0(t) = \omega_0(t) + M_0(t) = (\omega_0 + M_0)t + \psi_0 = \dot{\psi}t + \psi_0$ to get a new mathematical form for the total radial orbit error:

$$\Delta_T r(t) = \Delta_1 r(t) + a_0 + a_1 \cos \psi_0(t) + a_2 \sin \psi_0(t) + a_3 \Delta t \cos \psi_0(t) + a_4 \Delta t \sin \psi_0(t) + a_5 \Delta t \sin 2\psi_0(t) + a_6 \Delta t^2 \cos \psi_0(t) + a_4 \Delta t^2 \sin \psi_0(t) \qquad (7.37)$$

The new form in (7.37) is also used in Denker (1990) and Denker and Rapp (1990). The $\dot{\psi}$ value is the frequency associated with the 1 cyc/rev. However, the model in (7.37) is still not the final form adopted for this study. A slight modification for (7.37) will be made when we discuss the correlation between the 1 cyc/rev term in (7.37) and the origin shift of satellite tracking system.

7.2 Mathematical Models for Simultaneous Radial Orbit Error Reduction and Geoid-SST Estimation

7.2.1 Some Existing Problems in the Simultaneous Solution

Before we present the models of simultaneous solution, some important issues will be discussed first. By definition, the (stationary) SST is the departure of the sea surface from the geoid. The geoid itself is a surface whose determination still requires effort to date. Ideally, the geoid, and in turn the SST, will refer to a geocentric system. In satellite altimetry, the SST will be a quantity that is implicitly defined as the difference between the satellite-derived sea surface height and the geoid height implied by some geopotential model. In the simultaneous scheme, such a geopotential model will be the one that is used for orbit integration. The geoid according to such a model will refer to a geocentric system if the first degree harmonic terms vanish. However, for many obvious reasons (such as configuration of satellite tracking stations, instrument accuracies, etc.), the satellite orbit cannot be a perfect geocentric system and thus the satellite-derived sea surface height will also not be in a perfectly geocentric system. Since two reference systems are implied in the SST determination problem, the transformation between them must be considered. Assuming that the geoid system is geocentric and the satellite system is non-geocentric, then the SST error due to such a system inconsistency is (see also Rapp, 1989a, Vol. II, p. 134)

$$\Delta\zeta = \Delta Z \sin \phi + \Delta X \cos \phi \cos \lambda + \Delta Y \cos \phi \sin \lambda \qquad (7.38)$$

where we have assumed that such an inconsistency is only due to a translation vector (ΔX, ΔY, ΔZ) between two systems (disregarding rotations and scaling factor). In (7.38), ΔX, ΔY and ΔZ can be defined as the coordinates of the origin of the satellite system in the ideal geocentric system. It is easy to derive (7.38) if we project the translation vector on the unit normal vector of SST, ($\cos\phi \cos\lambda$, $\cos\phi \sin\lambda$, $\sin\phi$). The maximum error occurs when $\partial\Delta\zeta/\partial\phi = \partial\Delta\zeta/\partial\lambda = 0$, or $\tan\lambda = \Delta Y/\Delta X$ and $\tan\phi = \Delta Z/(\cos\lambda\,\Delta X + \sin\lambda\,\Delta Y)$. If we use ΔX = 30cm, ΔY = -25cm, ΔZ = -5cm, we find that $\phi = 7°17'47''$, $\lambda = 320°11'40''$ and the maximum SST error is 39cm. Such a geocentric shift problem has been pointed out by numerous researchers, e.g., Wagner (1986), Koblinsky (1989), Denker and Rapp (1990). To eliminate this type of SST error, obviously we need to get a geocentric satellite system, or we have to include it in the simultaneous estimation model.

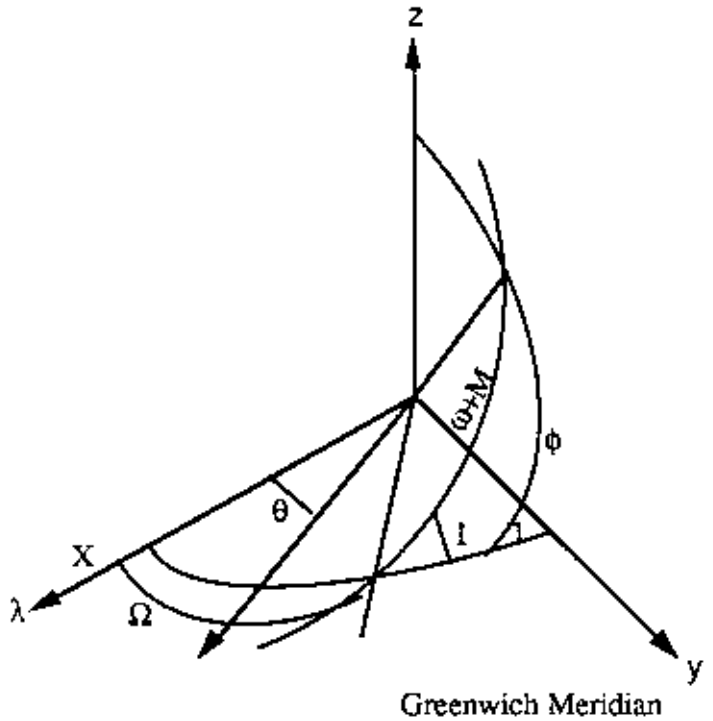


Figure 7.1 Satellite Orbit Geometry in a Near-Circular Orbit

Another problem is related to the 1 cyc/rev orbit error. From Figure 7.1, we have

$$\frac{\sin\phi}{\sin I} = \frac{\sin(\omega+M)}{\sin 90°} \tag{7.39}$$

Since in the linear orbit theory discussed in Section 7.1, we treat I as a constant, (7.39) becomes

$$\sin\phi = c\sin(\omega+M) \tag{7.40}$$

where c is a constant. Now from (7.37), the sine term of 1 cyc/rev radial orbit error is

$$\Delta r_s = a_2 \sin \psi_0(t) = a_2 \sin(\omega + M) = \frac{a_2}{c} \sin \phi \qquad (7.41)$$

Obviously from (7.38) and (7.41) we know that the errors due to the ΔZ-component of the geocentric shift and the sine term of 1 cyc/rev error are exactly the same (up to a constant). Therefore these two errors are indistinguishable in the adjustment process and must be modeled in a single functional form - either using $\sin\phi$ or $\sin(\omega + M)$. Consequently, it is only possible to determine the ΔX and ΔY components of the geocentric shift. Nevertheless, the error due to ΔZ has been taken care of by the mixed term, $\sin\phi$ or $\sin(\omega + M)$. In this study, we decide to use $\sin\phi$.

An additional problem arises from the use of spherical harmonic representation of SST. The problem is in the degree 1 terms. Using $P_0(t) = \sin\phi$, $P_1^1(t) = \cos\phi$, the degree 1 terms of spherical harmonic are

$$\zeta_1 = a_{10} \sin\phi + a_{11} \cos\phi \cos\lambda + b_{11} \cos\phi \sin\lambda \qquad (7.42)$$

Therefore, the degree 1 terms share exactly the same functional form as the error introduced by the geocentric shift. Furthermore, from (7.41) and (7.42), the functional form of the sine term of 1 cyc/rev error is exactly the same as that of a_{10} term, and this explains why numerous researchers have found a near-100% correlation between these two terms in the simultaneous estimation model (e.g., Wagner, 1986; Denker and Rapp, 1990). However, we must emphasize that such a result is due to the assumption that I = constant in the linear orbit theory.

The functional resemblance between the sine term of 1 cyc/rev error and the a_{10} term of spherical harmonic in (7.42) is somewhat artificial, in that it exists because we model the SST using the spherical harmonics. It can be taken care of by a different representation for SST, such as the ON functions or the Fourier-Tschebyscheff series discussed in Chapter 4. The functional resemblance between the ΔZ-component and the sine term of 1 cyc/rev error is quite real; since they all have physical meanings and cannot be replaced by other functional forms. Fortunately, they all belong to the "error" part of the simultaneous estimation model, and are theoretically solvable by error modeling with no harm to the SST.

A further trouble in the simultaneous model is caused by the mixed effect of the geocentric shift, 1 cyc/rev orbit error and the SST. The trouble is related to the geometry of the altimetric observation—the range between a satellite and the sea surface. It appears that the three quantities are all mixed in the radial direction and hence a distinction between them requires some effort. Furthermore, they are all purely geometric quantities and lack the dynamical property such as that of the geoid which can be connected to satellite orbit mechanics through potential coefficients. The situation is almost like a case where we only measure the total length of a line having three marked segments, but we still try to determine the individual lengths of the three segments. Therefore, it is only possible to estimate any of the three quantities when some a priori information is available. Exactly how we apply the a priori information will be discussed in the next section.

Although many uncertainties exist in the simultaneous model, the results from the works by, e.g., Engelis and Knudsen (1989), and Denker and Rapp (1990), still demonstrated the capability of such a model in improving the radial orbit accuracy and determining the geoid and SST even if an altimetric system (including the corrections to

observations, the orbits used, etc.) of moderate accuracies is used. As pointed out by Engelis (1987a, p. 131), in order to get a good result from the simultaneous model a necessary requirement is the existence of a well scaled and reliable geopotential error covariance matrix. Also, it is believed that a starting orbit of good accuracies is needed to lessen the uncertainties. With the state-of-the-art GEM-T2 orbits (Haines et al., 1990) to be used in the coming analyses, such requirements should be somewhat met. These points will be further enhanced when we analyze the results from the experiments in Section 7.5.

Despite the aforementioned problems, the experiments of simultaneous estimation will still be conducted using the GEOSAT altimeter data. The emphasis of such experiments will be on the applications of the ON functions to both geoid and SST. A slightly modified (with respect to that used by Denker and Rapp (1990)) orbital error model will be developed based on the above discussion.

7.2.2 Models for the Simultaneous Solution

Now we turn to the description of the mathematical models used for the simultaneous solutions. We first present the observation geometry in Figure 7.2 where we assume that the altimeter observations are made only over the oceans. Our goal here is to use the range observation ρ to solve for the corrections to the geopotential coefficients used in the orbit and geoid calculations, the empirical coefficients $a_0, \ldots, a_7$, the geocentric shift components and finally the SST which are represented by some basis functions (orthonormal functions or spherical harmonics). The meanings of the quantities shown in Figure 7.2 are:

ρ: range observation of altimeter
v: noise of ρ.
h_c: computed satellite's ellipsoidal height, according to some force model and satellite tracking station coordinates. h_c refers to an ellipsoid consistent with the force model and the satellite tracking station coordinates.
T: height correction due to the geocentric shift of the satellite coordinate system, containing the effect of ΔX and ΔY only.
Δr: radial orbit error due to incorrect force model, incorrect initial state vector, etc., as defined in (7.37) with a_2 changed according to the discussion in Section 7.2.1.
ζ: (quasi) sea surface topography, represented by some basis functions.
N_c: computed geoid undulation from the geopotential model used for orbital calculation, up to a maximum degree Nmax.
ΔN_0: omitted undulation due to spherical harmonics Nmax + 1,...,∞. Practically it is computed from an existing high degree geopotential model.
ΔN: correction to N_c, due to the incorrect geopotential model used for computing N_c.

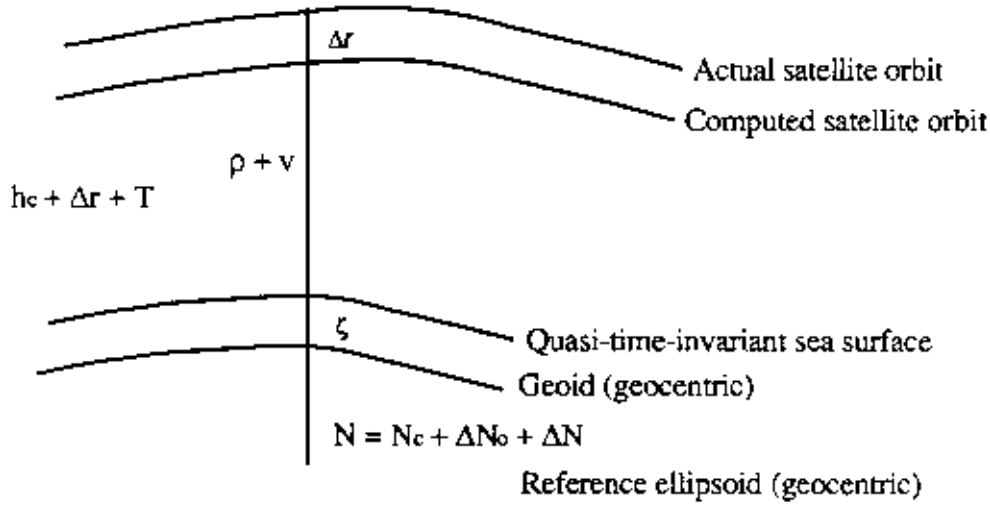


Figure 7.2 Geometry of Altimeter Observations Over the Oceans

Considering the total radial orbit error in (7.37), the discussion about the inseparability of the ΔZ-component of the geocentric shift and the sine term of 1 cyc/rev error, and all the parameters implied by the quantities shown in Figure 7.2, we can obtain the observation equations as follows:

$$h_c + \Delta r + T - (\rho+v) - \zeta - (N_c + \Delta N_0 + \Delta N) = 0 \tag{7.43}$$

If we write

$$\Delta ssh = h_c - \rho - N_c - \Delta N_0 \tag{7.44}$$

$$\Delta h = \Delta r - \Delta N + T - \zeta = \Delta h(X) \tag{7.45}$$

then the observation equations in a matrix form are:

$${}_nV_1 = {}_nA_u X_1 + {}_nL_1 \tag{7.46}$$

where n is the number of observations, and u is the number of unknowns. Vector V contains the noises, X all the parameters to be estimated, L the observations. The derived observation Δssh in (7.44) is called "residual sea surface height" which is also an element of vector L (in fact Δssh is an "incremental" sea surface height). The unknown vector X consists of four parts, namely,

$${}_uX_1 = \begin{bmatrix} X_g \\ X_s \\ X_t \\ X_e \end{bmatrix} = \begin{bmatrix} \text{geopotential corrections} \\ \text{SST coefficients} \\ \text{geocentric shift} \\ \text{empirical coefficients} \end{bmatrix} \tag{7.47}$$

where $X_g = (\Delta C_{nm}, \Delta S_{nm})^T$, $X_s = (a_{nm}, b_{nm})^T$, $X_t = (\Delta X, \Delta Y)^T$, $X_e = (a_0, a_1, a_2, \ldots a_7)^T$. For the moment, the SST coefficients refer to either the ON function expansion

or the spherical harmonic expansion (the spherical harmonic expansion is for the purpose of comparison, the emphasis here is on the ON function expansion). In vector X_c, the coefficient a_2 corresponding to the sine term of 1 cyc/rev error has been replaced by the coefficient of the mixed term $\sin\phi$ (see the previous section); for the geocentric shift, only ΔX and ΔY are solvable (see also the previous section). Using $N_{max}^{pot} = 50$, $N_{max}^{SST} = 24$ (using the ON function expansion) and using only one single altimetric arc (to be discussed later), the total number of unknowns u, is 3226. One additional arc will introduce another 8 parameters (for that particular arc only) assuming SST remains the same.

The elements of the design matrix A are obtained from

$$a_{ij} = \partial(\Delta h_i) / \partial X_j \tag{7.48}$$

where i is the index for the rows and j the index for the columns of matrix A, X_j are the elements of vector X. Due to the definitions of functions Δr, ΔN, T and ζ in (7.45), Δh is already linear in terms of the unknown vector X. Thus elements a_{ij} are just obtained by identifying the functions associated with the parameters.

The stochastic model to be used for estimating X is the "random effects model" (Schaffrin, 1989, p. 396):

$$\left.\begin{array}{l} V = AX + L \ , \ E\{X\} = X - e = 0 \text{ given} \\ \begin{bmatrix} V \\ e \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} , \ \sigma^2 \begin{bmatrix} D & 0 \\ 0 & C_x \end{bmatrix}^{-1} \right) \end{array}\right\} \tag{7.49}$$

where e is the residual to the prior value of X, and C_x and D are the covariance matrices of X and V respectively (C_x and D are the prior information to be described later). σ^2 is the a-priori variance of unit weight (chosen to be 1 in this study). To get the estimate for X, we employ the least-squares principle:

$$X^T C_x^{-1} X + V^T D^{-1} V = \underset{X,\, V}{\text{minimum}} \tag{7.50}$$

Schaffrin (1989, p. 395) pointed out that by the Least-squares principle (7.50) the estimate for X is numerically identical to that from the collocation method with a finite dimensional form (Moritz, 1980, Section 21) in which X is "signal" and V is "noise". The estimate for X is (Schaffrin, 1989, p. 397):

$$\tilde{X} = -(A^T D^{-1} A + C_x^{-1})^{-1} A^T D^{-1} L \tag{7.51}$$

To find the covariance matrix for the error $(\tilde{X} - X)$, we first derive the covariance matrix of L:

$$C_\ell = \text{cov}(L,L) = \text{cov}(V\text{-}AX, V\text{-}AX) = AC_xA^T + D \tag{7.52}$$

where we have assumed no correlation between X and V. Furthermore, the cross-covariance matrix between the observation L and signal X is

$$C_{\ell x} = \mathrm{cov}(L,X) = \mathrm{cov}(AX + V, X) = AC_x$$
$$C_{x\ell} = C_{\ell x}^T = C_x A^T \tag{7.53}$$

Thus the covariance matrix for $(\tilde{X} - X)$ is (Moritz, 1980, p. 105):

$$\begin{aligned} C_{\tilde{X}-X} &= E_x = C_x - C_{x\ell} C_\ell^{-1} C_{\ell x} \\ &= C_x - C_x A^T (AC_x A^T + D)^{-1} AC_x \\ &= (A^T D^{-1} A + C_x^{-1})^{-1} \end{aligned} \tag{7.54}$$

where we have employed a matrix identity found in Uotila (1986, p. 157).

The reason why we need to use conditioning or a priori information in the simultaneous solution has been given in the previous section. A detailed account on the conditioning of the system is given in Engelis (1987a, Chapter 8). Wagner (1986) also pointed out the need for the conditioning of the system. Then the next question is how we get the C_x and the D matrices. For D, we assume no correlation among the observations so that only the diagonal elements exist and are equal to the variances of the noises. It is more problematic to find C_x due to four kinds of unknowns in vector X (see (7.47)). It is practically impossible to find the cross-covariances among the four kinds of unknowns, thus we assume that the cross-covariances are zero. Within each unknown set, the covariances are determined below.

For the covariance matrix of geopotential coefficients, one way to obtain the needed elements in the matrix is the so-called power rule. Such a method is based on a global covariance model and a covariance propogation technique (cf. Moritz, 1980, p. 160):

$$\left.\begin{aligned} &\mathrm{cov}(\bar{C}_{nm}, \bar{C}_{nm}) = \mathrm{cov}(\bar{S}_{nm}, \bar{S}_{nm}) = \frac{\sigma_n^2}{2n+1} \\ &\mathrm{cov}(\bar{C}_{nm}, \bar{S}_{nm}) = 0,\ \mathrm{cov}(\bar{C}_{nm}, \bar{C}_{pq}) = \mathrm{cov}(\bar{S}_{nm}, \bar{S}_{pq}) = 0 \\ &\text{if } n \neq p \text{ or } m \neq q \text{ or both} \end{aligned}\right\} \tag{7.55}$$

where σ_n^2 is the degree variance of geopotential coefficients of degree n:

$$\sigma_n^2 = \sum_{m=0}^{n} (\bar{C}_{nm}^2 + \bar{S}_{nm}^2) \tag{7.56}$$

where $\bar{C}_{nm}$ and $\bar{S}_{nm}$ may be found from some existing model or σ_n^2 may be directly obtained from Kaula's rule (Kaula, 1966, p. 98). Thus by this technique, we will get a diagonal covariance matrix for the geopotential part. Such a power rule has been used by, e.g., Marsh et al. (1990), Nerem et al. (1990). It is also known from Schwarz (1974, p. 40) that C_x in (7.49) is the covariance matrix of the remaining signal (with respect to a priori estimated signal), thus use of power rule $\sigma_n^2/(2n + 1)$ is meaningful when the expectation of the signals (coefficients) are zero.

A more promising covariance matrix of geopotential coefficients will be a well-calibrated C_x which truly reflects the accuracies of the coefficients. The calibration technique can be found in, for example, Marsh et al., (1989), Lerch (1991). If the corresponding geopotential coefficients are used in the orbit integrations for altimetric arcs, the well-calibrated covariance matrix is suitable for the a priori information needed in (7.49). In fact, such a choice is made in the present study.

For the covariance matrix of SST coefficients of ON functions, we shall use a power rule by analogy with that of spherical harmonics, since a well-calibrated covariance matrix is not available. For such a choice, the SST covariance matrix is a diagonal matrix with elements computed by

$$v_{nm} = \frac{\tau_n^2}{2n+1} \ , \ m=0,1,...,n \tag{7.57}$$

where τ_n^2 is the degree variance of the ON coefficients defined in (6.22) (for the actual values used in the experiments, see Section 7.5.1), n and m are the degree and order of the ON functions, respectively. If a spherical harmonic expansion for the SST is desired, one can simply replace τ_n^2 in (7.57) by the degree variances of spherical harmonic coefficients. Use of (7.57) implies that the coefficients of the same degree receive an "average" power due to the obvious form in (7.57). This is done in a similar way as for the spherical harmonics (i.e., (7.54) and (7.55)).

Since no information is available for the statistical properties of the geocentric shift and the empirical coefficients, the covariances of these two quantities are assumed to be infinity. Using the choices of covariance matrices of X_g, X_s, X_t and X_e, we finally obtain the desired matrix needed in (7.49).

In Chapters 5 and 6, we have emphasized that if a power rule is needed for conditioning a system, then it should be applied in a compatible domain, otherwise, the solution will result in a degraded resolution. For the SST power rule, the use of ON functions will be a suitable one since the ON functions have a compatible domain with that of SST. The power rule from spherical harmonics is normally based on a global analysis in which the SST on land are assumed to be zero, providing a conditioning that results in an inferior resolution at the continental boundaries, as experimentally shown in Chapter 6. As shown by Reigber (1989, p. 221), a disadvantage of using the power rule (in both ON functions and spherical harmonics) is that it reduces the absolute values of the coefficients because the implicit assumption of using model (7.49) is that the expected values of the unknowns are zero. In later experiments, we really find that the ON coefficients of higher degrees (larger than 13 roughly) from the Geosat solution approach zero faster than those from the Levitus SST. The result is the loss of high resolution SST. With the use of spherical harmonics, the situation is even worse.

Finally we have to discuss the problem of truncation effect due to the limited terms used to represent the SST. In a case where the SST is the only signal, such an effect, especially for the accuracy estimates of the expansion coefficients, has been discussed in Chapter 6. In the simultaneous model, in addition to the SST, we have other signals. Unlike the geoid undulation of higher degree (than N_{max}^{pot}), i.e., ΔN in Figure 7.2, that can be taken care of by some existing high degree geopotential model, the SST of higher degrees (than N_{max}^{SST}) so far really do not have a promising model. They are even neglected in the simultaneous solutions, such as those conducted by Engelis et al. (1989), Denker

and Rapp (1990), Nerem et al. (1990). As shown in Chapter 6, an ON function expansion to degree 10 or a spherical harmonic expansion to degree 10 (using least-squares fit) really cannot pick up the high frequency SST signal (see Figures 6.7 and 6.8). In the simultaneous solution, such a low-degree expansion (below 15) of SST then produces residual SST signals that are mixed with V in (7.46) which are supposed to be "noises" only. Apparently, with a low-degree expansion, the residual SST in the areas of energetic circulation should have relatively larger effect on the noises V. Consequently, we are minimizing the "noise + residual SST" instead of the "noise" alone. To avoid such a disturbing effect by the residual SST, it is suggested that we model the SST to a maximum degree as high as possible. According to the tests in Chapter 6, ON degree 24 is a sufficiently high number. Obviously, one has to worry about if we really can obtain the SST signal to the degree we model in the simultaneous solution, since the errors of geoid and others probably do not allow us to do so. This can be overcome by truncating the SST expansion after the simultaneous solution is made based on the signal-to-noise ratio, provided that the correlations of the expansion coefficients are negligibly small.

In the simultaneous model, we are in fact dealing with a problem of approximating a function (now SST) in the presence of other signals (geoid correction and radial orbit errors). Recalling the least-squares error principle in finding the ON coefficients in Chapters 2 and 5, we can easily see that minimization of $(X^TC_x^{-1}X + V^TD^{-1}V)$ cannot lead to an approximation of SST with minimum errors. Therefore, in the simultaneous model, the approximation to SST by any basis functions is not the "best" in the least-square approximation error sense. The expansion coefficients are simply treated as parameters, like rectangular coordinates of a point, to be recovered in a model such as (7.49) (see the discussion in Section 6.4.2). Treating the expansion coefficients in such a way will be more statistically meaningful in the parameter estimation problem under two conditions: (1) the signal (now SST) to be approximated is bandlimited with respect to the same basis functions used for the approximation and (2) the expansion is made up to the "highest frequency" term. Under these conditions, the signal is thus "represented" by an exact form through the basis functions and the rest of the problem in a model such as (7.49) will be taken care of by statistics (of course the estimated parameters will still be subject to the assumption in the model).

A totally different way of using a priori information for SST in the simultaneous solution will be to incorporate the equations of ocean dynamics into the solution, as proposed in, e.g., Wagner (1989), Marshal (1985). Under steady-state condition, the equations of motion of water particles are (Officer, 1974, p. 126):

$$\begin{aligned} 2\omega v_y \sin\phi &= \frac{1}{\rho}\frac{\partial P}{\partial x} \\ -2\omega v_x \sin\phi &= \frac{1}{\rho}\frac{\partial P}{\partial y} \end{aligned} \tag{7.59}$$

or

$$2\omega v \sin\phi = \frac{1}{\rho}\frac{dP}{dn} \tag{7.60}$$

where x and y are local horizontal coordinates along the east and north directions, ω is the earth's rotational velocity, ρ is the water density, P is the pressure of water particles,

$v = \sqrt{v_x^2 + v_y^2}$, and n is a coordinate direction perpendicular to v with its positive direction to the right of v.

Then Officer showed that the dynamic height, ζ, which is the SST we analyzed in Chapter 6, is (ibid., p. 128)

$$\zeta = \frac{1}{g}\int_{p}^{P_0} \frac{1}{\rho}\, dp \tag{7.61}$$

where g is the gravity, and P_0 is the pressure at the depth of no motion. What we now have is the "product" - ζ, not the equations of motion, which involve more quantities. Thus we have to look at the ζ value in (7.61) in a different way. Assuming that ζ , which results from the equations of motion, can be treated as an observation and can be expanded with respect to some basis functions:

$$V_\zeta = \zeta(X_2) - \zeta = B_2X_2 - \zeta \tag{7.62}$$

where we have treated ζ as a vector containing all the observations of SST. X_2 is a vector containing the expansion coefficients and V_ζ is a noise vector. The expansion degree must be sufficiently high (for the Levitus SST, ON expansion to degree 24 will be enough) to make V_ζ purely "noise". Now we group all parameters other than the SST coefficients in (7.47) into vector X_1 , so that (7.46) can be rewritten as

$$V = \Delta h(X_1, X_2) + L = (A_1\ A_2)\begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix} + L \tag{7.63}$$

Now we assume that we have the a priori weight matrix P_1 for the unknown vector X_1 and a solution is sought with the condition

$$\phi = X_1^T P_1 X_1 + V^T D^{-1} V + V_\zeta^T P_\zeta V_\zeta = \text{a minimum} \tag{7.64}$$

Then we have

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} A_1^T D^{-1} A_1 + P_1 & A_1^T D^{-1} A_2 \\ A_2^T D^{-1} A_1 & A_1^T D^{-1} A_2 + B_2^T P_\zeta B_2 \end{pmatrix}^{-1} \begin{pmatrix} -A_1^T D^{-1} L \\ -A_2^T D^{-1} L + B_2^T P_\zeta \zeta \end{pmatrix} \tag{7.65}$$

In the above discussion, we need to find the accuracies of the SST in order to form the weight matrix P_ζ. For the Levitus SST, it is believed that the error upper bound is 25 cm (see also Nerem et al., 1990, p. 3168), due to the unknown effects of the level of no motion. Since the Levitus SST data are scattered with respect to time (a period of almost 70 years), it is no doubt that the SST from the one-month or even one-year Geosat solution will not be compatible with the Levitus' SST values. Thus using the Levitus SST as additional observations in the simultaneous solution (in the form of normal equations of the coefficients of some basis functions) is more or less to provide the a priori information for

enhancing the separation of the geoid and the SST. Consequently, in addition to assigning pessimistic accuracies for the Levitus SST, downweighting the normal of the Levitus SST, i.e., $B_2^t P_\zeta B$ is also needed. These will be discussed in the experimental part.

7.3 Applications of ON Functions in the Simultaneous Solution

One obvious application of the ON functions is to serve as basis functions for the SST representation in the simultaneous solution. Using the oceanic a priori information, specifically that derived from the Levitus SST, conditioning of the system is possible through the use of the power rule - the so called "mild" constraint (Wagner, 1986) or "light" constraint (Reigber, 1989) for the SST ON coefficients. The other application is on the spectral analysis for the signal/error of geoid and SST. For the SST, the spectral application of the ON functions is clear, as we have done in Chapter 6; for the geoid, we obviously can only concentrate on the oceanic part, i.e., the oceanic geoid, and it is only the oceanic domain where we can take advantage of the orthonormality of the ON functions for spectral studies. Since the geoid is not modeled directly by the ON functions in the simultaneous solution, a spectral analysis for the signal/error using the ON functions requires some transformations.

To get the ON coefficients of the oceanic geoid, one can follow the steps illustrated in Chapter 6: we first perform harmonic synthesis from the spherical harmonic coefficients obtained from the simultaneous solution to get mean undulations (strictly speaking, they are undulation corrections) at the equiangular blocks over the oceans (for this current study, the blocksize is chosen to be 1° x 1°); then the ON coefficients are obtained by harmonic analysis using these mean values. Or, one can perform direct transformations as follows (now we use system $\{Z_j\}$, see (5.76)):

$$\hat{\alpha}_j^N = (\Delta N, Z_j) = R_e \sum_{n=2}^{Nmax} \sum_{m=0}^{n} \{\Delta\bar{C}_{nm}(\bar{R}_{nm}, Z_j) + \Delta\bar{S}_{nm}(\bar{S}_{nm}, Z_j)\} \tag{7.66}$$

where $\hat{\alpha}_j^N$ are the ON coefficients with respect to system $\{Z_j\}$, and ΔN is the geoid undulation correction. The inner products $(\cdot,\cdot)$ are carried out by integration over the oceans. Again we have used an approximate formula for ΔN as in (6.111).

A more convenient way to calculate the accuracy estimate of $\hat{\alpha}_j^N$ will be that based on the first method, namely, we first obtain the ON coefficients from

$$\begin{aligned}\hat{\alpha}_j^N &= \frac{1}{a_\sigma}\iint_\sigma \Delta N(\theta,\lambda)\, Z_j(\theta,\lambda)\, d\sigma,\ \sigma=\text{oceans} \\ &= \frac{1}{a_\sigma}\sum_{\ell=1}^{M} \overline{\Delta N}_\ell I Z_j^\ell\end{aligned} \tag{7.67}$$

where a_σ is the area of the oceans, M the number of equiangular blocks, ΔN_ℓ the mean undulation at block ℓ, and IZ_j^ℓ the integration of function Z_j over block ℓ, see also (6.9) and (6.29) for better understanding (where we dealt with system $\{X_j\}$). Written in a matrix form, (7.67) is

$$\begin{bmatrix} \widehat{\alpha}_1^N \\ \widehat{\alpha}_2^N \\ \vdots \\ \widehat{\alpha}_\mu^N \end{bmatrix} = \frac{1}{a_\sigma} \begin{bmatrix} IZ_1^1 & IZ_1^2 & \cdots & IZ_1^M \\ IZ_2^1 & IZ_2^2 & \cdots & IZ_2^M \\ & & \vdots & \\ IZ_\mu^1 & IZ_\mu^2 & \cdots & IZ_\mu^M \end{bmatrix} \begin{bmatrix} \overline{\Delta N_1} \\ \overline{\Delta N_2} \\ \vdots \\ \overline{\Delta N_M} \end{bmatrix}$$

or

$$X_z = \frac{1}{a_\sigma} (B_z^*)^T N_z \tag{7.68}$$

where B_z^* is a matrix containing IZ_j^ℓ, N_z is a vector containing $\overline{\Delta N_\ell}$, and μ is the highest "degree" of the ON function expansion. Now $\overline{\Delta N_\ell}$ can be obtained from the geopotential coefficient corrections in the simultaneous solution as follows:

$$\begin{aligned} \overline{\Delta N_\ell} &= \frac{1}{\Delta\sigma_\ell} \sum_{j=0}^{\nu} \overline{\Delta C_j} \iint_{\Delta\sigma_\ell} L_j(\theta,\lambda) d\sigma \\ &= \frac{1}{\Delta\sigma_\ell} \sum_{j=0}^{\nu} \overline{\Delta C_j} IL_j^\ell \end{aligned} \tag{7.69}$$

where $\overline{\Delta C_j}$ is a single-index form of the geopotential coefficient corrections $\overline{\Delta C_{nm}}$ and $\overline{\Delta S_{nm}}$, ν is a number corresponding to Nmax of the geopotential coefficients and IL_j^ℓ is the integration of spherical harmonics over block ℓ. Eqn. (7.69) can be written in a matrix form as:

$$\begin{aligned} N_z &= \begin{bmatrix} \frac{1}{\Delta\sigma_1} & & 0 \\ & \frac{1}{\Delta\sigma_2} & \\ & & \ddots \\ 0 & & \frac{1}{\Delta\sigma_M} \end{bmatrix} \begin{bmatrix} IL_1^1 & IL_2^1 & \cdots & IL_\nu^1 \\ IL_1^2 & IL_2^2 & \cdots & IL_\nu^2 \\ & & \vdots & \\ IL_1^M & IL_2^M & \cdots & IL_\nu^M \end{bmatrix} \begin{bmatrix} \overline{\Delta C_1} \\ \overline{\Delta C_2} \\ \vdots \\ \overline{\Delta C_\nu} \end{bmatrix} \\ &= \Lambda A_z Y_z \end{aligned} \tag{7.70}$$

where the definitions of matrices Λ, A_z and Y_z are clear in the equation. Vector Y_z contains the corrections to the geopotential coefficients. Using the same argument for the relationship in (6.36), we can show that

$$(B_z^*)^T = CB_z^T = C\begin{bmatrix} \mathrm{IL}_1^1 & \mathrm{IL}_1^2 & \cdots & \mathrm{IL}_1^M \\ \mathrm{IL}_2^1 & \mathrm{IL}_2^2 & \cdots & \mathrm{IL}_2^M \\ & & \vdots & \\ \mathrm{IL}_\mu^1 & \mathrm{IL}_\mu^2 & \cdots & \mathrm{IL}_\mu^M \end{bmatrix} \tag{7.71}$$

where C is the matrix containing the combination coefficients e_{ij} in (5.76), B_z is a matrix containing the integrations of spherical harmonics. Combining equations (7.68) and (7.71), we get

$$X_z = \frac{1}{a_\sigma} CB_z^T \Lambda A_z Y_z \tag{7.72}$$

From the simultaneous solution, we can get the covariance matrix of Y_z, which is now denoted as Σ_y. From the definition of Y_z, Σ_y is the covariance matrix of the corrections to the geopotential coefficients. However, from the adjustment theory, we also know that Σ_y is the covariance of the complete coefficients (see Uotila, 1986, p. 64). Using error propagation, we can compute the error covariance matrix of X_z as

$$\Sigma_x = \frac{1}{a_\sigma^2} CB_z^T \Lambda A_z \Sigma_y A_z^T \Lambda B_z C^T \tag{7.73}$$

If we perform Cholesky decomposition for Σ_y, namely let $\Sigma_y = RR^T$, where R is a lower triangular matrix, then (7.73) becomes

$$\Sigma_x = \frac{1}{a_\sigma^2} E^T E \tag{7.74}$$

where

$$E = R^T A_z^T \Lambda B_z C^T \tag{7.75}$$

If we are just interested in the standard deviations of the ON coefficients $\hat{\alpha}_j^N$ then we have

$$\sigma_j = \left(e_j^T e_j\right)^{1/2} / a_\sigma \tag{7.76}$$

where e_j is the jth column vector of matrix E and σ_j is the standard deviation of $\hat{\alpha}_j^N$. The most difficult part in computing σ_j is the formation of matrix E in (7.75), which in turn requires the formation of matrix $A_z^T \Lambda B_z$. If we write

$$A_z^T \Lambda B_z = [P_{ij}]_{\nu x \mu} \tag{7.77}$$

where P_{ij} is an element of $A_2^T \Lambda B_2$, then P_{ij} has an "inner product" form of spherical harmonics over the oceans, since

$$P_{ij} = \sum_{\ell=1}^{M} \frac{1}{\Delta\sigma_\ell} \Pi L_i^\ell \Pi L_j^\ell \tag{7.78}$$

Eq. (7.78) is merely a formal expression in that the exact harmonics corresponding to indices i,j must be identified according to the definition of the orthonormal system {Zj} in (5.76). Note that in this chapter and in Chapters 5 and 6, we have consistently used the notation ℓ to indicate the block sequence of the ocean blocks and M to indicate the number of such blocks. After obtaining P_{ij} , the computation of matrix E is easy since R and C are simply two triangular matrices.

Having computed σ_j in (7.76), we then can compute the error degree variances of the ON coefficients of the undulation corrections. This can be done using the property shown in (2.23), even if correlations among the ON coefficients exist. By performing the ON expansion and finding the accuracy estimates of the coefficients, we are now able to compare the two major signals - the oceanic geoid and the SST in the same domain - the oceans. We thus can disregard whatever happens on land. The spectral comparison of the two signals both in the signal part and the error part should be more reasonable in that the disturbance from land values has been eliminated. Such an advantage holds not only in the case of using purely altimeter data but also in the case of simultaneously using altimeter data, surface gravity data, other tracking data, etc.

In view of the non-uniformity of the distribution of the geoid error (see Denker and Rapp, 1990, Figure 9) in the altimetric solutions (or even in the solutions with multiple-types of data), an error spectrum concerning only the domain of oceans should provide more realistic error assessment for the resulting signals (mainly the geoid and the SST), since now we are dealing with a unique domain - the oceans. However, due to two obvious reasons, correlations between the ON coefficients of the SST exist: (1) The domain formed by the altimeter data points cannot exactly match the domain of ON functions. (2) The SST is not the unique signal to be found in the simultaneous solution. The correlations introduced by the presence of other signals can be explained as follows: The normal matrix of the system consists of two parts, namely, the SST part and the part that includes geopotential coefficient corrections, empirical coefficients and geocentric shift. Formally, the normal matrix and its inverse can be written as:

$$N^{-1} = \begin{pmatrix} A^TA+P_1 & A^TB \\ B^TA & B^TB+P_2 \end{pmatrix}^{-1} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \tag{7.79}$$

where (B^TB+P_2) is the SST part, P_1 and P_2 are a priori weight matrices discussed in Section 7.2.2. Matrix Q_{22} provides the error covariance of the ON coefficients of the SST and can be analytically written as:

$$Q_{22} = \left(B^TB+P_2 - B^TA(A^TA+P_1)^{-1}A^TB\right)^{-1} \tag{7.80}$$

Thus even if $(B^TB+ P_2)$ is close to a diagonal matrix, the presence of the other matrix in (7.80) will destroy such a characteristic. However, later on we will find that the correlations among the ON coefficients are only concentrated at the low-degree parts (lower than degree 10), thus a signal decomposition of the SST from the simultaneous solution is still possible after a certain degree (to be discussed later).

7.4 The Altimeter Data Used in the Solutions

The altimeter data to be used in the experimentally simultaneous solutions, with emphasis on the applications of the ON functions, are the Geosat data. Starting from the GDR's in Cheney et al. (1987), the Geosat have evolved from the low-accuracy NAG orbits that used the GEM10 potential coefficients as the force model, to today's GEMT2 orbits (Haines et al., 1990). The original GDR orbits (in Cheney et al., 1987) have radial accuracies of 4 m while GEM-T2 are claimed to have radial accuracies of 35 cm. The orbits used by Denker and Rapp (1990) were based on GEMT1 which, according to Haines et al. (1990), have radial accuracies of 85 cm.

The starting orbits to be used in this study are the ones based on the GEM-T2 orbits (Haines et al., 1990). In addition to the U.S. NAVY's OPNET tracking system, DMA's TRANET tracking system has also been incorporated in the orbit determinations that lead to the GEM-T2 orbits to be used here. Unlike the GEM-T1 orbit's 17 day arcs, the GEM-T2 orbits are determined on a 6-day basis with roughly one-day overlap (Koblinsky et al., 1990). For the purpose of experiments, 6 6-day arcs will be used. These 6 arcs have a duration of 2 ERMs, or roughly 32 days and are suitable for testing some factors that will be described later. The arc numbers (defined by Koblinsky et al., 1990) and the start/stop times of the six arcs are listed in Table 7.2.

Table 7.2 Start/stop times of the 6 GEMT2 arcs used for experiments

Arc #	Start time*	Stop time*	No. of normal points
22	870220-1103	870226-0000	10598
23	870225-0000	870303-0000	11134
24	870302-0000	870308-0000	11528
25	870307-0000	870313-0000	10864
26	870312-0000	870318-0000	10718
27	870317-0000	870323-0000	11280

* Year-month-day-hour-minute

These arcs are created at 1 minute time intervals while the original GDR's have data spaced at 1 second intervals so that interpolations are needed to replace the original orbits in the original GDR's (those by Cheney et al., 1987). For the simultaneous solution, data editing is required before the simultaneous solutions are made.

Normally altimeter data editing for a simultaneous solution consists of two steps. In the first step, we can eliminate the spurious altimeter data existing in shallow-water areas, land, etc. Also, altimeter data over some rough-gravity field areas will not be suitable for generating observables since current high degree models may not be able to remove the "omitted" undulation ΔN. Numerous other factors that should be considered may be found in Denker and Rapp (1990). These factors could serve as editing criteria and the altimeter data to be used in the simultaneous solutions must pass the criteria first. On

the other hand, if all the accepted data (after testing against the criteria) are used, the resolution implied by such data will be higher that implied by the modeled geopotential field (say, to degree 50) and the SST field (say, to ON degree 24), resulting in a waste of computer time. So the second step for editing is to compress the raw data into "normal points" (Denker and Rapp, ibid.). A normal point in this study is selected as follows: A linear model was least-squares fitted to 20 successive altimeter points and the residuals are found. Then a 3σ criterion is used for an iterative outlier rejection (see Denker, 1990, p. 13 and Cheney et al., 1987, p. 9). The convergence of such iterations is achieved when no more points are rejected. Then, the normal points values are computed from the linear fit at the central time of the interval.

Since the experiments on the simultaneous solutions were done in parallel with the research carried out by Rapp et al. (1991), the editing criteria for the Geosat data adopted in this study are the same as those used by Rapp et al. (ibid, pp. 3-5). Such edited raw altimeter data are then used to generate the 20-second normal points by the algorithm stated above. The number of 20-second normal points of the six selected Geosat arcs can be found in Table 7.2. Figure 7.3 shows the point distribution of the 20-second normal points from arcs 22, 23 and 24. One particular issue is related to the treatment of the permanent tidal effects. Following Rapp's (1989c) suggestion, in order to obtain the "mean geoid" consistent with the definition of the SST to be determined here, we need to add a correction of 9.33×10^{-9} to the J_2 coefficient from the GEM-T2 field to get the so-called "zero" geoid undulation, and add

$$\Delta N_2 = -0.198(3/2 \sin^2\phi - 1/2) \qquad \text{(meters)} \qquad (7.81)$$

to get the mean geoid (i.e., N_c in Figure 7.2). See also Rapp (ibid.) for more details.

As shown in Figure 7.3, the data in the Mediterranean Sea have been edited out, while in the full-year solution of Rapp et al. (ibid), some selected data in that area have been used.

7.5 Numerical Experiments of Simultaneous Solutions and Results

7.5.1 Experiments

In connection with the models of simultaneous solutions presented in Section 7.2.2 and the data described in Section 7.4, we shall classify the experiments by first defining the types of solution:

a. 1-arc solution: using data of one 6-day arc
b. 3-arc solution: using data of arcs 22, 23, 24 or 25, 26, 27 (each solution covers 17 days). The information about these arcs has been given in Table 7.2.
c. 6-arc solution: using data of arcs 22, 23, 24, 25, 26 and 27 (32 days)

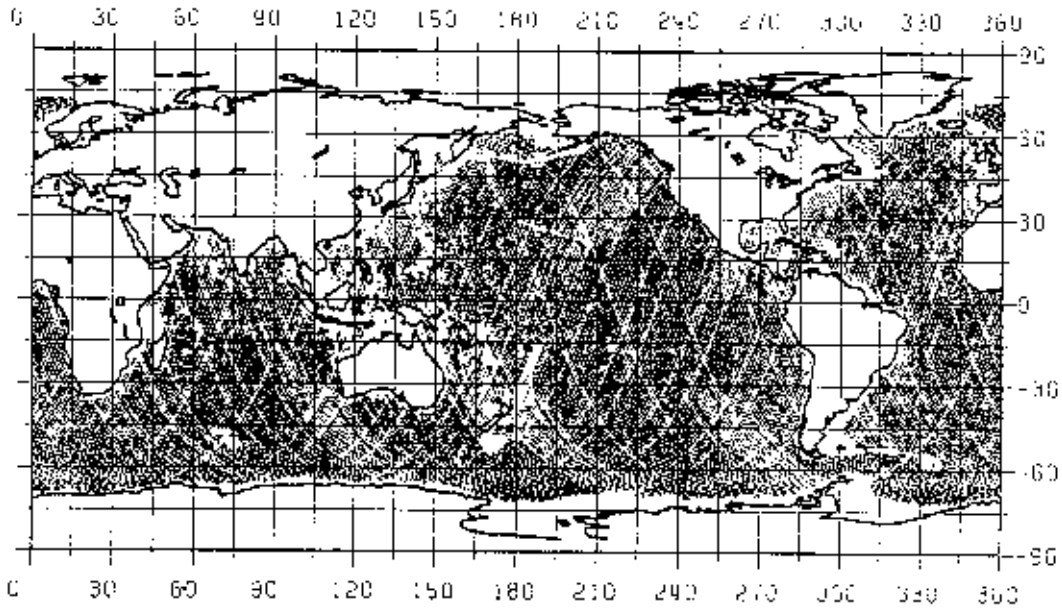


Figure 7.3 Distribution of 20-second Normal Points Form Arcs 22, 23, and 24.

Then the factors considered in the various experiments are:

a. Geopotential coefficient corrections $\overline{\Delta C}_{nm}$, $\overline{\Delta S}_{nm}$ for GEM-T2 field to degree 50 and orbit errors of gravitational origin and due to initial state vector error (2595 potential coefficients + (8 arc coefficients) x number of arcs)
b. Orthonormal function representation of SST to degree 24 (621 coefficients)
c. Spherical harmonic representation of SST to degree 15 or 24
d. Geocentric shift of the orbit coordinate system (ΔX and ΔY)
e. Degree variances of the ON coefficients from the Levitus SST as "mild constraint"
f. Degree variances of the spherical harmonic coefficients from the Levitus SST as "mild constraint"
g. Downweighting altimeter normals
h. Use of surface gravity anomalies
i. Use of the Levitus SST as additional observables in the form of the normal equations of the ON coefficients (see (7.62) and (7.63))
j. GEM-T2 error covariance matrix

Regarding factor a, the starting potential field, GEM-T2, are only complete to degree 36, with extended coefficients to degree 50. Thus the corrections ΔC_{nm}, ΔS_{nm} from the simultaneous solution for the absent coefficients are not really "corrections" but the complete coefficient. The use of the spherical harmonic representation of SST is for the purpose of comparison, since the spherical harmonics are currently the most popular basis functions for doing SST expansion. The downweighting process is now almost a standard practice in the simultaneous solutions, since it is believed that the long-wave length characteristics of the satellite model (now GEM-T2) will not be destroyed by doing this and also the altimeter data are basically repeated in space if more than one ERM data are used. A theoretical justification of using downweighting for various data may be found in Lerch (1991). In Engelis and Knudsen's (1989) treatment of downweighting, the best result is achieved if a factor of 1/4 was used for one 17-day Seasat arc, while Rapp et al. (1991) used 1/96 for 24 ERM data to obtain an optimum solution. Thus both seemed to downweight the one 17-day arc by a factor of 1/4. In the current study, various downweighting factors will be tried and the results will be described in the next section.

The use of surface gravity data, in the form of the normal matrix and the "U" vector, is increasingly popular and even is thought to be necessary for separating the geoid and the SST. Furthermore, the use of gravity data tends to avoid obtaining a "tailored" geopotential model over the oceans, since we now have data related to earth's potential field on land. However, it is also possible that the gravity data may deteriorate locally the oceanic geoid if the accuracies of the gravity data are not comparable with the altimeter data. Overall speaking, more geophysically meaningful geopotential coefficients should be obtained by incorporating the gravity data in the simultaneous solutions. For the experiments performed in this section, the surface gravity normal equations are formed by using the gravity anomalies continued to the ellipsoid, denoted as "V2" solution in Rapp et al. (1991, pages 14, 22, 23). The gravity data distribution can be found in Figure 2 of Rapp et al. (ibid.). The V2 solution was also used in the final solution FYS10.W96GW in Rapp et al.

As stated above, some coefficients above degree 36 are absent in the GEM-T2 model, therefore a priori information is needed to augment the original GEM-T2 error covariance matrix. To do this, we first invert the original error covariance matrix to get the weight matrix. Then the minimum value of the diagonal elements of the weight matrix is found. Finally a value that is 100 times smaller than this minimum value is assigned to the diagonal elements of the augmented weight matrix that correspond to the absent coefficients. The off-diagonal elements pertaining to these absent coefficients are simply assumed to be zero. This procedure will give more freedom for the absent coefficients to be adjusted since their weights are relatively small. Although this is not quite justifiable it is a necessary step to get a "complete" error covariance matrix to degree 50. Note that the degree 0, 1 terms and the C_{21}, S_{21} terms are not included.

For the degree variances of the ON coefficients and spherical harmonic coefficients that are needed for constraining the SST coefficients, we shall use the results from the analyses of the Levitus SST (see Section 6.4.1). These values are listed in Table 7.3. For the ON functions, we shall use system $\{Z_j\}$ which have no degree 0 and 1 terms. The degree variances of spherical harmonics in Table 7.3 are obtained from a global analysis assuming SST = 0 on land (solution shnqto24 in Section 6.4.1). Therefore, these values are different from the degree variances used by Rapp et al. (1991, Table 5). This is done because we will model the SST in the experiments to spherical harmonic degree 24 and the a priori degree variances of spherical harmonics of decreasing magnitudes (especially after

degree 10) are possible only through a global analysis (see Chapter 6 for more discussions).

Table 7.3 Square Roots of the SST a priori Degree Variances for Conditioning the Simultaneous Solutions in Meters

Degree	ON*	SH†	Degree	ON	SH
1		0.182	13	0.029	0.043
2	0.430	0.221	14	0.020	0.044
3	0.115	0.094	15	0.017	0.030
4	0.176	0.071	16	0.017	0.029
5	0.120	0.072	17	0.015	0.029
6	0.100	0.137	18	0.017	0.022
7	0.076	0.109	19	0.014	0.021
8	0.081	0.076	20	0.013	0.026
9	0.063	0.061	21	0.011	0.023
10	0.034	0.030	22	0.010	0.025
11	0.028	0.036	23	0.010	0.022
12	0.029	0.051	24	0.010	0.021

* : Orthonormal functions, System $\{Z_j\}$.
† : Spherical harmonic functions.

Considering the types of solutions and factors, we divide the experiments into five categories, as shown in Table 7.4. This is for a better explanation of the results in the next section and for finding the optimum solution strategy that can be recommended for future work.

Table 7.4 Categories of Simultaneous Solutions

Category	Factors	Types of solutions	Remarks
I	a,b,d,e,g,j	1-arc, 3-arc, 6-arc	standard solution
II	a,b,e,g,j	1-arc, 3-arc, 6-arc	w/o geocentric shift
III	a,b,d,e,g,h,j	1-arc, 3-arc, 6-arc	standard sol. + Δg
IV	a,b,d,e,i,j	1-arc, 3-arc, 6-arc	standard sol. + SST ("obs")
V	a,c,f,g,h,j	1-arc, 3-arc, 6-arc	spherical harm. rep. of SST

7.5.2 Results

7.5.2.1 SST Models from the Solutions and Geostrophic Currents

In this section we will present the SST models from the simultaneous solutions and the geostrophic currents implied by these models.

In Category I, we basically follow the strategy used in Denker and Rapp (1990) where only Geosat altimeter data were used. As proved in (ibid.), the goal of satellite radial orbit error reduction and geoid-SST determination can still be achieved with promising results in such a strategy. Figure 7.4 shows the SST to ON degree 15 from the ON (24, 24) 6-arc solution with altimeter normals downweighted by 1/6 (to be discussed later). Although such a solution strategy yields a reasonable signature (this can be visually inspected from Figure 7.4) a serious problem found is the extremely large standard deviations of geoid undulations (larger than 10 meters) on land. This is due to the lack of

observations on land. Such excessively large standard deviations were reduced when incorporating surface gravity data as in Category III.

The difference between the strategies in Category I and Category II is the elimination of ΔX and ΔY in Category II. Figure 7.5 shows the SST to ON degree 15 from the ON (24, 24) 6-arc solution with the same downweighting factor as used for the SST in Figure 7.4. Figure 7.6 shows the difference between the SST in Figure 7.4 and the SST in Figure 7.5. The difference between these two sets is of long wavelength nature and is attributed to the differences in the low degree ON coefficients which may be caused by solving for ΔX and ΔY in one of the solutions. Table 7.5 show the RMS differences by degree (see eq. (6.102)) between these two sets of SST. In general the SST from Category I and Category II yield almost the same SST signatures. However, one could question that how reliable the solved parameters ΔX, ΔY are, due to the complex simultaneous model and the poor observation geometry (see the discussion in Section 7.2.1). One way to judge the reliability of ΔX and ΔY is to investigate the correlations between these two components and the SST coefficients and other parameters. This will be discussed in Section 7.5.2.3.

Table 7.5 RMS Differences by Degree Between the SST in Figure 7.4 and the SST in Figure 7.5 Using ON Functions in Meters

Degree	diff.	Degree	diff.
2	0.044	13	0.001
3	0.039	15	0.000
4	0.036	18	0.000
5	0.027	20	0.000
6	0.019	24	0.000

In Category III, we incorporated the gravity normal equations (see the previous section for the exact normal equations that are used). As compared to the strategy used in Category I, the standard deviations of geoid undulations on land in Category III have been significantly reduced (see later discussion). Figure 7.7 shows the SST to ON degree 15 from the ON (24, 24) 6-arc solution with altimeter normals downweighted by 1/6. The comparison between the SST ON coefficients in Category I and Category III (the cases without/with surface gravity anomalies) show that the cumulative difference is 6.1 cm up to ON degree 15 and is 6.2 cm up to ON degree 24, indicating that the large differences in the ON coefficients are concentrated at the low degree terms. In general the SST signatures from the solution with/without gravity data (those in Figure 7.4 and Figure 7.7) agree well, but deviations can be found in the Kuroshio Current, the Peru Current, and some other areas. An experiment was made to downweight the gravity normal equations by 1/2 in such a way

$$\begin{bmatrix} N_1 + \frac{1}{2}N_g & N_{12} \\ N_{21} & N_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} U_1 + \frac{1}{2}U_g \\ U_2 \end{bmatrix} \tag{7.82}$$

where X_1 contains the geopotential coefficients, X_2 contains other parameters including the SST coefficients, N_g is the gravity normal matrix and U_g is the gravity "U" vector. Figure 7.8 shows the SST resulting from such a downweighting factor. The differences between the SST in Figure 7.7 and Figure 7.8 amount to 2.2 cm at ON degree 15 and to 2.4 cm at

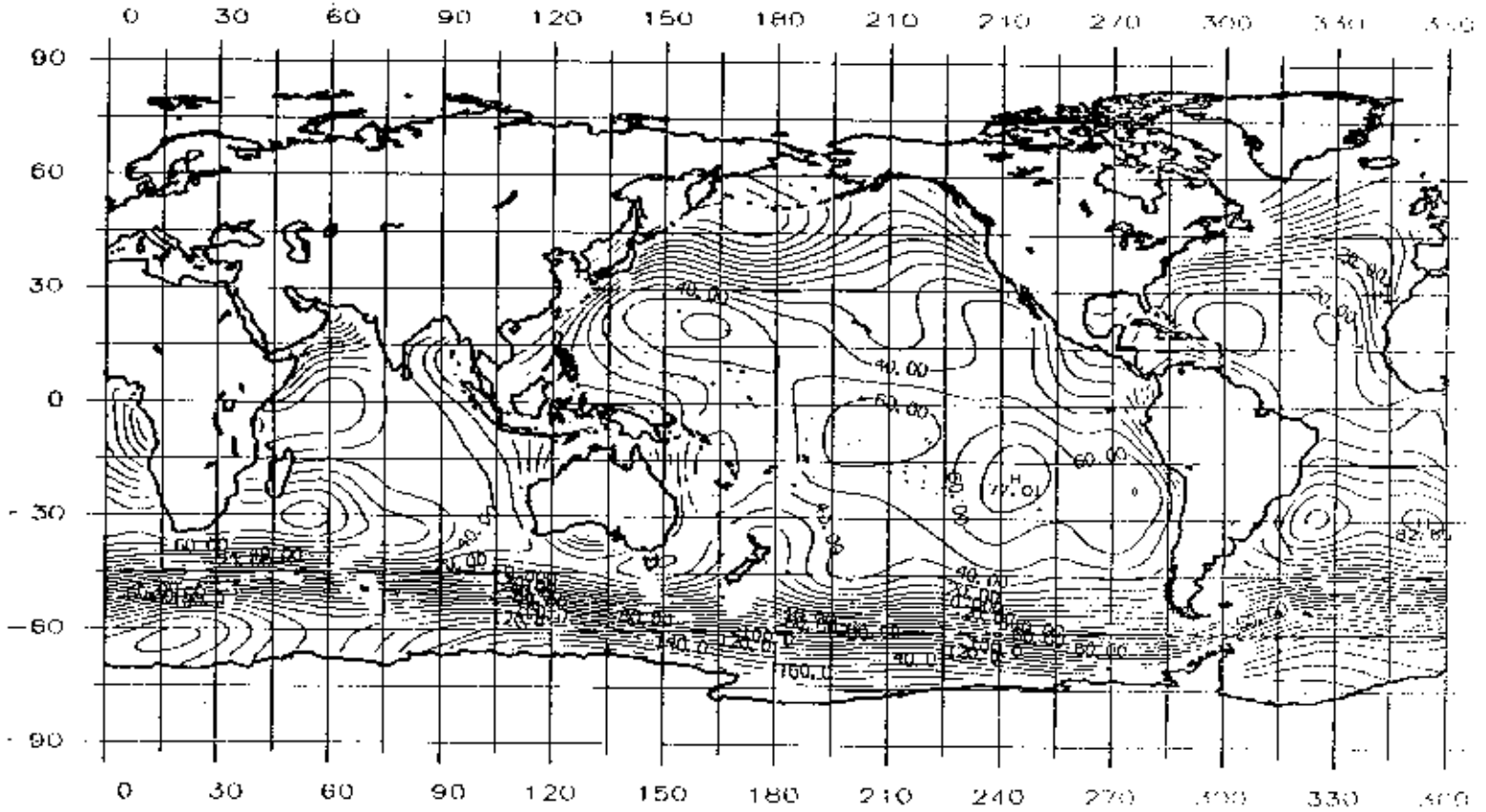


Figure 7.4 SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Category I, CI = 10cm

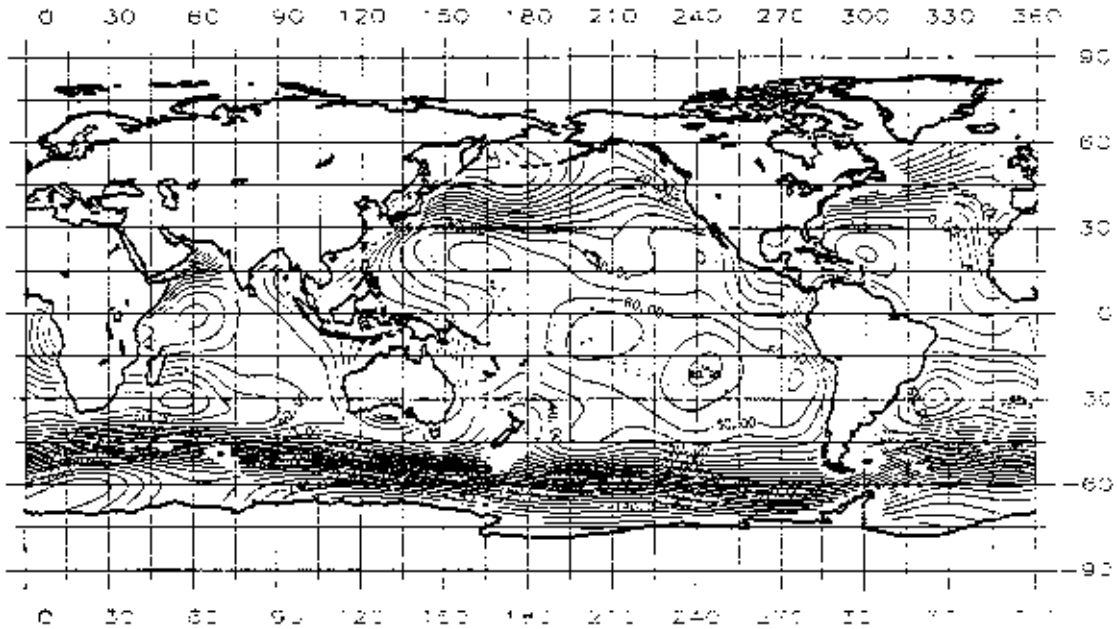


Figure 7.5 SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Category II, CI = 10cm

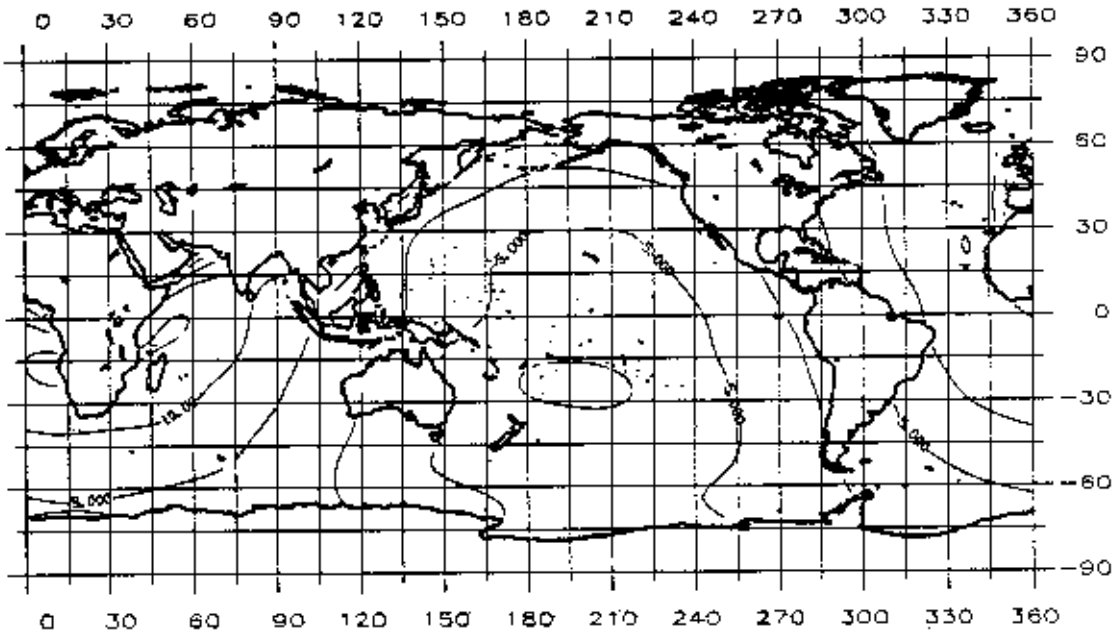


Figure 7.6 Difference Between the SST from ON (24, 24) 6-arc Solution, Category I and ON (24, 24) 6-arc Solution, Category II

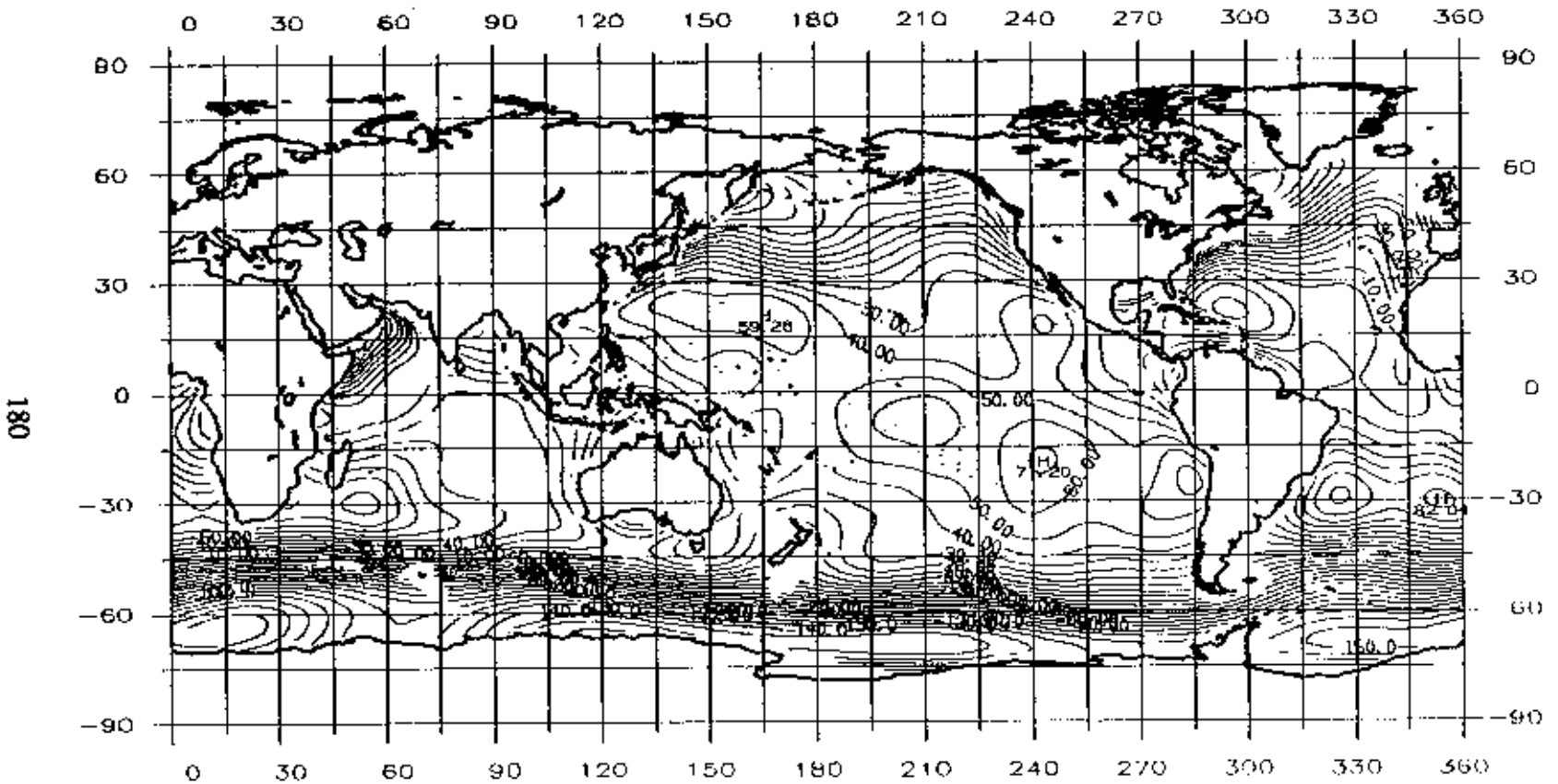


Figure 7.7 SST to ON Degree 15 from ON (24, 24) 6-arc Solution, No Downweighting to Gravity Normal, Category III, CI = 10cm

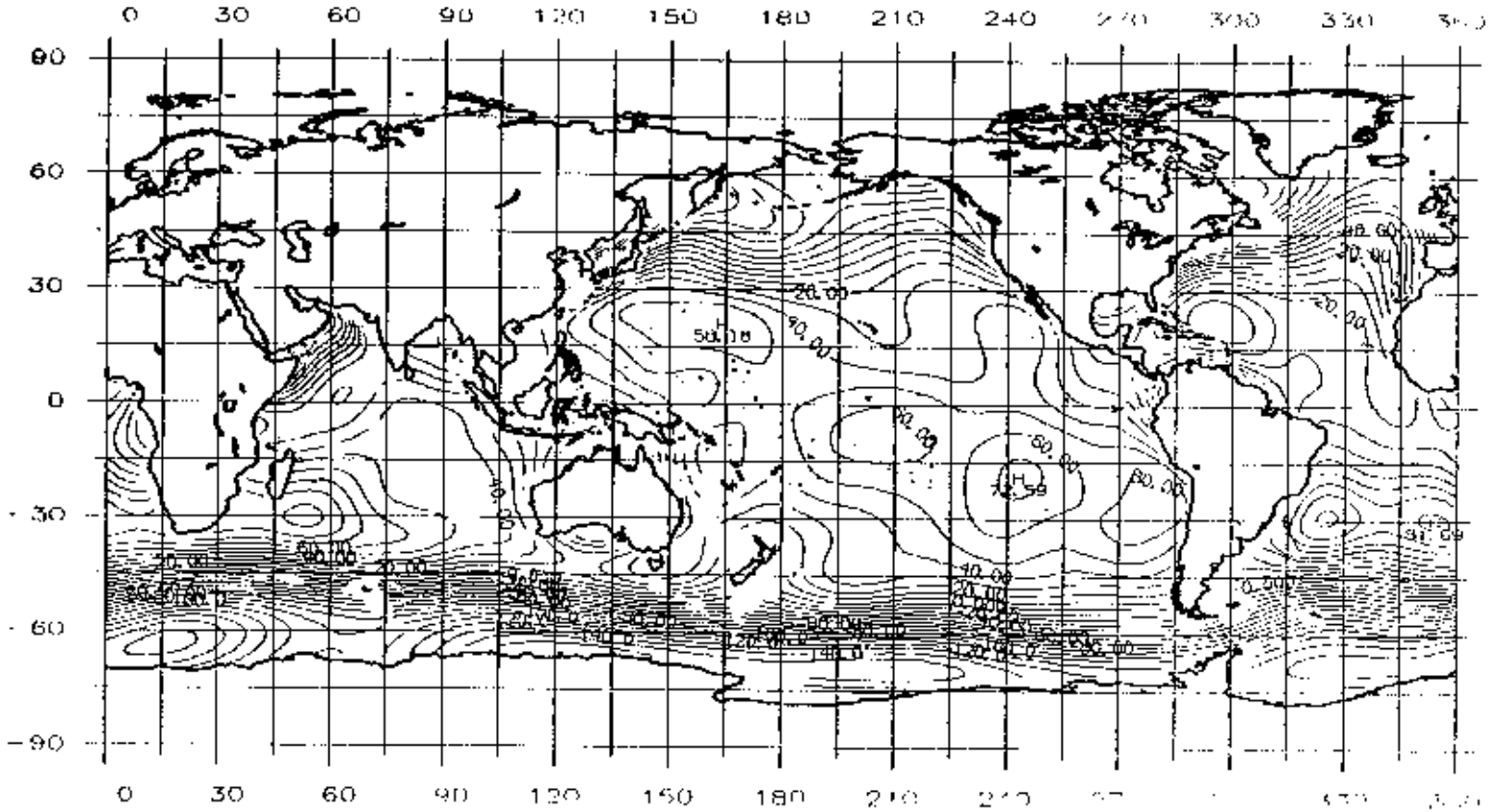


Figure 7.8 SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Downweighting Gravity Normal by 1/2, Category III, CI = 10cm

ON degree 24 and thus the change of SST due to the downweighting of the gravity normal is about 4% of the signal up to ON degree 24 (the RMS SST value in Category III is 0.58 meter based the ON coefficients). Also, based on the ON function analysis of the geoid undulation error (see Section 7.3), we found that the increase in the geoid undulation errors over the oceans is 0.5 cm up to ON degree 24 (about 5% increase) when using the factor 1/2. In addition, the crossover analysis (see Section 7.5.2.5) using the 6 Geosat arcs shows that exactly the same RMS crossover discrepancies can be obtained from using downweighting and from not using downweighting. Therefore, the downweighting factor 1/2 to the gravity normal seems to have no substantial impact on the solution.

The idea of using oceanic equations of motion as constraints for separating the geoid and the SST have long been proposed, but not been tried before (see the discussion in Section 7.2.2 on how to use the oceanic equations of motion). In Category IV, we try a "modified" idea of using the oceanic equations by incorporating the Levitus SST as additional observables in terms of normal equations of the ON coefficients (see (7.65)). This is done by analogy with the surface gravity data which are treated as additional observables related to the gravity field. Figure 7.9 shows the SST to ON degree 15 from ON (24, 24) 6-arc solution in Category IV. For the SST shown in Figure 7.9, the Levitus SST are assumed to have a uniform accuracy of 25 cm and a downweighting factor of 1/4 is applied to the SST normal equations (for the reason see Section 7.2.2). The assumption of 25 cm uniform accuracy for the Levitus SST is somewhat unrealistic, but it is so determined simply due to the fact that the accuracies of the Levitus SST are hard to obtain because of the level of motion in the oceans and some other factors. A somewhat "unfortunate" situation is that the resulting SST from the simultaneous solution are too "close" to the Levitus SST: the cumulative RMS difference between the Geosat SST in this category and the Levitus SST is 7 cm up to ON degree 15 and 9 cm up to ON degree 24. The small RMS differences show that an independent SST model cannot be obtained if the Levitus SST are treated as additional observables in the simultaneous model. However, based on the correlation analysis in Section 7.5.2.6, extremely low correlations (below 0.05) between the above mentioned parameters were found in this category and this suggests that the oceanic equations of motion should have great potential in efficiently separating the geoid and the SST and in decorrelating the parameters of interest. Oceanographers may provide a better idea of how exactly the oceanic equations of motion can be used in the simultaneous solution.

In comparison to the use of the ON functions, we use the spherical harmonic functions as the basis functions for the SST in the simultaneous solutions in Category V where both Geosat data and the gravity data were used (as in Category III). Figure 7.10 shows the SST from the spherical harmonic (15, 15) solution using the 6 Geosat arcs. The SST in Figure 7.10 can be compared to the SST in Figure 7.7. Figure 7.11 shows the SST differences between these two SST models. From Figure 7.11, we find that the large difference occur in the coastal areas. Over the oceans, the RMS difference between these two models is 13 cm, the mean difference is -0.7 cm and the maximum difference is 83 cm.

Using the SST models discussed above, we have computed the geostrophic currents using (6.97) and (6.98). Due to limited space, only selected current plots are presented. Figure 7.12, 7.13 and 7.14 show the geostrophic currents implied by the SST models in Figure 7.4, 7.7 and 7.10 respectively. Of particular interest is the differences between the current patterns in Figure 7.13 and Figure 7.14, which correspond to the solutions using the ON functions and the spherical harmonic functions respectively (Δg were also used in both models). Overall speaking, the use of the ON functions yields more

realistic current patterns in the coastal areas, e.g., in the Peru Current. However, in general the current patterns implied by these two models agree well in the open oceans. It should be noted that the ON function representation of the SST not only provides a more realistic current flow, as shown in Figure 7.12 and Figure 7.13, but also gives more theoretically justifiable basis for the signal/error spectral analyses for the results from the simultaneous solutions, as will be carried out in Section 7.5.2.4. Further comparisons on the results from the use of the ON functions and the use of the spherical harmonic functions will be made in later development.

A comment will be made on the downweighting factors for the altimeter normals. The problem of determining the optimum downweighting factor for the altimeter normals in the simultaneous solutions has been studied in detail by Denker and Rapp (1990) and Rapp et al. (1991). Denker (1990, p. 25) also showed numerically the necessity of the downweighting factor using the calibration technique developed by Lerch (1985). Rapp et al. (1991) studied the problem in greater detail and concluded that a factor of 1/96 is the optimum number for downweighting the altimeter normals from 24 Geosat ERM data. The factor used by Rapp et al. (ibid.) thus implied a factor of 1/4 for each ERM data. These analyses have provided useful information on the downweighting factor in the solution performed here. Various downweighting factors ranging from 1/2 to 1/8 for the normal from the 6 arcs were tested and it was found that the factor 1/6 yields the most oceanographically meaningful SST signature and the best-defined ocean currents. Although such a decision is somewhat "subjective", it still relies on the results from the previous investigators such as Denker and Rapp (1990). A more "objective" approach of determining the optimum downweighting factor should be based on the procedure in Rapp et al. (1991). Due to such a decision, we have used 1/6 as the downweighting factor for all the 6-arc solutions and 1/3 for the 3-arc solutions discussed in this section.

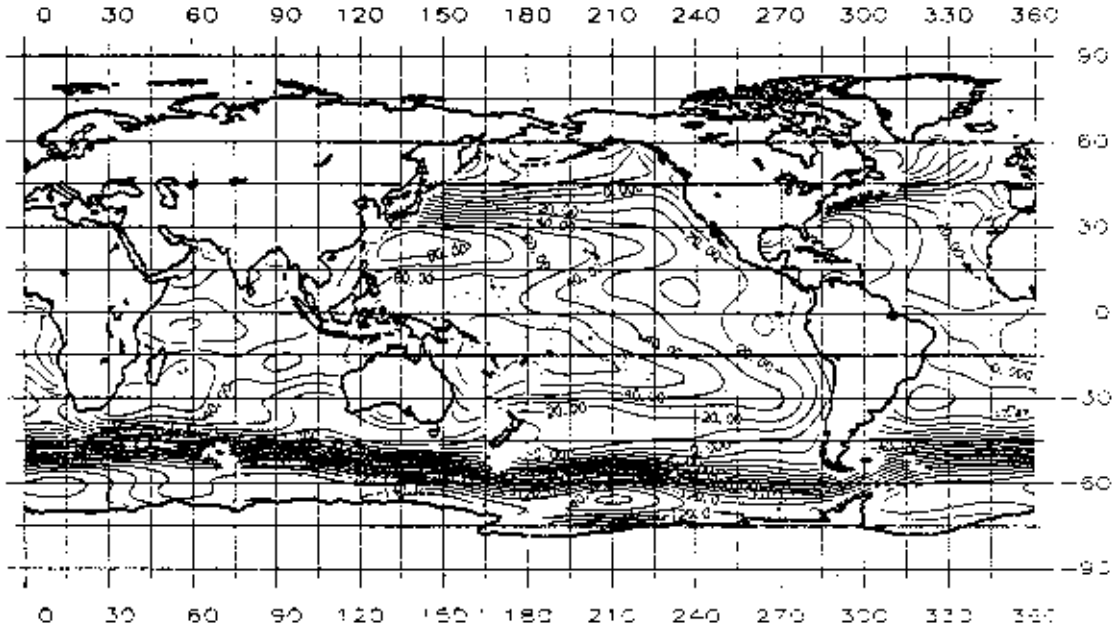


Figure 7.9 SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Downweighting Levitus's Normal by 1/4, Category IV, CI = 10cm

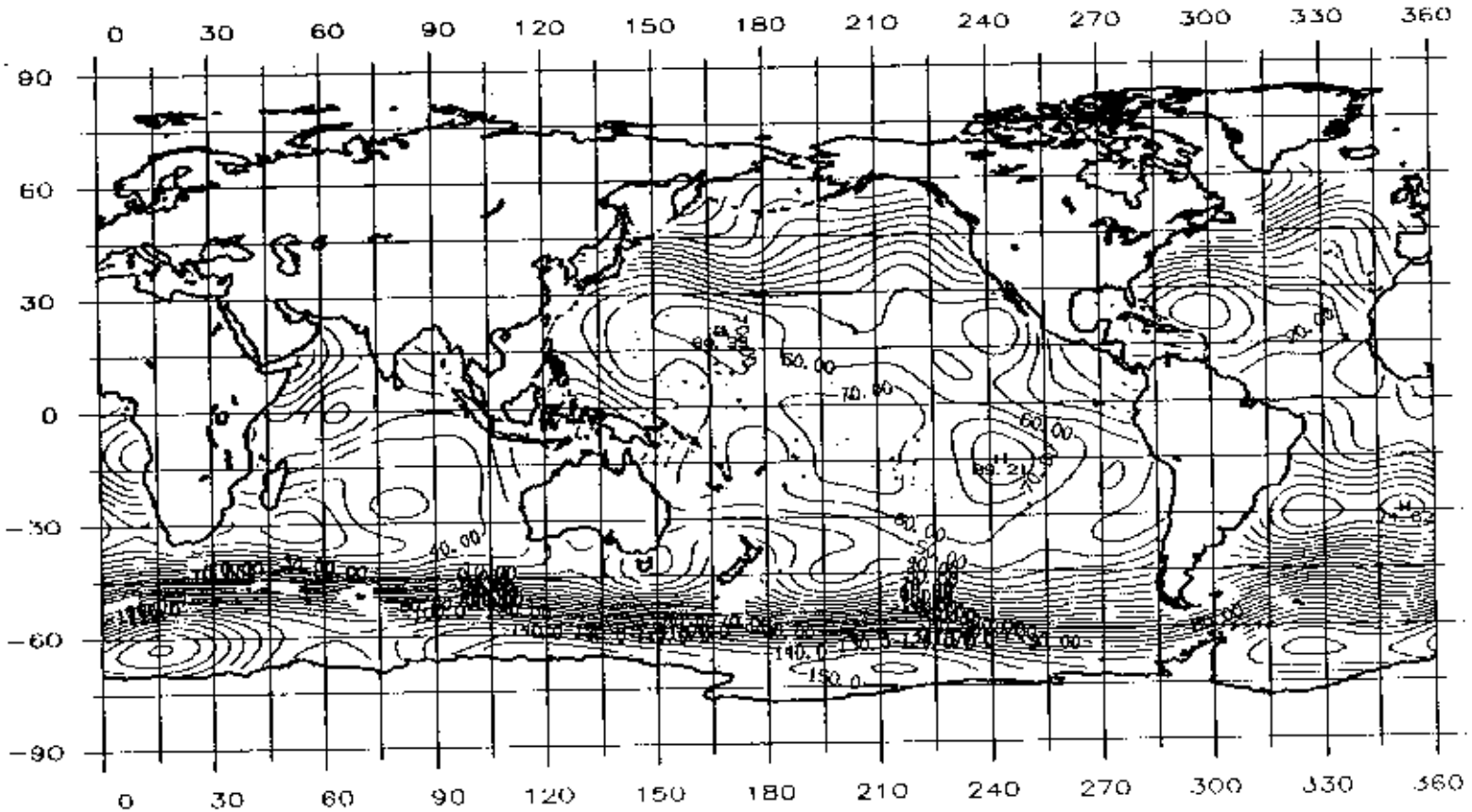


Figure 7.10 SST from Spherical Harmonic (15, 15) 6-arc Solution, Category V, CI = 10cm

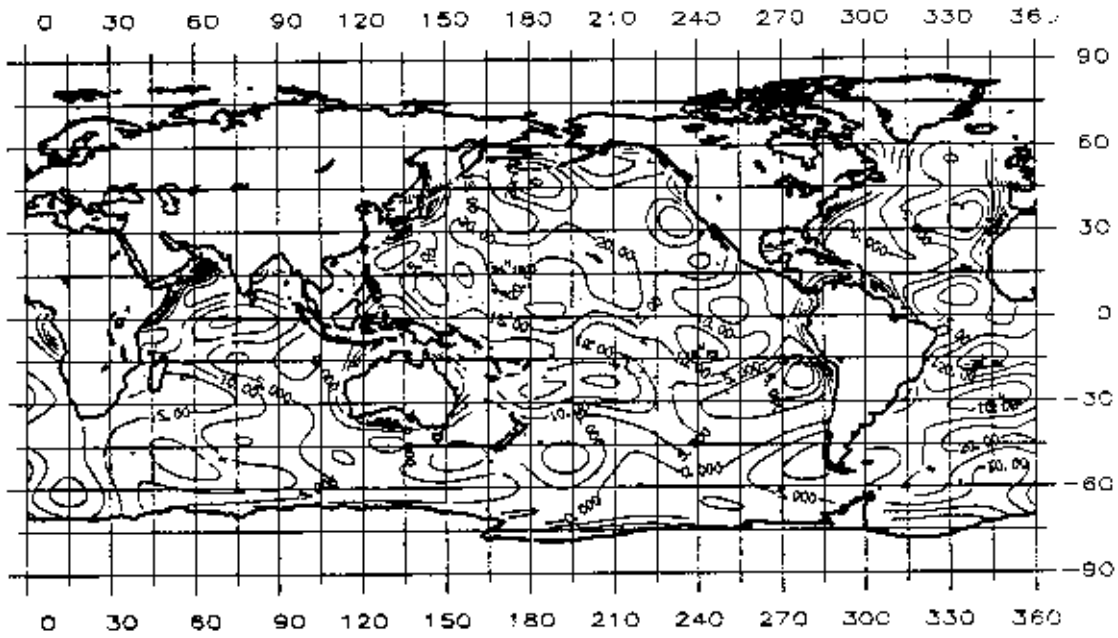


Figure 7.11 Difference Between the SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Category III and the Spherical Harmonic (15, 15) Solution, Category V, CI = 5 cm

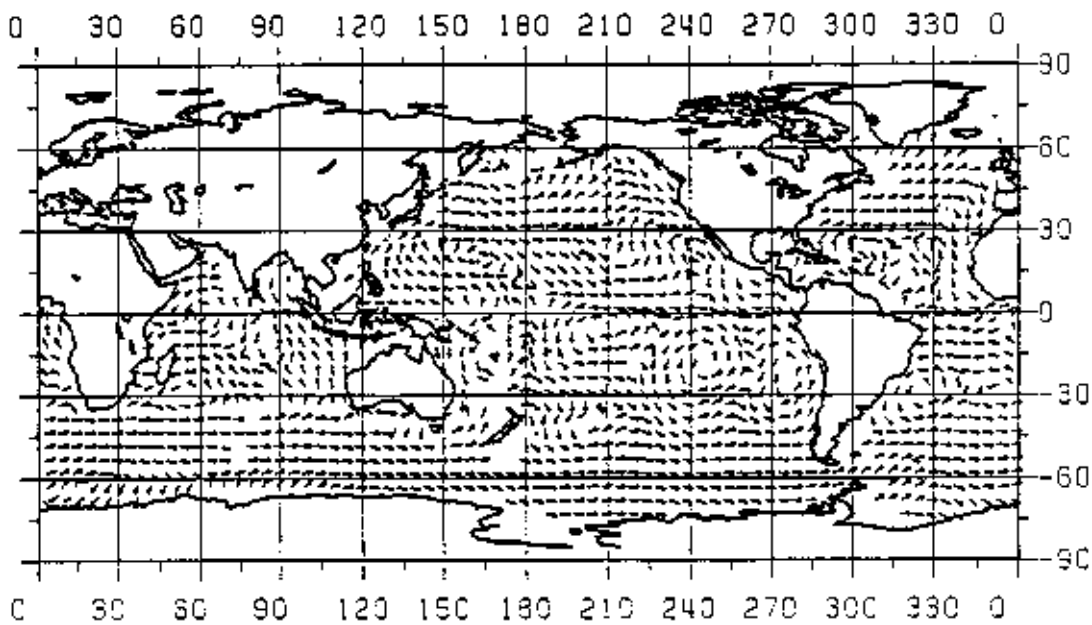


Figure 7.12 Geostrophic Currents Implied by the SST to ON Degree 15 from ON (24, 24) 6-arc Solution, Category I, CI = 10cm

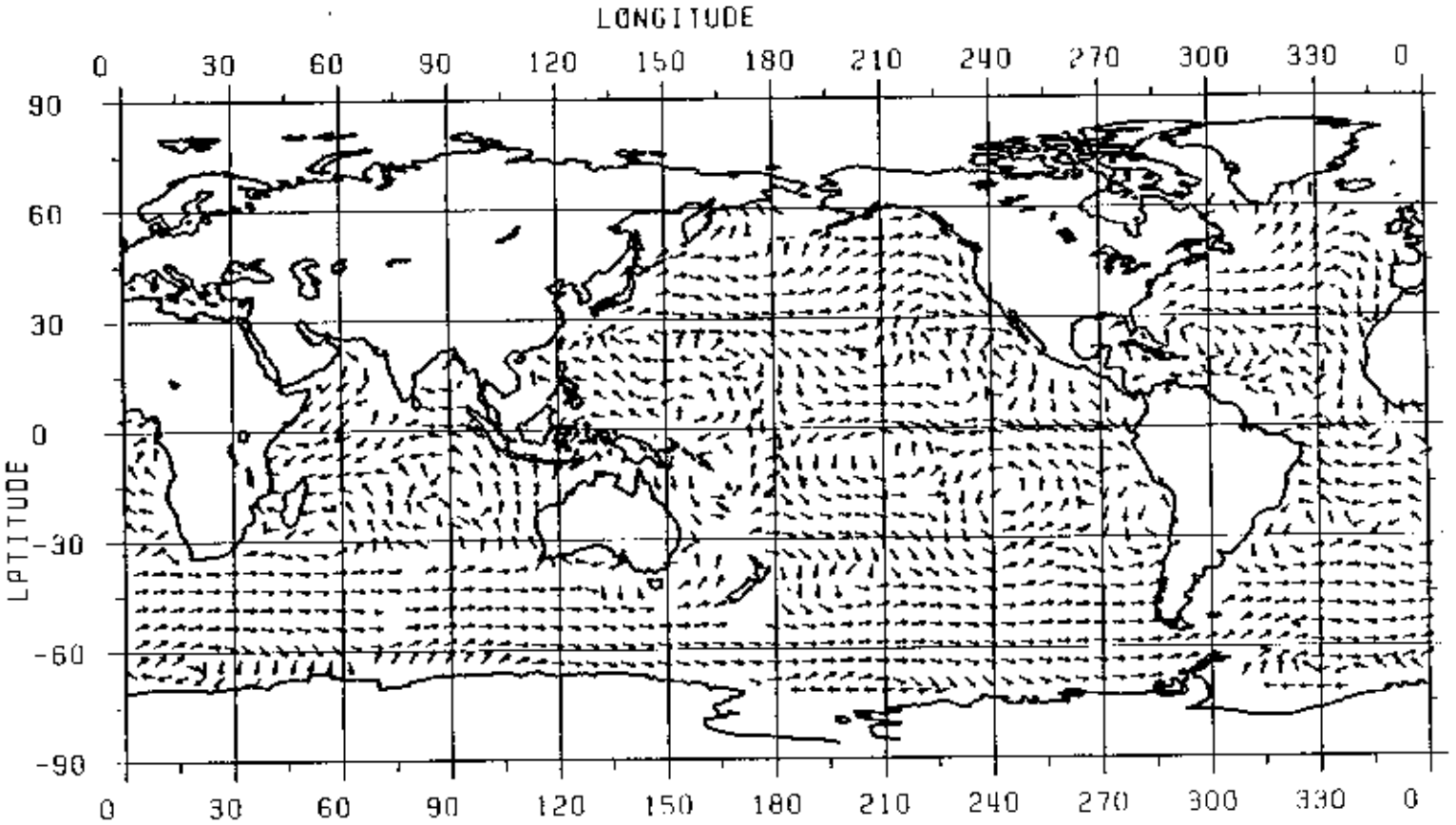


Figure 7.13 Geostrophic Currents Implied by the SST to ON Degree 15 from ON (24, 24) 6-arc Solution, No Downweighting to Gravity Normal, Category III, CI = 10cm

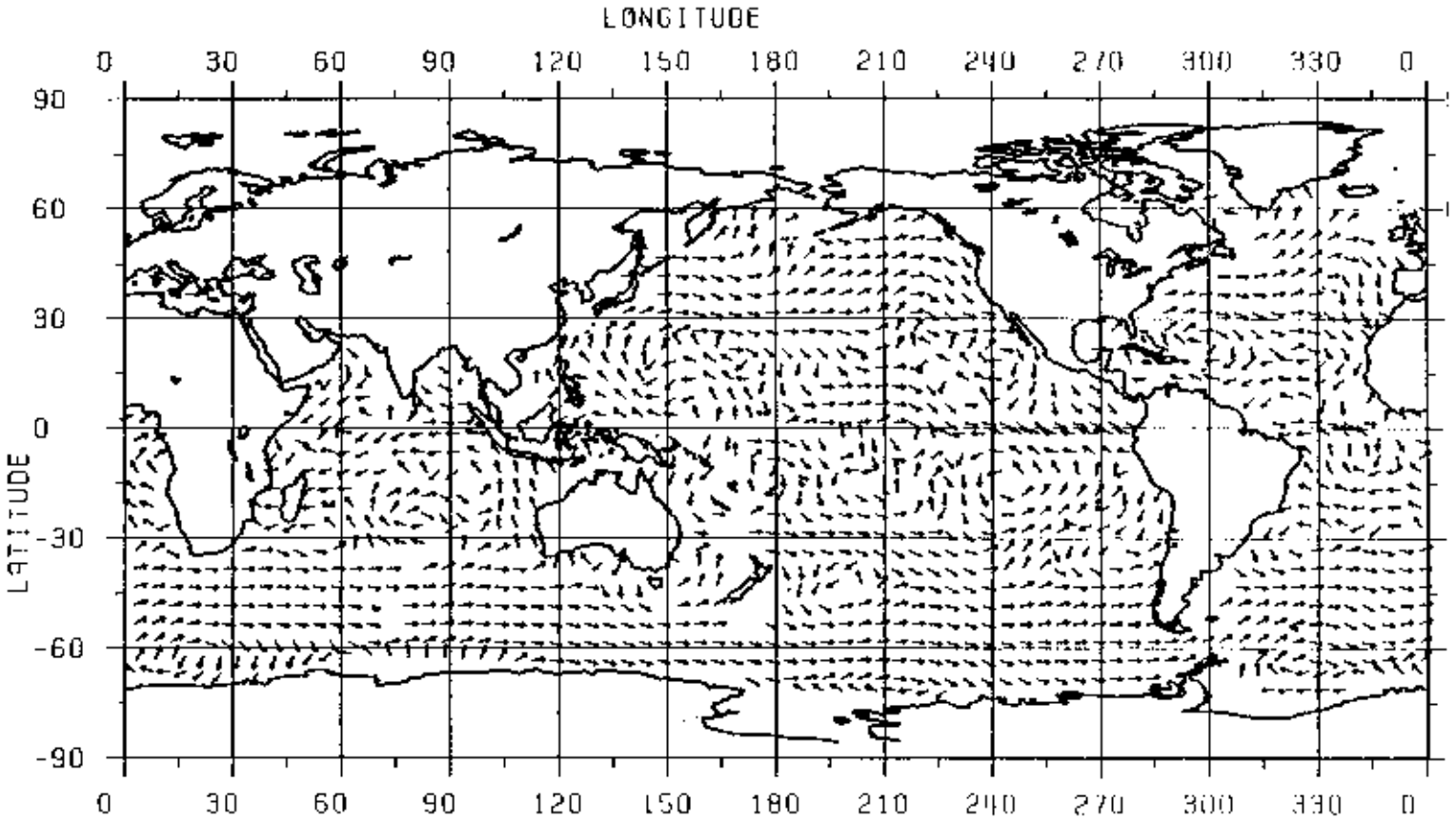


Figure 7.14 Geostrophic Currents Implied by the SST from Spherical Harmonic (15, 15) 6-arc Solution, Category V, CI = 10cm

7.5.2.2 Comparison with the Levitus SST and The One-year Geosat Solution

It is customary to compare the SST from the altimetric solutions to the hydrographic data. In Table 7.6, we list the correlations by degree between the five 6-arc solutions in Category I to V and the Levitus SST. The correlation by degree is computed according to the definition of correlation in (6.66). Specifically, let two signals at degree n be expressed as

$$f_{1n}(\theta, \lambda) = \sum_{m=0}^{n} \left(\hat{\alpha}_{1nm}\hat{O}_{nm}(\theta, \lambda) + \hat{\beta}_{1nm}\hat{Q}_{nm}(\theta, \lambda)\right)$$
$$f_{2n}(\theta, \lambda) = \sum_{m=0}^{n} \left(\hat{\alpha}_{2nm}\hat{O}_{nm}(\theta, \lambda) + \hat{\beta}_{2nm}\hat{Q}_{nm}(\theta, \lambda)\right) \tag{7.83}$$

then the correlation at degree n for the ON coefficients is

$$\rho_n = \frac{\int_\sigma f_{1n} f_{2n} d\sigma}{\left(\int_\sigma f_{1n}^2 d\sigma \int_\sigma f_{2n}^2 d\sigma\right)^{1/2}} \tag{7.84}$$
$$= \frac{\sum_{m=0}^{n} \left(\hat{\alpha}_{1nm}\hat{\alpha}_{2nm} + \hat{\beta}_{1nm}\hat{\beta}_{2nm}\right)}{\left[\sum_{m=0}^{n} \left(\hat{\alpha}_{1nm}^2 + \hat{\beta}_{1nm}^2\right) \sum_{m=0}^{n} \left(\hat{\alpha}_{2nm}^2 + \hat{\beta}_{2nm}^2\right)\right]^{1/2}}$$

where σ is the oceans. If we replace σ by the entire earth and the ON functions by the spherical harmonics, the definition (7.84) can also be used for the correlation at degree n for the spherical harmonic coefficients.

From Table 7.6, it is found that in all the 6-arc solutions the degree 2 terms have the highest correlations with those of Levitus SST models. After degree 10 (for both ON function and spherical harmonic), the correlations between the Geosat SST and the Levitus SST become relatively low as compared to those before 10, reflecting the fact that the high degree SST components from the Geosat solutions have larger deviations (than the low degree components) from the Levitus model. The high correlations between the 6-arc solution in Category IV and the Levitus model confirm that the Geosat solution in that category (treating the Levitus as additional observables, but with very pessimistic accuracies) and the Levitus model are indistinguishable (see also the previous discussion). It is interesting to see that the degree 4 and degree 5 terms of the spherical harmonic coefficients have negative correlations with those of Levitus. These negative correlations may be caused by the fact that the correlations by (7.84) for the spherical harmonic coefficients are evaluated on the entire globe and the meaningless SST land values can also affect the calculations. From this view point, the use of the ON functions in calculating the

correlation coefficients should be more meaningful since the domain of concern is only the oceans.

The comparison between the SST maps presented in this section and those in Chapter 6 show that significant deviations between the Geosat SST and the Levitus SST exist in some areas. Such differences are expected, since the Levitus SST are based on data over a period of 70 years and contain various error sources. A good summary of the reasons for the differences between the Geosat SST and the Levitus may be found in Nerem et al. (1990., p. 3173). However, qualitative comparisons reveal that the ocean currents in the five solutions in Table 7.6 are comparable with the currents defined by the Levitus SST. Except in Category IV, the SST from the Geosat solution yield a high centering at about $\phi = -20°$, $\lambda = 250°$. Such a high was also found in the solution of Rapp et al. (1991, p. 38) who used one-year of Geosat data and a spherical harmonic representation.

Table 7.6 Correlations by Degree Between the 6-arc SST Solutions in Category I to V and the Levitus SST† in Percentage

Degree	I*	II	III	IV	V
1					71.2
2	99.5	98.8	99.7	100.0	99.9
3	58.2	63.4	63.1	98.6	90.4
4	63.1	72.0	69.8	99.0	-20.6
5	46.3	62.9	57.5	98.6	-38.5
6	55.9	62.3	68.4	98.1	89.5
7	55.6	64.7	54.3	97.1	71.1
8	70.6	72.5	67.1	98.3	86.5
10	22.7	24.8	32.8	80.5	3.6
15	10.3	10.5	9.3	53.2	10.6
20	22.2	22.3	17.8	80.5	
24	3.2	2.6	3.1	66.5	
Ave. to Deg. 15	45.7	49.8	49.4	87.3	46.8
Ave. to Deg. 24	34.2	36.6	36.1	83.5	

†: The Levitus SST models are from the solution onlsfto24 (for the ON functions, see Figure 6.17) and from shnqto24 (for the spherical harmonics, see Figure 6.23). Except in Category V, all refer to ON coefficients.
*: In Categories I, II, III, IV and V, the SST models correspond to those in Figures 7.4, 7.5, 7.7, 7.9 and 7.10, respectively.

In order to compare the FYS10.W96GW solution (the spherical harmonic (10, 10) solution) in Rapp et al. (1991), we perform the unitary transformation from the SST model from the 6-arc solution in Category III (see Figure 7.7) to a SST model corresponding to system {Yj} which is comparable to the spherical harmonic SST model of Rapp et al. (1991) in terms of solution strategy and the SST function space spanned by the elements. The transformation technique can be found in Section 5.3 and Section 6.4.2. Figure 7.15 shows the transformed SST model up to ON degree 10 with respect to system {Yj} (see (5.76) for the definition). The SST model in Figure 7.16 may be compared with that in

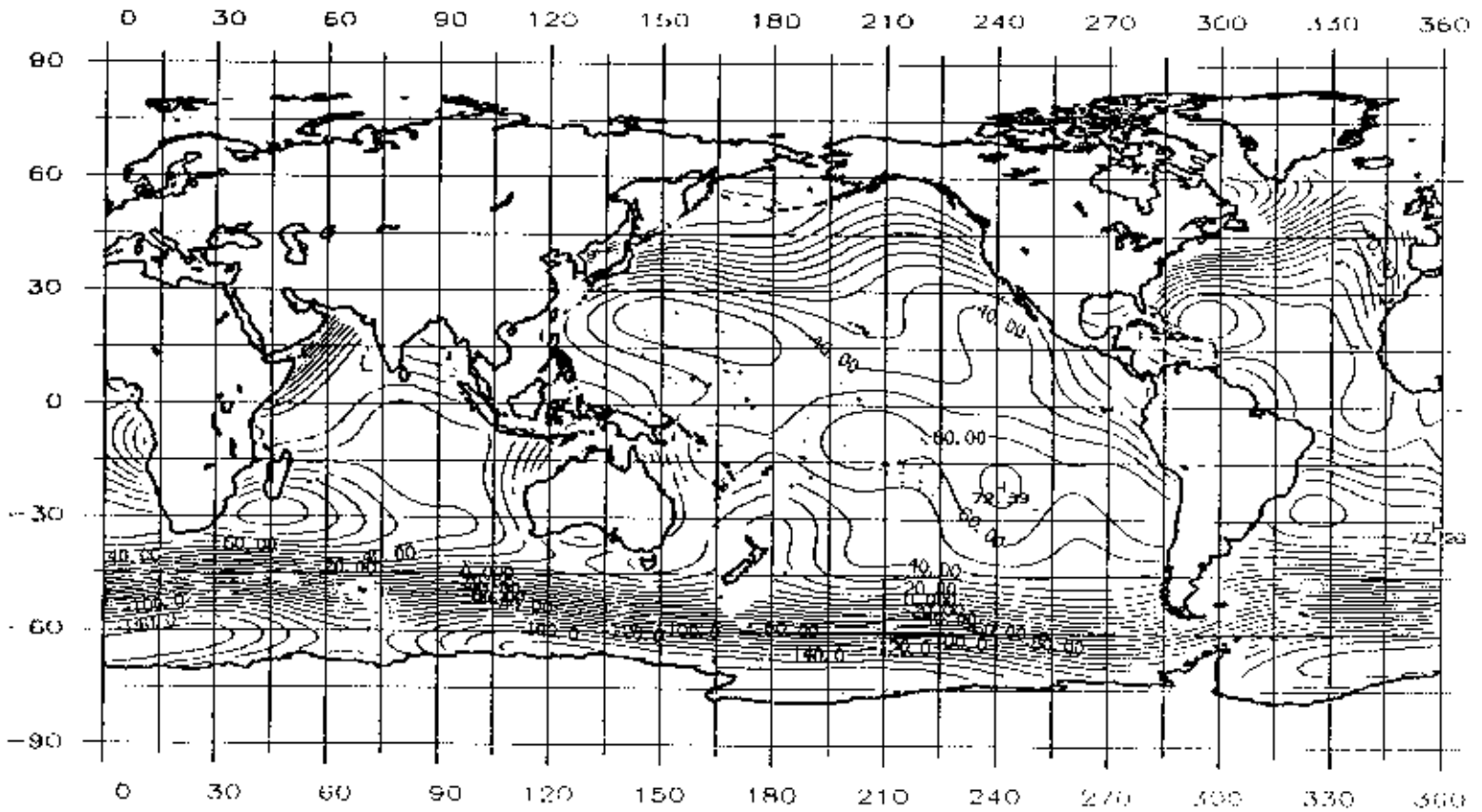


Figure 7.15 SST to ON Degree 10 Using ON System $\{Y_j\}$ from Unitary Transformation of ON (24, 24) 6-arc Solution in Category III, CI = 10 cm

Figure 10 of Rapp et al. (ibid.). Both models reveal the similar SST signature. However, deviations between the SST magnitudes exist, especially in the area of Kuroshio Current, where the difference can amount to 20cm. The SST in FYS10.W96GW show greater details in the current structure (or "streamline" as seen from a SST map). In some other areas, such as the Peru Current, the Gulf Stream, etc., the SST magnitudes in the two models agree very well. The differences in the SST magnitudes and the local SST structures between the two models seems to be reasonable since the model of Rapp et al. (ibid.) used one-year of Geosat data while the 6-arc solution used only 32-days of Geosat data. Study will be needed to find out the reason of large difference in the SST magnitude in the area of Kurishio Current. In addition, from the current flow map in Figure 7.13, we see that the currents missing in the map of Rapp et al., such as the Agulhas Current, still cannot be detected in the 6-arc solution of Category III and it could be attributable to the lack of normal point data in those areas (see Figure 7.3).

7.5.2.3 The Estimability of Geocentric Shift Components ΔX and ΔY

As an experiment, we modeled the geocentric shift components ΔX and ΔY in Category I, III and IV. It will be necessary now to discuss the estimbility of these quantities. The first issue to be discussed is the correlations between the geocentric shift components ΔX, ΔY and other parameters from the simultaneous solutions. It was found that the correlations between ΔX, ΔY and SST coefficients, geopotential coefficients are all below 0.10. However, the ΔY component has relatively high correlations with the orbit error a_0, a_2 terms, except in Category V. Table 7.7 shows the correlations found in various categories. Also included in Table 7.7 are the correlations between the above terms in case that a priori information for the ΔX and ΔY are used in Category III (the case of using Δg). The correlations are at the 60% level and it seems that the ΔX and ΔY components may be recovered with some degree of confidence in the simultaneous solution. In addition, these components from the solutions in Category I and Category III are quite consistent and are about $\Delta X = 0.17 \pm 0.08$m, $\Delta Y = 0.05 \pm 0.08$m. However, the solution in Category IV shows that $\Delta X = -0.14 \pm 0.04$m, $\Delta Y = 0.10 \pm 0.03$m.

Table 7.7 Correlations Between the Geocentric Shift Components and a_0, a_2 Terms of Orbit Errors

term	I	III	III*, $\sigma_{\Delta X} = \sigma_{\Delta Y} = 0.20$m	III*, $\sigma_{\Delta X} = \sigma_{\Delta Y} = 0.10$m	IV
$\rho(\Delta Y, a_0)$	0.63	0.67	0.65	0.59	0.05
$\rho(\Delta Y, a_2)$	0.59	0.63	0.61	0.55	0.03
$\rho(\Delta X, \Delta Y)$	0.19	0.19	0.17	0.13	0.06

* : $1/\sigma^2_{\Delta X}$ and $1/\sigma^2_{\Delta Y}$ are used as a priori weights for ΔX and ΔY

To further investigate the estimability of the ΔX and ΔY components, a test was performed below. We assume that the ΔX and ΔY are known and have the values

$$\Delta X = 0.50 \text{ m} \ , \ \Delta Y = 0.40 \text{ m}$$

These values are then used to generate errors to the altimeter-derived sea surface height in such a manner (see also (7.38)):

$$\Delta h_e = \Delta X \cos\phi\cos\lambda + \Delta Y \cos\phi\sin\lambda \qquad (7.85)$$

where Δh_e is error due to ΔX and ΔY. Using arcs 22, 23, and 24 and incorporating the surface gravity normal equations, we then performed the simultaneous solutions with/without Δh_e applied to the original data (Category III). The results for the ΔX and ΔY components are:

Case I: Δh_e is not applied
$\Delta X = 0.20 \pm 0.08$ meters
$\Delta X = -0.02 \pm 0.08$ meters

Case II : Δh_e is applied
$\Delta X = -0.30 \pm 0.08$ meters
$\Delta X = -0.42 \pm 0.08$ meters

Amazing enough, the differences in ΔX and ΔY between these two cases are precisely the known ΔX and ΔY values! In addition, the resulting SST values are exactly the same. This shows that the incorporation of the ΔX and ΔY components in the simultaneous solution is not totally unrealistic, although the geometry of the altimetric observation is not quite strong for such a purpose. It should be noted that such a determination of the ΔX and ΔY components is only possible through the use of the ON functions or other basis functions than the spherical harmonics for the SST representation and the use of a priori SST information.

7.5.2.4 Accuracies of the SST and Geoid Undulations from the Solutions

Up to this point, we have presented the results (mainly the SST maps and the current flow patterns) from the 6-arc solutions in the five categories. We shall begin to assess the qualities of the various solutions and determine the optimum strategy among the five presented in Table 7.4. In addition, some "old issues" may now be discussed by the use of the ON functions.

We start with the accuracy assessments of the SST and geoid undulations from the solutions. Due to the large amount of tests and limited space, some results will be descriptively presented and some will be illustrated in figures. Using the full error covariance matrix of the parameters (namely potential coefficients, SST coefficients, orbit error parameters, and geocentric shift components) from the 5 aforementioned 6-arc solutions, the standard deviations of the geoid undulations and the SST have been computed at regular 5° x 5° grid points and the results were then interpolated at 1° x 1° grids. Figure 7.16 shows the geoid undulation errors from the 6-arc solution in Category III. From Figure 7.16, we find that the geoid undulation errors are not uniformly distributed on the entire globe. However, over the oceans the geoid undulation error is quite uniform and the RMS errors are about 10 cm. In the coastal areas, the geoid undulation error increase significantly, due to the lack of altimeter data there. In Category I, II and IV where no gravity anomalies have been used, the geoid undulation errors from the 6-arc solution on land are totally unacceptable since the magnitudes can reach 10 meters in some land areas. However, the RMS errors in these three categories over the oceans are still about 10 cm which is about the same magnitude as obtained in Category III (Δg used). Thus, the non-uniformity of the geoid undulation errors in Category I, II and IV is even more serious than that in Category III. For the geoid undulation errors in Category V where the spherical harmonic representation for SST is employed and the gravity data are

used, the error distribution patterns and RMS error are almost the same as those obtained in Category III. From the above discussion, it seems that for the purpose of obtaining geophysically meaningful geopotential coefficients, the inclusion of surface gravity anomalies, especially on land, is necessary or the resulting geopotential coefficients will be "tailored" to an unacceptable extent. Therefore, from the point of view of the geoid undulation errors, the solution strategy in Category III is the optimum one.

Figures 7.17 and 7.18 show respectively the SST errors from the 6-arc solutions in Category III and V (the difference in these two categories is only on the SST basis functions). From the standard deviations of the ON coefficients, it was found that the RMS SST error shown in Figure 7.17 is 11 cm (over the oceans). The RMS SST error shown in Figure 7.18 is about 13 cm (over the oceans). The larger RMS SST error found in Figure 7.18 is attributable to the large standard deviations of the spherical harmonic (1, 0) term (about 10 cm). To reduce such a error, a standard treatment is to fix the (1, 0) term in the course of adjustment (e.g., Denker and Rapp, 1990 and Rapp et al., 1991). Such a treatment should be reasonable in that the (1, 0) term represents the signal component of the "longest" wavelength and it is believed that the average of this term should not change significantly over 1 or 2 years and it should be reliably determined by the oceanographic method. However, smaller oscillations of this term should be expected and it is unfortunate that this term cannot be reliably determined by the altimetric method shown here. The SST model obtained from Category V, as shown in Figure 7.10, does reveal most of SST signatures, although some of them are not as oceanographically meaningful as those obtained from Category III (the ON function case). In addition, Figure 7.19 shows the degree variances of the SST signal/error and the geoid undulation error in case of using spherical harmonics as basis functions for the SST (namely Category V). Due to the non-uniformity of the errors of the SST and geoid undulations over land and the oceans, and due to the fact the degree variances of the spherical harmonic coefficients are evaluated over the entire sphere, any conclusion on the "cut-off" frequency of the SST (the transition point where the signal-to-noise ratio is 1) from a plot such as Figure 7.19 could be misleading. This point has been made very clear in Denker and Rapp (1990).

Due to the use of the ON functions, we shall now re-investigate the problem of the maximum determinable degree for the Geosat SST in terms of the ON functions. Figure 7.20 shows the degree variances and error degree variances of the SST and the geoid undulations using the orthonormal functions from the 6-arc solution in Category I. The ON error degree variances of the geoid undulations have been computed by the technique developed in Section 7.3. To show a more realistic geoid error estimates over the oceanic area using the ON functions, we present the error degree variances using the spherical harmonic functions (the basis functions used for modeling the geoid in the simultaneous solution) and using the ON functions (the functions for estimating the geoid error posterior the solutions using the technique in Section 7.3) in Figure 7.21. Also shown in Figure 7.21 are the error degree variances of the geoid undulations in case of including the gravity data and in case of downweighting the gravity normal. Based on Figure 7.21, substantial improvement on the accuracies of the geoid undulations has been achieved when incorporating the gravity data. Also, the downweighting factor 1/2 has no substantial impact on the accuracies of the geoid undulations. It can be seen from Figure 7.21 that after ON degree 8, the error degree variances (of the ON coefficients) start to decrease,

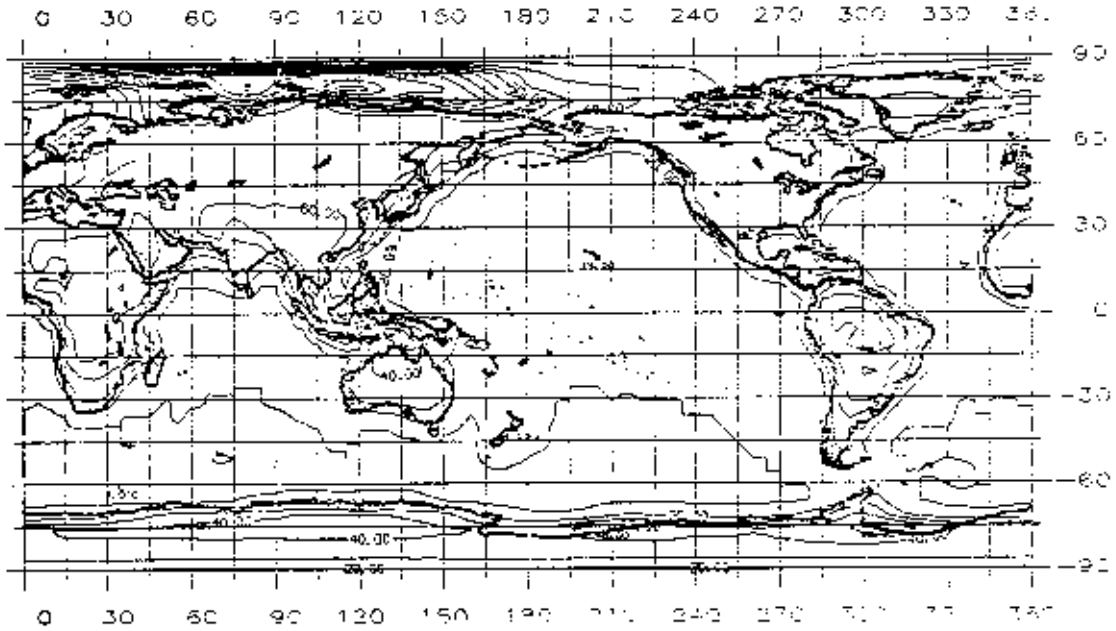


Figure 7.16 Geoid Undulation Errors Corresponding to ON (24, 24) 6-arc Solution, Category III, CI = 5cm

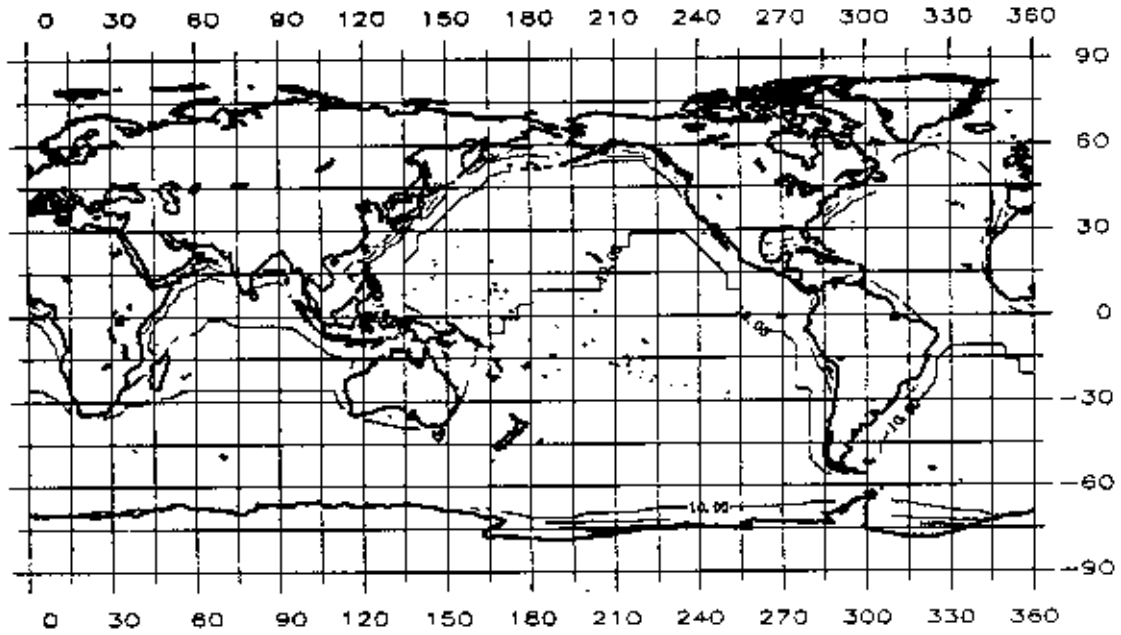


Figure 7.17 SST Errors Corresponding to the SST to ON Degree 15 from (24, 24) 6-arc Solution, Category III, CI = 5cm

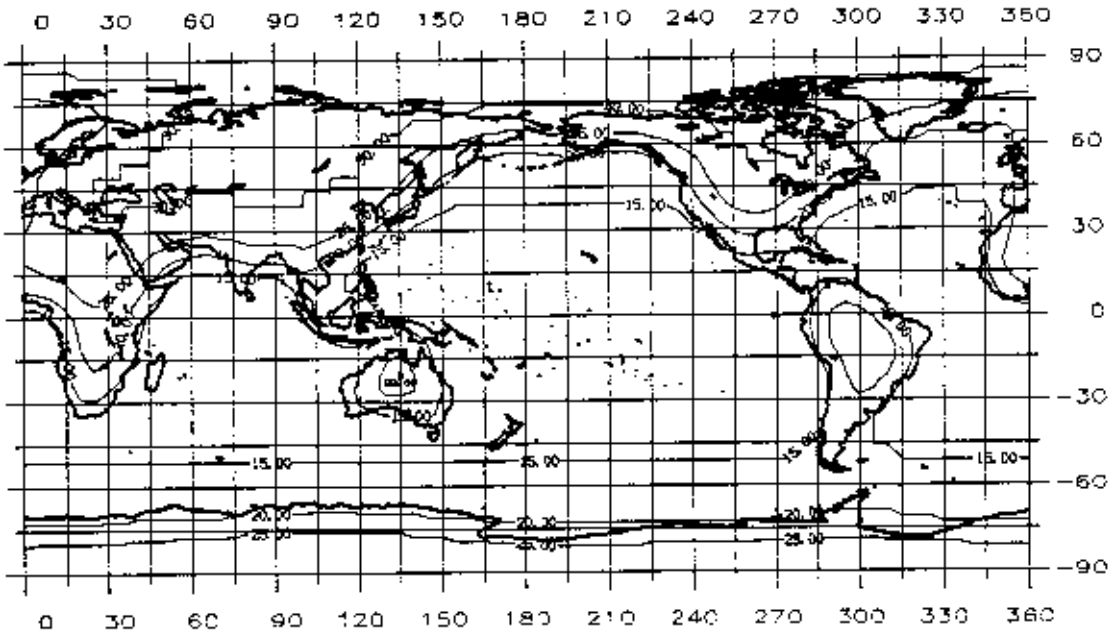


Figure 7.18 SST Errors Corresponding to the Spherical Harmonic (15, 15) Solution, Category V, CI = 5cm

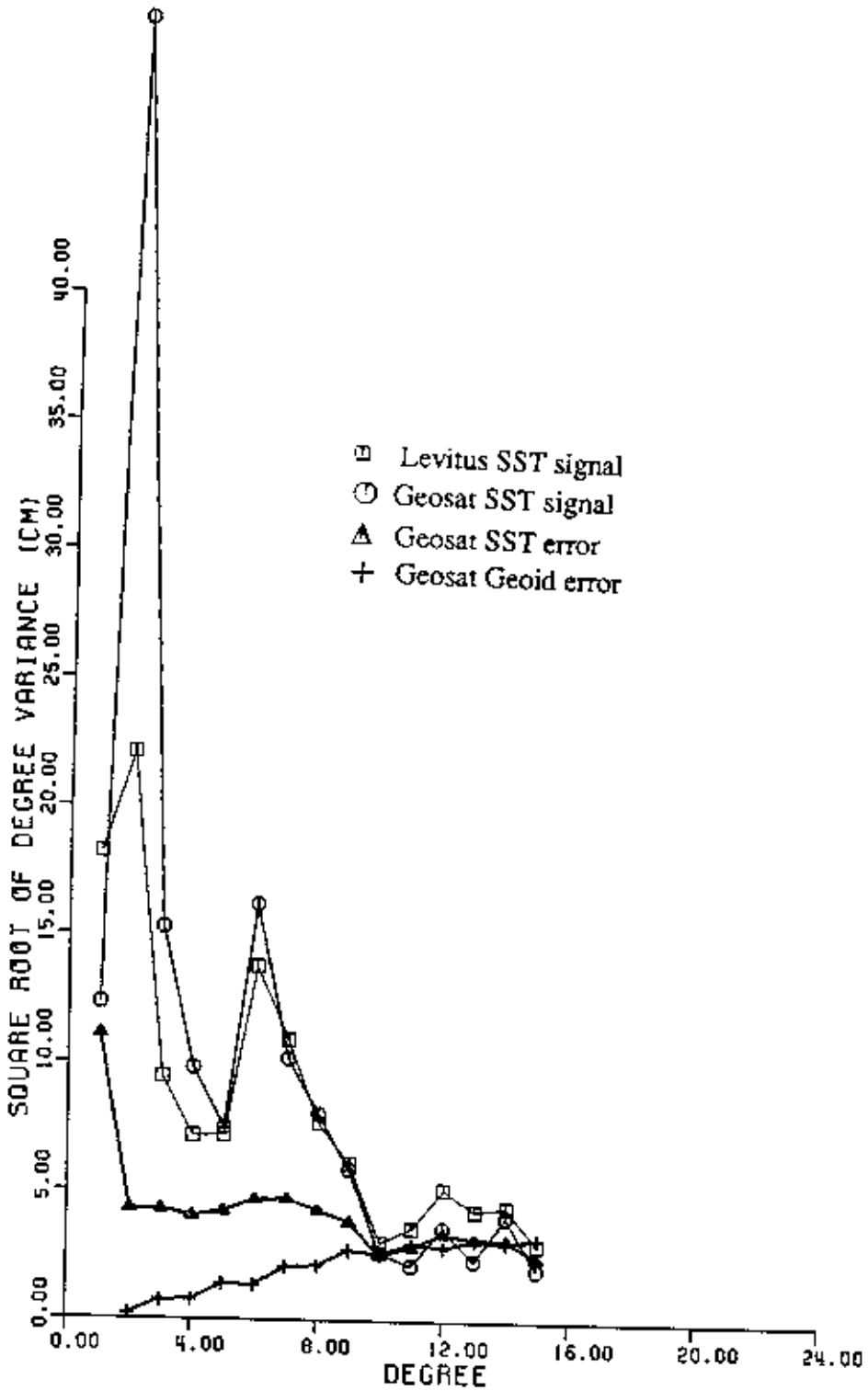


Figure 7.19 Degree Variances and Error Degree Variance From Geosat 6-arc Solution Using Spherical Harmonic Functions, Category V

reflecting the real situation that the altimeter data can provide a better resolution for the short wavelength geoid undulations over the oceans. On the other hand, the error degree variances of spherical harmonic coefficients always increase with degree due to the fact that the spherical harmonic functions are global functions and their error degree variances must reflect the global accuracies of the geoid which could be superior over the oceans and and inferior on land. Therefore, the use of the ON functions should provide a better way of determining the maximum SST degree (of the ON functions) since the domain of concern is only oceans where the SST are defined. The non-uniformity of the geoid undulation errors over oceans and land can be found in Figure 7.16, as mentioned before.

To further show the numerical results for the signal/error analyses using the ON functions, we list the signal/error degree variances of the SST and the geoid undulations from the 6-arc solution in Category III in Table 7.8. In addition, in Figure 7.22, we plot the signal/error degree variances from the 6-arc solution in Category IV, although this solution provides a SST model that could be dependent on the Levitus' model (see the analysis in Section 7.5.2.2). From Figure 7.20, Figure 7.22 and Table 7.8, we see that the SST signal/error and the geoid error are comparable in magnitudes from degree 11 to degree 15. However, after ON degree 15, the SST error and geoid error start to obscure the SST signal considerably. Based on such an analysis, we conclude that the current Geosat system (including the GEM-T2 orbits, the corrections to the raw observables such as tropospherical corrections, tidal corrections, etc.) may not yield a SST model with resolution higher than ON degree 15. Note that such a conclusion is based on the data in the 6 Geosat arcs. The resolution may be higher than ON degree 15 if the one-year or two-year of data are used. It is because of such a cut-off frequency (ON degree 15) that we present the SST discussed so far only up to ON degree 15 although they were modeled up to ON degree 24. In fact, based on the energy distributions of the SST signal components in terms of the ON functions, only about 0.02% of the energy (relative to the energy up to ON degree 24) is contained within ON degrees 16 to 24 for the 6-arc solutions in Category I to III.

Table 7.8 Square Roots of SST Signal/Error and Geoid Error Degree Variances from Geosat 6-arc Solution Using ON Functions, Category III

Degree	SST signal	SST error	Geoid error
2	0.482	0.042	0.007
3	0.192	0.039	0.012
4	0.209	0.035	0.015
5	0.059	0.032	0.020
6	0.076	0.029	0.022
7	0.077	0.029	0.026
8	0.081	0.030	0.027
9	0.053	0.030	0.028
10	0.033	0.024	0.024
11	0.016	0.022	0.023
12	0.019	0.022	0.022
13	0.020	0.022	0.023
14	0.012	0.018	0.019
15	0.007	0.015	0.018
16	0.006	0.015	0.018

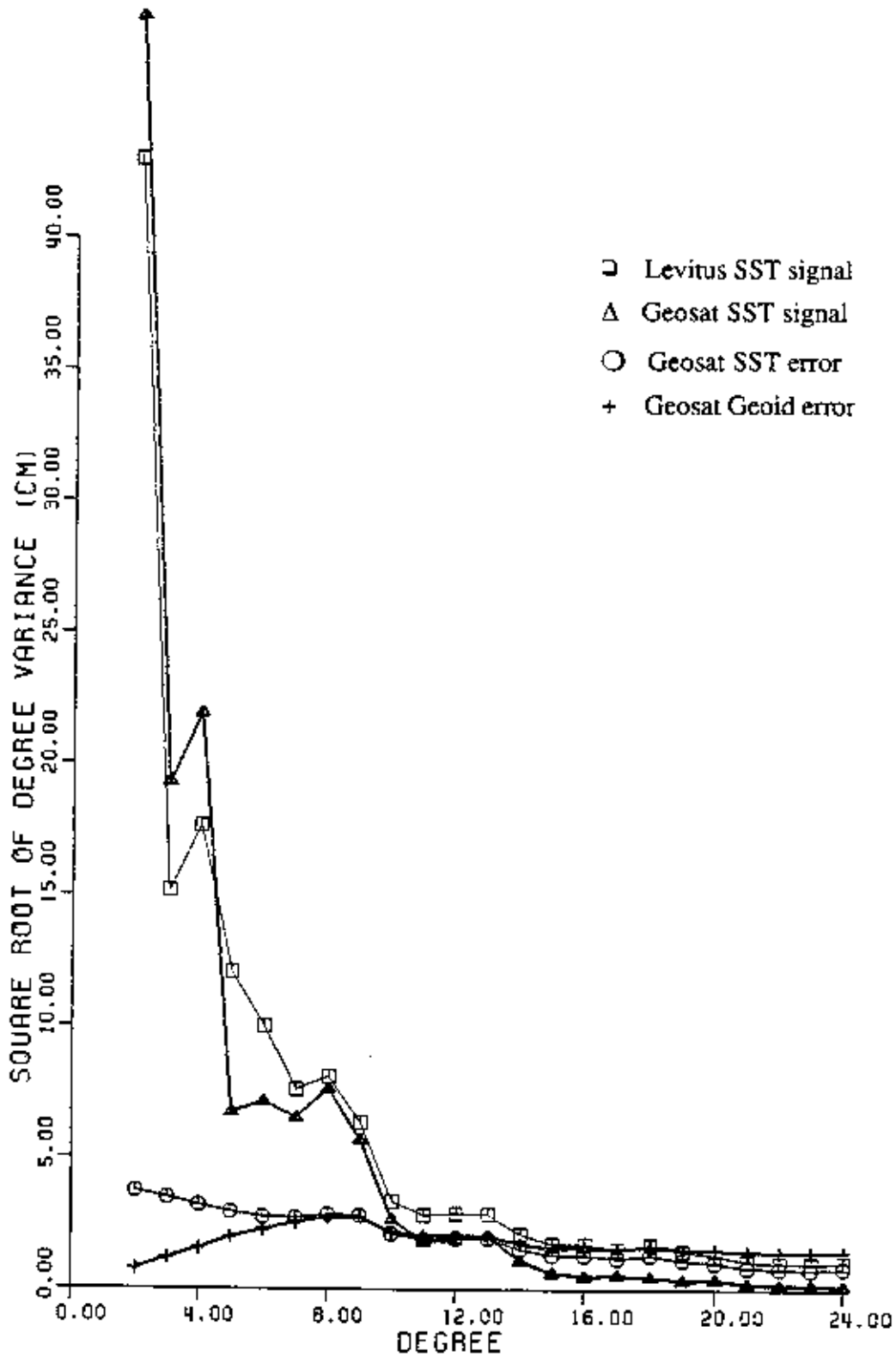


Figure 7.20 Degree Variance and Error Degree Variance from Geosat 6-arc Solution Using Orthonormal Functions, Category I

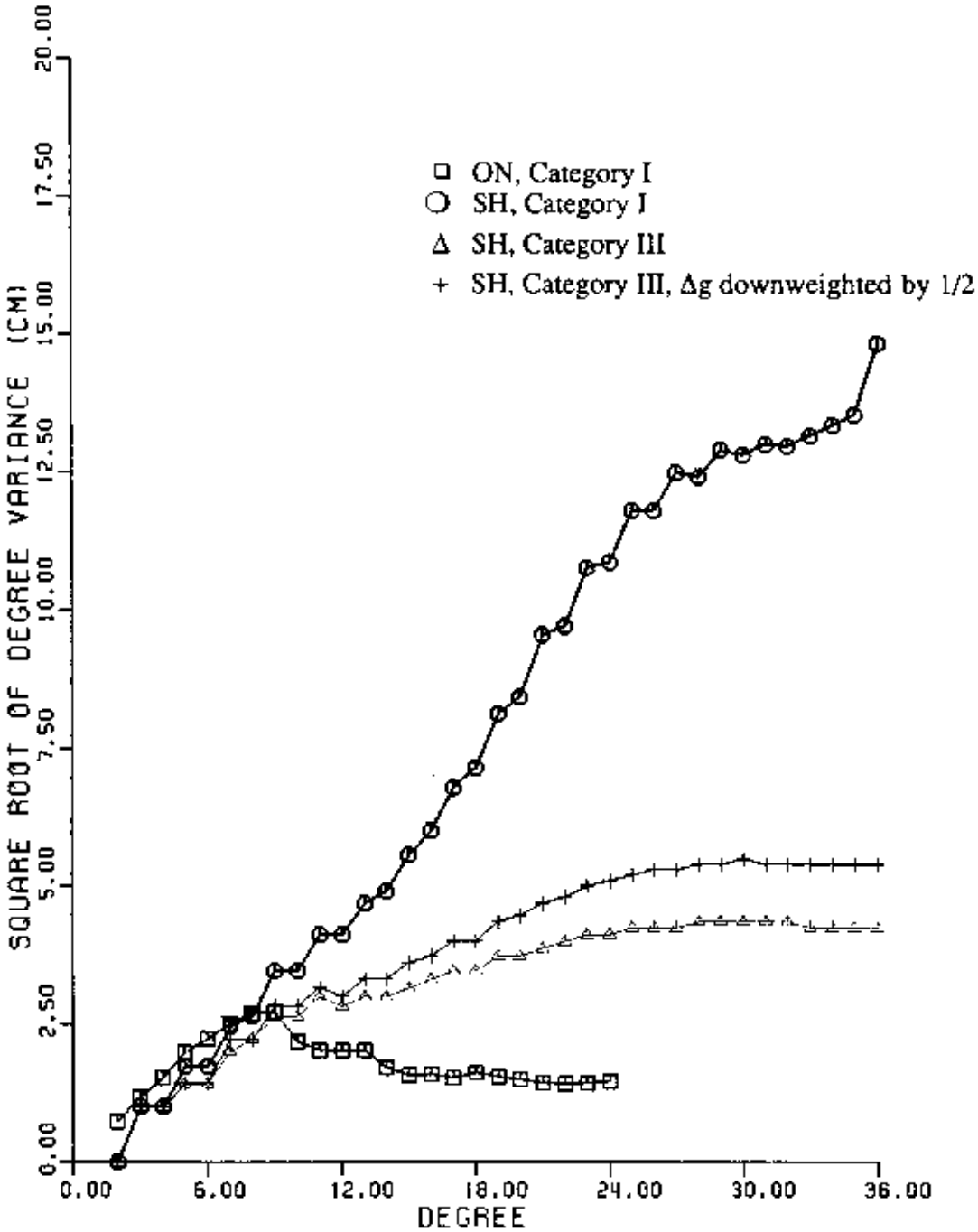


Figure 7.21 Error Degree Variances of Geoid Undulations from Geosat 6-arc Solutions Using Spherical Harmonics (SH) and Orthonormal Functions (ON), Category I and III

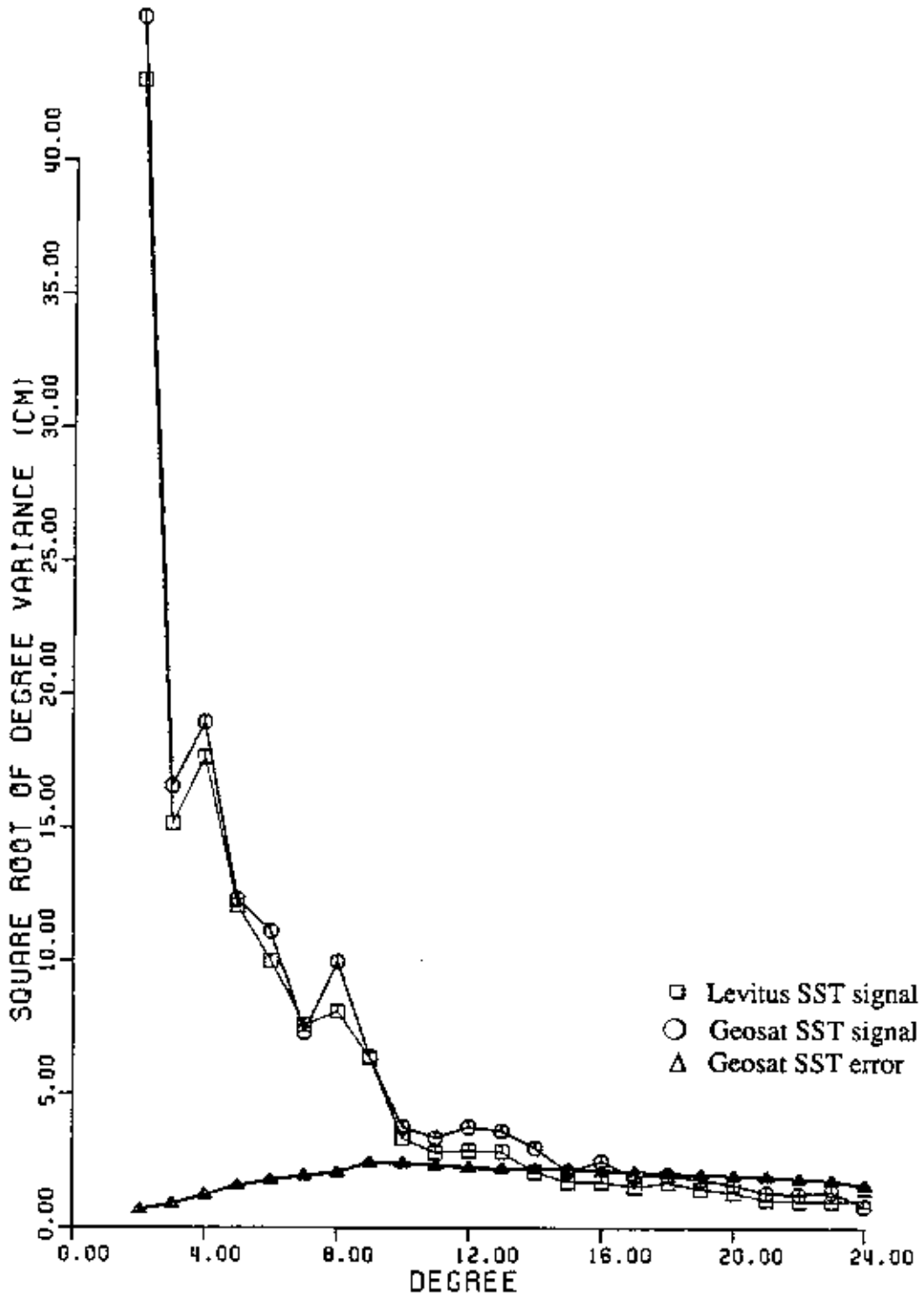


Figure 7.22 Degree Variance and Error Degree Variance from Geosat 6-arc Solution Using Orthonormal Functions, Category IV

7.5.2.5 Orbit Accuracies from the Solutions and Some Statistics

We shall now assess the orbit accuracies from the various solution strategies and present some statistics about all the 6-arc solutions. First of all, we show the RMS values of corrections and some other interested quantities in Table 7.9; Table 7.10 shows the RMS values of orbital errors. These RMS values are computed from the corresponding values at the normal points (see Table 7.2 for the number of normal points in the 6 Geosat arcs). The RMS values for ΔN and SST here will be slightly different from those implied by the ON coefficients since they are evaluated in two different ways and the "domains" are not quite the same. In these two tables $\Delta_T r$ is the total radial orbit error defined in (7.37), $\Delta_I r$ is the error of gravity origin whose definition may be found in (7.37). T is the error due to the ΔX and ΔY components of the geocentric shift. Comparing Table 4.8 in Denker (1990) and Table 7.9, we can see that the GEM-T1 orbits contain larger orbit errors. The undulation corrections based on GEM-T1 orbits are also larger than those based on GEM-T2 orbits. The RMS values of the SST obtained from this study are generally larger than that obtained by Rapp and Denker (1991). Figure 7.23 shows the geoid undulation corrections corresponding to the 6-arc solution in Category III (for the corresponding SST model, see also Figure 7.7). The RMS undulation correction of 1.16 meter obtained here is consistent with the GEM-T2 error estimates and consistent with the results in Rapp et al. (1991, Table 7).

From Table 7.9, we find that for the 6-arc solution, the largest crossover discrepancies occur in Category IV the next in Category III, and then in Category I where the ON functions were used for representing the SST. The adjusted residual sea surface heights are also the largest in Category IV. This shows that the Levitus SST indeed are not quite compatible with the 32-days of Geosat data used here (recalling that in Category IV we treat the Levitus SST as additional observables). The increase of crossover discrepancies from Category I to Category III is 1 cm, about 5% of the magnitude of the average discrepancy. Using the formula $\sigma_X/\sqrt{2} \le \sigma_T \le \sigma_X$ where σ_X is the crossover discrepancy and σ_T is the RMS error due to geopotential, we find that the total orbital error after the 6-arc solution ranges from 13 to 21 cm.

The most important comparison will be on the results from the 6-arc solutions in Category III and Category V (ON functions versus spherical harmonics). From Table 9(c) and 9(d), it is found that the use of spherical harmonics increases the RMS residual sea surface heights (after the adjustments) by 1 cm or about 5% as compared to the model of using the ON functions. Thus the latter model fits the observations better than the former. The crossover discrepancies after the adjustments in these two models are the same at arcs 22, 23, 26 and 27, but at arcs 24 and 25 the use of spherical harmonics yields 1 cm increase (about 5%). The recovered ΔN values in these two models are exactly the same except at arc 23. For the individual orbital error terms, the maximum deviation between the two models occurs at the 1 cyc/rev terms. According to the comparisons on the RMS residual sea surface heights and the crossover discrepancies after the adjustments in these two models, it is evident that these 1 cyc/rev terms should be more properly recovered in the case of using the ON functions than the case of using the spherical harmonics. In addition, we found that the sine term of 1 cyc/rev error has a standard deviation of 17 cm in the case of using the spherical harmonics, proving that these terms cannot be accurately recovered in such a model. Based on these discussions, we conclude that the case with the ON functions yields a better orbit accuracy than the case with the spherical harmonics.

Table 7.9 RMS Values of Corrections to the Residual Sea Surface Heights and Adjusted Crossover Discrepancies in Meters

(a) 6-arc solution, Category I

arc	$\Delta_T r + T$	ΔN	SST	Δh model	Δh resid.	Crossover discr.		No. of Crossovers
						before	after	
22	0.78	1.16	0.69	1.63	0.18	0.36	0.19	1646
23	0.79	1.15	0.70	1.61	0.17	0.41	0.20	1858
24	0.84	1.15	0.71	1.67	0.17	0.50	0.20	2083
25	0.75	1.15	0.70	1.58	0.17	0.42	0.18	1749
26	0.77	1.16	0.69	1.59	0.17	0.37	0.18	1724
27	0.83	1.15	0.69	1.63	0.18	0.47	0.21	1898

(b) 6-arc solution, Category II

arc	$\Delta_T r$	ΔN	SST	Δh model	Δh resid.	Crossover discr.(after)
22	0.77	1.16	0.70	1.63	0.18	0.19
23	0.78	1.15	0.70	1.61	0.17	0.20
24	0.84	1.15	0.72	1.67	0.17	0.20
25	0.75	1.15	0.71	1.58	0.17	0.18
26	0.76	1.16	0.70	1.59	0.17	0.18
27	0.83	1.15	0.70	1.63	0.18	0.21

(c) 6-arc solution, Category III

arc	$\Delta_T r + T$	ΔN	SST	Δh model	Δh resid.	Crossover discr. (after)
22	0.76	1.17	0.69	1.62	0.18	0.20
23	0.76	1.16	0.69	1.61	0.18	0.21
24	0.81	1.16	0.70	1.67	0.17	0.20
25	0.73	1.15	0.69	1.58	0.17	0.19
26	0.74	1.16	0.69	1.60	0.17	0.19
27	0.81	1.16	0.69	1.63	0.18	0.22

(d) 6-arc solution, σ_{SST} = 0.25m, Category IV

arc	$\Delta_T r + T$	ΔN	SST	Δh model	Δh resid.	Crossover discr. (after)
22	0.62	1.19	0.69	1.61	0.19	0.20
23	0.63	1.18	0.69	1.59	0.19	0.21
24	0.69	1.19	0.71	1.66	0.19	0.20
25	0.61	1.18	0.70	1.57	0.18	0.20
26	0.61	1.19	0.69	1.58	0.18	0.19
27	0.68	1.18	0.69	1.62	0.19	0.22

Table 7.9 (continued)

(e) 6-arc solution, N_{max}^{SST} = 15, Category V

arc	$\Delta_T r$	ΔN	SST	Δh model	Δh resid.	Crossover discr. (after)
22	0.74	1.17	0.73	1.62	0.19	0.20
23	0.75	1.16	0.73	1.61	0.18	0.21
24	0.80	1.16	0.74	1.67	0.18	0.21
25	0.73	1.15	0.73	1.58	0.17	0.20
26	0.73	1.16	0.73	1.59	0.17	0.19
27	0.80	1.16	0.72	1.63	0.19	0.22

Table 7.10 RMS Values of Orbital Errors (Solution Categories are the Same as Those in Table 7.9) and Geocentric Shift in Meters

(a)

arc	$\Delta_1 r$	const.	1 cyc/rev+ ΔZ	1cyc/rev(t)	1cyc/rev (t^2)	2 cyc/rev(t)	ΔX	ΔY
22	0.22	0.57	0.31	0.06	0.02	0.01		
23	0.23	0.61	0.24	0.09	0.06	0.00	0.17	0.05
24	0.23	0.59	0.29	0.24	0.01	0.00	±0.08	±0.09
25	0.23	0.60	0.20	0.13	0.06	0.00		
26	0.23	0.63	0.19	0.03	0.03	0.01		
27	0.24	0.63	0.24	0.15	0.04	0.00		

(b)

arc	$\Delta_1 r$	const.	1 cyc/rev + ΔZ	1 cyc/rev (t)	1 cyc/rev (t^2)	2 cyc/rev (t)
22	0.21	0.53	0.33	0.06	0.02	0.01
23	0.23	0.57	0.27	0.09	0.06	0.00
24	0.22	0.55	0.32	0.24	0.01	0.00
25	0.22	0.56	0.23	0.13	0.06	0.00
26	0.22	0.59	0.22	0.03	0.03	0.01
27	0.23	0.59	0.26	0.15	0.04	0.00

(c)

arc	$\Delta_1 r$	const.	1 cyc/rev + ΔZ	1 cyc/rev(t)	1 cyc/rev(t^2)	2 cyc/rev(t)	ΔX	ΔY
22	0.23	0.54	0.29	0.06	0.02	0.01	0.17	0.04
23	0.24	0.58	0.23	0.09	0.06	0.00	±0.07	±0.08
24	0.23	0.56	0.29	0.24	0.01	0.00		
25	0.23	0.58	0.19	0.12	0.06	0.00		
26	0.23	0.60	0.18	0.03	0.03	0.01		
27	0.24	0.60	0.23	0.15	0.04	0.00		

Table 7.10 (continued)

(d)

arc	$\Delta_1 r$	const.	1 cyc/rev + ΔZ	1 cyc/rev (t)	1 cyc/rev (t^2)	2 cyc/rev (t)	ΔX	ΔY
22	0.24	0.51	0.21	0.06	0.02	0.01	-0.14	0.10
23	0.24	0.55	0.14	0.09	0.06	0.01	±0.04	±0.03
24	0.25	0.53	0.20	0.24	0.01	0.01		
25	0.23	0.54	0.11	0.12	0.06	0.00		
26	0.25	0.57	0.11	0.03	0.03	0.01		
27	0.24	0.57	0.15	0.15	0.04	0.00		

(e)

arc	$\Delta_1 r$	const.	1 cyc/rev + ΔZ	1 cyc/rev (t)	1 cyc/rev (t^2)	2 cyc/rev (t)
22	0.23	0.55	0.24	0.06	0.02	0.01
23	0.24	0.59	0.18	0.09	0.06	0.01
24	0.23	0.57	0.23	0.24	0.01	0.00
25	0.23	0.58	0.14	0.12	0.06	0.00
26	0.23	0.61	0.14	0.03	0.03	0.01
27	0.24	0.61	0.18	0.15	0.04	0.00

In all the categories, three types of solution, namely, the 1-arc, 3-arc and 6-arc solutions, have been tested. The crossover differences (for the numbers of crossovers in individual arcs, see Table 7.9) after adjustment increase from about 15 cm in the 1-arc solution to about 20 cm in the 6-arc solution. This is reasonable since in a multi-arc solution the least-squares solution must simultaneously take care of the residuals pertaining to individual arcs, resulting in a poorer fit of the residual sea surface heights to the unknown parameters in some arcs. However, such a slightly inferior accuracy of orbits will not create problems in future use of the arcs (such as the prediction of gravity anomalies), since remaining errors can be further treated by a simple model in a local area.

The orbit errors of GEM-T2 orbits are of different nature as compared to those of GEM-T1 found in Denker and Rapp (1990). For the purpose of comparison, we show the RMS values of the orbit errors from one of Denker's 6-day arcs and from arc 27 (see Table 7.10(c)) in Table 7.11. Although such a comparison is not quite valid, it provides a general feeling of how the errors of these two kinds of orbits behave. Indeed, one can find that before the adjustment process, the GEM-T2 orbits are superior to GEM-T1 orbits. In particular, the errors due to resonant effects in GEM-T2 orbits are much less those in GEM-T1 orbits. The time-dependent 2 cyc/rev terms are almost zero in both GEM-T1 and GEM-T2 orbits. The constant terms of the orbit errors from the two orbits are very consistent and are about 60 cm. This term is normally interpreted as the correction to equatorial radius of the ellipsoid implied by the satellite field (now GEM-T2 or GEM-T1). However, from the radial error analysis in Section 7.1, we know that this term should have included the mixed effect of the equatorial radius correction, the initial state vector error and the error of gravity origin, but it is believed that the first one has the dominant effect.

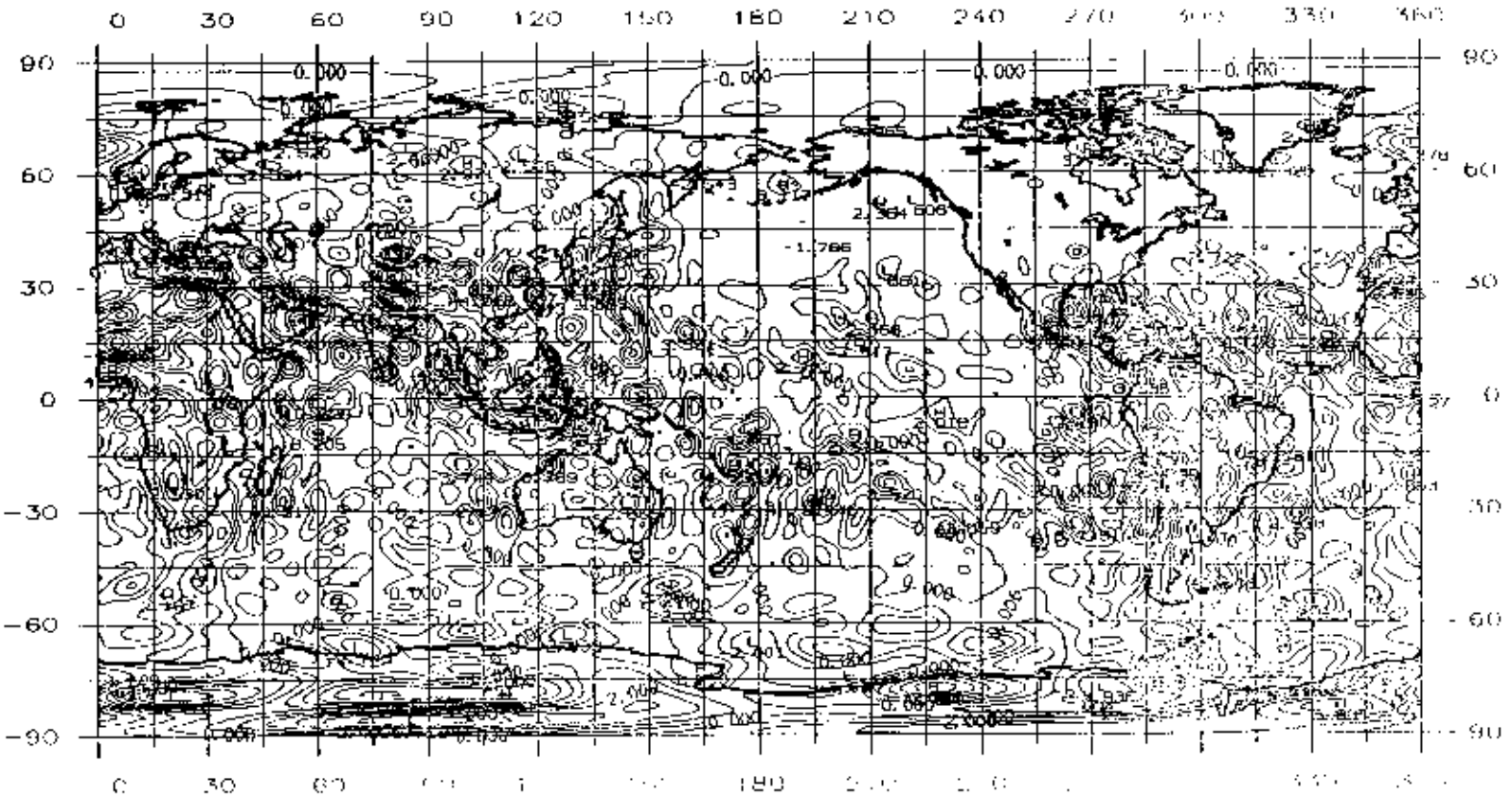


Figure 7.23 Geoid Undulation Corrections from ON (24, 24) 6-arc Solutions, Category III, CI = 1m

Table 7.11 Comparison of Orbit Errors of GEM-T1* and GEM-T2† Orbits

orbit	$\Delta_1 r$	const.	1 cyc/rev	1 cyc/rev (t)	1 cyc/rev (t^2)	2 cyc/rev (t)
GEMT1	0.40	0.60	0.47	0.65	0.16	0.02
GEMT2	0.24	0.60	0.23	0.15	0.04	0.00

Unit = meters
* GEM-T1 orbit from one 17-day solution at epoch 870416
† GEM-T2 orbit from the result in Table 7.10(c), arc 27

7.5.2.6 Correlation Analysis

Correlation analyses using the covariance matrix from the 6-arc solutions show that there is no significant (below 50%) correlation between the SST ON coefficients and geopotential coefficients. However, significant correlations (about 60%) were found between the constant terms of orbit errors and the ON degree 2 terms of the SST (in Category I). Use of gravity data has reduced such a correlation to an acceptable level (below 50%). High correlations between the constant terms (i.e. a_0 terms) and the spherical harmonics degree 1 and 2 terms (Category V) were also found and such a phenomenon is consistent with Engelis and Knudsen's (1989, p. 200) finding.

To calculate the correlation coefficients between the geoid undulation correction and the SST, we write formally

$$ {}_mN_1 = {}_mA_{uu}X_1 \tag{7.86} $$

$$ {}_m\zeta_1 = {}_mB_{vv}X_1 \tag{7.87} $$

where N is the vector of geoid undulation corrections, ζ is the vector of SST values, X is the voctor of geopotential coefficients, Y is the vector of SST coefficients, A is a matrix containing the spherical harmonics, B is a matrix containing the basis functions of SST, m is the number of points where the N and ζ are evaluated, u and v are associated with the maximum expansion degrees of the geopotential coefficients and the SST coefficients. From the adjustment process, we can obtain the error covariance matrix between X and Y as

$$ \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \tag{7.88} $$

Then we can get the error covariance between N and ζ at the m points using error propogation:

$$ \begin{bmatrix} \Sigma_{NN} & \Sigma_{N\zeta} \\ \Sigma_{\zeta N} & \Sigma_{\zeta\zeta} \end{bmatrix} = \begin{bmatrix} A^T\Sigma_{XX}A^T & A\Sigma_{XY}B^T \\ B\Sigma_{YX}A^T & B\Sigma_{YY}B^T \end{bmatrix} \tag{7.89} $$

then the correlation coefficient between N and ζ at one point (ϕ,λ) is

$$\rho(\phi,\lambda) = \frac{\sigma_{N\zeta}}{\sigma_N \sigma_\zeta} \tag{7.90}$$

where $\sigma_{N\zeta}$ is from $\Sigma_{\zeta N}$ and the standard deviations of N and ζ, namely σ_N and σ_ζ, are from the matrices in (7.89). Such a procedure for determining $\rho(\phi, \lambda)$ has been used by Nerem et al. (1990). To evaluate the $\rho(\phi, \lambda)$ at 5° x 5° grids on the entire globe, m will be 36 x 72 = 2592. Using such a procedure and 5° x 5° grids, five sets of correlation coefficients between N and ζ for the 6-arc solutions in Categories I to V are calculated. Based on these sets, the lowest correlations between the geoid undulation corrections and the SST are found in Category IV, the next is found in Category III, then in Category I (all ON function cases). The mean values in these sets are all negative showing the geoid undulation corrections and the SST are negatively correlated. Of particular interest is the comparison between the correlation coefficients in Categories III and V (all 6-arc solution). It was found that the correlation coefficients in Category III have a mean value of -0.373, and a RMS value of 0.386, while in Category V, the corresponding values are -0.034 and 0.071. Both sets of values show that good separation between N and ζ are achieved in the cases of using spherical harmonics and ON functions. Despite the advantage of relatively low correlations in Category IV, that the oceanographic SST will dominate the resulting SST from the simultaneous solutions does not look like a preferable situation since we want to obtain an independent SST estimate from satellite altimetry. Therefore, the best solution strategy seems to be that in Category III in this regard since in this category we can obtain the geophysically meaningful geopotential coefficients, reasonable SST signatures and low correlations between the undulation corrections and the SST.

7.5.2.7 Conclusions

The primary goal of the experiments has been to investigate the capability of the ON functions in representation of the SST and spectral analyses of the signals from the simultaneous solutions. The above analysis in various aspects have shown that the ON functions have the required capability. However, some problems are basically inherent in the simultaneous solutions and any kind of basis functions for the SST cannot resolve these problems. One problem is related to the geometry of the functions of the unknowns in the observables - the residual sea surface heights. All these functions lie in the radial direction and creates difficulty in separation of the individual components. Although additional data, such as gravity data, can improve the situation, the results still rely on the accuracies of the incorporated data. The other problem is in the linear orbit theory. Although such a theory provides a good insight into the orbit error characteristics, it has certain limit. Wagner (1986) and Wagner (1991, private communication) pointed out that the linear theory has a defect in recovering the orbit error arising from the odd-degree zonal terms of potential coefficients. Unless an improved analytical orbit theory is made, the weakness in the simultaneous solution still exists. Nevertheless, it is believed that a priori information and a good starting altimeter system should have reduced such a defect.

Despite all the possible problems in the simultaneous solution, the ON functions still provide a good tool for assessing the accuracies of the estimated quantities. The ON functions offer an advantage over the spherical harmonic functions that our analyses on the estimability and the resolution of the SST will not be affected by the situation on land and hence the analyses can lead to correct conclusions. Also, the accuracy of the estimated geoid, which could be superior over the oceans and inferior on land, now can be compared with the signal/error of the SST in the same domain using the ON functions and the transformation technique given in Section 7.3.

Several important conclusions can be drawn from the results of the experiments described in this section:

1. The solution using altimeter data, surface gravity data (Category III) is the optimum among all the solutions, considering all factors such as SST signatures, geoid undulation accuracies, etc.

2. The SST resolution from the 6-arc solutions should not exceed ON degree 15 based on the signal/error analysis. To improve the resolution, the Geosat needs a refined system taking into account better orbits and better correction models to the raw data.

3. The geocentric shift components ΔX, ΔY can be estimated with 60% correlation existing between ΔY and the orbit error a_0 and a_2 terms. The correlation may be reduced by some a priori information of these two components.

4. The use of ON functions for modeling the SST in general yield better defined ocean currents and better resolutions in the coastal area as compared to the spherical harmonic functions.

5. If correct oceanic equations of motion are incorporated in the simultaneous solution, the separation between the SST and geoid is the best in terms of correlation analysis.

8. Conclusions and Recommendations

In this study, a set of orthonormal functions have been constructed using one of the methods proposed in Chapter 3. Tests have shown that the spectral behavior of a function in such a orthonormal function expansion is remarkably good as compared to that in a spherical harmonic expansion over the oceans. Experiments in the simultaneous model with the orthonormal functions as basis functions for the SST also demonstrated the capability of these functions in modeling the SST.

The ON functions provide an opportunity for signal decomposition of oceanic signals such as the SST and the oceanic geoid. They also provide an excellent tool for assessing the accuracies of signal components over the oceans. These analyses can also be carried out using the spherical harmonic expansions, but theoretical justifications cannot be found and often the analyses lead to incorrect conclusions.

Many other othonormal functions have been developed in Chapter 3 and Chapter 4. Of particular interest are the spherical cap harmonics which are orthonormal over a spherical cap and satisfy Laplace's equation. The idea of finding the spherical cap harmonics was formed one century ago (see Thomson et al., 1879); the latest development on this subject is given by Haines (1985a). However, all these investigators have their own definitions of spherical cap harmonics which are not consistent with the definition of the fully normalized spherical harmonics, and detailed computational formula such as those presented in Chapter 3, cannot be readily found. The work done in Chapter 3 has provided all necessary definitions and software for future applications of these functions. The applications could be expanding the gravity anomalies in the Arctic and Antarctic areas, expanding the SST in the Pacific ocean for sea surface variation studies and expanding any signal within a spherical cap.

Also provided in Chapter 3 are the orthonormalized Fourier-Bessel functions, or in the author's definition, the generalized Fourier-Bessel functions, which are suitable for series expansion of a signal defined on a domain bounded by two parallels and two meridians. The applications of these functions are thus similar to those of spherical cap harmonics. One may employ these functions to approximate the surface topography in the United States or Europe. As the final method, the conformal mappings also contribute to this study in providing complex orthonormal functions. In comparison to the eigenvalue-eigenfunction method, the method of conformal mapping has greater flexibility in "choosing" an oceanic domain, such as a polygon. Unfortunately the mapping functions often do not have simple forms or even closed forms, yielding only "approximate" complex orthogonal functions in some cases. In addition to being a tool for finding orthogonal functions, conformal mapping can also be applied to the solution of boundary value problem (Churchill and Brown, 1984). The idea is to transform a supposedly complicated boundary and its boundary condition to a simple boundary and its associated boundary condition, then the problem is solved using the simple boundary and the associated boundary condition. Finally, the desired solution is found through the relationship implied by the mapping. Thus conformal mapping could be a tool for solving the geodetic boundary value problem.

The generalized spherical 2-D Fourier functions and the generalized Fourier-Tschebyscheff functions are proposed in Chapter 4. These orthonormal functions can be basis functions for representing a global signal, such as the surface topography, the geoid etc. Naturally they can be alternative functions in the simultaneous solution but they will probably suffer the same problem in the spectral domain as the spherical harmonics do.

The geodesist's old friends—spherical harmonics, apparently do not lose their importance in this study. They are the necessary functions for constructing the orthonormal functions that are used in the SST expansion and in the simultaneous solution. An unfortunate situation is that they are independent only up to a certain degree and order over the oceans. Due to such a function dependence, 3 orthonormal systems have been constructed using different combinations of the spherical harmonics. A compact FFT form for computing the inner products of the spherical harmonics over the oceans was developed. A relationship between one associated Legendre function and the product of two associated Legendre functions was found. Such a relationship thus provides a new technique for computing the integration of the product of two associated Legendre functions. However, in this new technique a numerical instability problem exists when using CRAY single precision or IBM double precision to compute the combination constants. Although such a problem can be solved by using Cray double precision, it will be an interesting topic to find a stable algorithm for computing constants in a CRAY single precision environment or an IBM double precision environment.

One important theorem which has been proved and numerically tested is related to a technique for detecting the maximum spherical harmonic degree of a global bandlimited signal given data only over the oceans. Another theorem states that the orthonormal functions constructed by the Gram-Schmidt process do not have eigenvalues with respect to the Laplace surface operator Δ^*. The second theorem has been a basis for arguing that there is also no eigenvalue with respect to any kernel function for the ON functions (See Section 6.2.1).

The results from the expansions of the Levitus SST show that to approximate the high frequency SST, the maximum expansion degree must be at least 15 using either the

ON functions or the spherical harmonics. Using only oceanic data, the spherical harmonic expansions with Nmax > 10 yield excessively large coefficients, thus the "optimum" Nmax was normally set to 10 in the past (Engelis, 1987b). However, as we just stated, such a low Nmax will lead to a poor fit to the data. What the ON functions have contributed in resolving this dilemma is to obtain explainable expansion coefficients without losing the approximation accuracies.

As a "by-product" of this research, a FFT method was developed to compute the elements of the normal matrix and the "U" vector (see (6.56)) in calculating the geopotential coefficients from gravity anomalies. It has been found (see Table 6.2) that for Nmax = 70 only 146 CPU seconds is needed to form the normal matrix and U vector on the CRAY Y-MP/864 machine, about 130 times faster than the conventional method (Rapp, 1991, private communication). Such a highly improved method will allow researchers to test various options of data weightings and selection strategies without worrying about the computer time. This method can be extended to take into account miscellaneous data such as gravity anomalies, geoid undulations. For such a purpose, the modification to the software developed in Chapter 6 should not be a problem.

The spectral analysis of the Levitus SST using the ON functions shows that 98.52% of the energy is contained within ON degree 10; up to ON degree 24, 99.90% of the energy is recovered. Thus we conclude that the Levitus SST are low frequency signal (with respect to the ON functions). This result is not surprising since a smoothing function has been employed by Levitus when he constructed his SST. It will be impossible to detect such a phenomenon using the spherical harmonic expansions either due to the theoretical problem or the practical problem. The theoretical problem is that the signal component obtained form the spherical harmonic expansion is not truly a "component" for an oceanic signal in the Fourier analysis sense. The practical problem is that the expansion coefficients do not have the permanent property since the coefficients change significantly as the Nmax values change. As a result, different conclusions on the energy distribution of the Levitus SST can be made as different Nmax values are used. Although the above theoretical and practical problems can be resolved by artificially setting SST = 0 on land, the results on the energy distribution from the spherical harmonic expansions are misleading due to a relatively slow convergence of the expansion coefficients and the artificial energy introduced at the continental boundaries.

In Chapter 7, the linearized Lagrange's equations of motion have been briefly reviewed with an aim to explain the origins of satellite radial orbit errors. In a detailed analysis of functional forms of radial orbit errors due to initial state vector, the geocentric shift components and the spherical harmonics, we analytically verified a correlation problem pointed out by other researchers in the past. The use of ON functions only removed the part of correlation between the sine term of 1 cyc/rev radial orbit error and the SST (1, 0) term, the problem of geocentric shift is still present. Due to a further complication of the functional relationship, it is proved that only the mixed error due to the ΔZ - component and the sine term of 1 cyc/rev orbit error is solvable. Therefore, in terms of geocentric shift, only the ΔX, ΔY components are independently modeled in the simultaneous model. The solutions show that the two components are on the order of 10 cm for the GEMT2 orbits. The correlation between them is low (about 16%) since the altimeter data cover a large portion of the earth and allow a relatively good estimability of the transformation parameters—ΔX, ΔY, between the geocentric system and the satellite system.

Simultaneous models other than the one used in Denker (1990) have also been tested in Chapter 7. It was found that the model which treats the Levitus SST as additional observables yields a SST model almost identical to the Levitus', even if a very pessimistic accuracy and a small downweighting factor are used. This particular model was tried in response to the ideas of using oceanic dynamic equations put forward by, for example, Wagner (1989), Marshall (1985), Engelis (1987a). However, it appears that a different way of trying this idea will be more appropriate. Another model incorporated gravity anomalies into the solution, yielding better geoid undulation accuracies and increasing the separability between the geoid and the SST. In all these models, we have used the ON functions to perform spectral analyses for the SST signal/error and for the geoid error. Again, it is more theoretically justifiable to use the ON functions for such analyses and it is believed that correct conclusions have been obtained by using the ON functions for spectral analyses. In terms of the SST resolution from the Geosat data, we found that the cutoff frequency is in the vicinity of ON degree 15, depending on the models used. As stated in Wunsch (1991), "Geosat was a crude system and there was no expectation that it could be used to study the large-scale ocean circulation," the analyses using the ON functions here may somewhat prove his statement (note that this is only based on 6-arc solutions). However, this current Geosat system may be improved in the future by using more accurate orbits and more reliable correction models to the raw altimeter data.

It is hoped that an improved Geosat system will be available soon or the TOPEX/POSEIDON mission (Fu et al., 1991) will fill the gap between system capabilities and the required accuracies. Then, the orthonormal functions used in all the numerical experiments in this study, as well as other proposed yet not tested orthonormal functions, will contribute to the geodetic and oceanographic communities in correctly performing spectral analyses over the oceans and extracting useful informations.

Appendix A

Proof of Orthonormality of Functions Obtained From the Gram-Schmidt Process

From the process shown in (3.1), we get

$$(\bar{f}_2, \bar{f}_1) = \frac{1}{\|h_2\|}(h_2, \bar{f}_1) = \frac{1}{\|h_2\|}(f_2, \bar{f}_1) - (f_2, \bar{f}_1)(\bar{f}_1, \bar{f}_1) = 0$$

$$\begin{aligned}(\bar{f}_3, \bar{f}_1) &= \frac{1}{\|h_3\|}(h_3, \bar{f}_1) = \frac{1}{\|h_3\|}\left(f_3 - \sum_{k=1}^{2}(f_3, \bar{f}_k)\bar{f}_k, \bar{f}_1\right) \\ &= \frac{1}{\|h_3\|}[(f_3, \bar{f}_1) - (f_3, \bar{f}_1)(\bar{f}_1, \bar{f}_1) - (f_3, \bar{f}_2)(\bar{f}_2, \bar{f}_1)] \\ &= \frac{1}{\|h_3\|}[(f_3, \bar{f}_1) - (f_3, \bar{f}_1)] \\ &= 0\end{aligned}$$

$$\begin{aligned}(\bar{f}_3, \bar{f}_2) = \frac{1}{\|h_3\|}(h_3, \bar{f}_2) &= \frac{1}{\|h_3\|}[(f_3, \bar{f}_2) - (f_3, \bar{f}_1)(\bar{f}_1, \bar{f}_2) - (f_3, \bar{f}_2)(\bar{f}_2, \bar{f}_2)] \\ &= \frac{1}{\|h_3\|}[(f_3, \bar{f}_2) - (f_3, \bar{f}_2)] \\ &= 0\end{aligned}$$

Following this line, we can prove that for $i \leq n - 1, j < i$, $(\bar{f}_i, \bar{f}_j) = 0$

Now for $i = n, j \leq n - 1$, we have

$$\begin{aligned}(\bar{f}_n, \bar{f}_j) = \frac{1}{\|h_n\|}(h_n, \bar{f}_j) &= \frac{1}{\|h_n\|}\left(f_n - \sum_{k=1}^{n-1}(f_n, \bar{f}_k)\bar{f}_k, \bar{f}_j\right) \\ &= \frac{1}{\|h_n\|}\left[(f_n, \bar{f}_j) - \sum_{k=1}^{n-1}(f_n, \bar{f}_k)(\bar{f}_k, \bar{f}_j)\right] \\ &= \frac{1}{\|h_n\|}[(f_n, \bar{f}_j) - (f_n, \bar{f}_j)] \\ &= 0\end{aligned}$$

Thus we prove that $(\bar{f}_i, \bar{f}_j) = \delta_{ij}$.

Appendix B

Green's Second Identity on a Curved Surface in an Orthogonal Curvilinear Coordinate System

1. Background

In the development of a set of orthogonal functions on the surface of a sphere, an important identity derived below, will be employed. The identity states that

$$\iint_{\sigma} \left(f\Delta^* h - g\Delta^* f\right) d\sigma = \oint_C \left(f\frac{\partial h}{\partial n} - h\frac{\partial f}{\partial n}\right) ds \tag{B.1}$$

where Δ^* is the Laplace surface operator on the unit sphere, σ is a region on the unit sphere, C is the boundary of σ, and n is the outer normal of line C, lying on the tangent plane of C. The situation is shown in Figure B.1. Δ^* is defined as

$$\Delta^* = \frac{1}{\sin\theta}\left[\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\lambda^2}\right] \tag{B.2}$$

where θ is the colatitude and λ is the longitude. On a 2-D rectangular coordinate system (B.1) becomes the well-known Green's second identity on a plane, which has been proven by many authors (e.g., Kaplan, 1981, p. 814). The purpose of this appendix is to extend Green's second identity to a curved surface that is defined by some curvilinear coordinates. The extension to the surface of the unit sphere will help some of the developments in Chapter 3.

2. Proof of the identity

Consider a regular parametric representation of a surface σ in E^3:

$$x^i = x^i(u^1, u^2) \qquad i = 1, 2, 3 \tag{B.3}$$

where u^1 and u^2 are two parameters. x^i form a rectangular coordinate system in E^3. On σ, the arc length form is

$$(ds)^2 = g_{ij}du^i du^j,\ i, j = 1, 2 \tag{B.4}$$

where g_{ij} is the covariant metric tensor. In (B.4), the summation convention of tensor calculus has been employed and we will do so in the following development. Let $r = (x^1, x^2, x^3)$, and $r_i = \partial r/\partial u^i$, g_{ij} may be obtained from

$$g_{ij} = r_i \cdot r_j,\ i,\ j = 1, 2 \tag{B.5}$$

(B.4) is the well-known First Fundamental Form of a surface.

Assume u^1 and u^2 are orthogonal, then g_{ij} will vanish if $i \neq j$. Under this condition, we re-write (B.4) as

$$(ds)^2 = g_{11}(du^1)^2 + g_{22}(du^2)^2 \tag{B.6}$$

Let e_1 be a unit tangent vector along u^1 and e_2 a unit tangent vector along u^2, then the unit tangent vector and unit normal vector of an arc C on σ can be written as

$$e_t = \sqrt{g_{11}}\frac{du^1}{ds}e_1 + \sqrt{g_{22}}\frac{du^2}{ds}e_2 \tag{B.7}$$

$$e_n = e_t x e_3 = \sqrt{g_{22}}\frac{du^2}{ds}e_1 - \sqrt{g_{11}}\frac{du^1}{ds}e_2 \tag{B.8}$$

where e_3 is a unit vector normal to the plane expanded by vectors e_1 and e_2. e_n is obtained by rotating e_t by 90° clockwise on the e_1 - e_2 plane. "x" denotes the cross-product.

Stokes' theorem (Spiegel, 1959, p. 106) states that the surface integral of the curl of a vector A taken over any surface σ is equal to the integral of A taken around the periphery of the surface. The theorem can be expressed as

$$\iint_\sigma (\nabla x A)\cdot e_3 d\sigma = \oint_C A\cdot e_t ds \tag{B.9}$$

Since e_1, e_2 and e_3 form a local orthogonal system in E^3, we can express vector A in the e_i-basis:

$$A = A_1e_1 + A_2e_2 + A_3e_3 \tag{B.10}$$

Let $\overline{A} = A_1e_1 + A_2e_2$. We rotate $\overline{A}$ on the e_1 - e_2 plane by 90° clockwise to obtain $\overline{B}$, defined as

$$\overline{B} = A_2e_1 - A_1e_2 \tag{B.11}$$

By the definitions of e_t and e_n, it is easy to see that

$$A\cdot e_t = \overline{A}\cdot e_t = \overline{A}_t = \overline{B}\cdot e_n = \overline{B}_n \tag{B.12}$$

where $\overline{A}_t$ is the projection of vector $\overline{A}$ on e_t, $\overline{B}_n$ is the projection of $\overline{B}$ on e_n.

For $(\nabla x A)\cdot e_3$, again we assume that we are working in the e_i-basis and we are able to obtain the "scale factor" $\sqrt{g_{33}}$ for vector e_3. With such an assumption, we can find $(\nabla x A)$ from (Spiegel, 1959)

$$\nabla \times A = \frac{1}{\sqrt{g_{11}g_{22}g_{33}}} \begin{vmatrix} \sqrt{g_{11}}\, e_1 & \sqrt{g_{22}}\, e_2 & \sqrt{g_{33}}\, e_3 \\ \dfrac{\partial}{\partial u^1} & \dfrac{\partial}{\partial u^2} & \dfrac{\partial}{\partial u^3} \\ \sqrt{g_{11}}\, A_1 & \sqrt{g_{22}}\, A_2 & \sqrt{g_{33}}\, A_3 \end{vmatrix} \tag{B.12}$$

where u^3 is the curvilinear coordinate associated with e_3. From (B.12), we get

$$(\nabla \times A)\cdot e_3 = \frac{1}{\sqrt{g_{11}g_{22}}}\left[\frac{\partial}{\partial u^1}(\sqrt{g_{22}}\, A_2) - \frac{\partial}{\partial u^2}(\sqrt{g_{11}}\, A_1)\right] \tag{B.13}$$

Clearly g_{33} and u^3 plays no role in our derivation. With (B.11) and (B.13), we can rewrite (B.9) as

$$\iint_\sigma \frac{1}{\sqrt{g_{11}g_{22}}}\left[\frac{\partial}{\partial u^1}(\sqrt{g_{22}}\, A_2) - \frac{\partial}{\partial u^2}(\sqrt{g_{11}}\, A_1)\right] d\sigma = \oint_C (A_2 e_1 - A_1 e_2)\cdot e_n\, ds \tag{B.14}$$

Assume $f(u^1,u^2)$ and $h(u^1,u^2)$ are two functions defined on σ. By the substitution

$$A_2 = f\frac{1}{\sqrt{g_{11}}}\frac{\partial h}{\partial u^1} \tag{B.15}$$

$$A_1 = -f\frac{1}{\sqrt{g_{22}}}\frac{\partial h}{\partial u^2} \tag{B.16}$$

and the definition of gradient on σ

$$\nabla = \frac{1}{\sqrt{g_{11}}}\frac{\partial}{\partial u^1} e_1 + \frac{1}{\sqrt{g_{22}}}\frac{\partial}{\partial u^2} e_2 \tag{B.17}$$

we get

$$\oint_C (A_2 e_1 - A_1 e_2)\cdot e_n ds = \oint_C f\frac{\partial h}{\partial n} ds \tag{B.18}$$

where $\partial h/\partial n$ is the derivative of h along e_n. On the other hand, the left-hand side of (B.14) is

$$\iint_\sigma \frac{1}{\sqrt{g_{11}g_{22}}}\left[\frac{\partial}{\partial u^1}\left(f\sqrt{\frac{g_{22}}{g_{11}}}\frac{\partial h}{\partial u^1}\right)+\frac{\partial}{\partial u^2}\left(f\sqrt{\frac{g_{11}}{g_{22}}}\frac{\partial h}{\partial u^2}\right)\right]d\sigma$$

$$=\iint_\sigma \frac{f}{\sqrt{g_{11}g_{22}}}\left[\frac{\partial}{\partial u^1}\left(\sqrt{\frac{g_{22}}{g_{11}}}\frac{\partial h}{\partial u^1}\right)+\frac{\partial}{\partial u^2}\left(\sqrt{\frac{g_{11}}{g_{22}}}\frac{\partial h}{\partial u^2}\right)\right]d\sigma$$

$$+\iint_\sigma \left(\frac{1}{g_{11}}\frac{\partial h}{\partial u^1}\frac{\partial f}{\partial u^1}+\frac{1}{g_{22}}\frac{\partial h}{\partial u^2}\frac{\partial f}{\partial u^2}\right)d\sigma \qquad \text{(B.19)}$$

Recalling that the Laplace operator in a curvilinear system is (Kay, 1988, p. 157):

$$\Delta f = \frac{1}{\sqrt{g}}\frac{\partial}{\partial u^i}\left(\sqrt{g}\, g^{ij}\frac{\partial f}{\partial u^j}\right) \qquad \text{(B.20)}$$

where g is the determinant of the covariant metric tensor g_{ij}, and g^{ij} is the contravariant metric tensor. g^{ij} can be obtained by inverting g_{ij}:

$$(g^{ij})_{m\times m} = (g_{ij})^{-1}_{m\times m} \qquad \text{(B.21)}$$

where m is the order of the matrix. For an orthogonal system,

$$g^{ij} = g_{ij} = 0, \quad i \neq j \qquad \text{(B.22)}$$

Thus it is easy to see that (for an orthogonal system)

$$g^{ii} = \frac{1}{g_{ii}} \qquad \text{(B.23)}$$

Now, we define the Laplace surface operator $\Delta^* f$ as

$$\Delta^* f = \frac{1}{\sqrt{g}}\frac{\partial}{\partial u^i}\left(\sqrt{g}\, g^{ij}\frac{\partial f}{\partial u^j}\right)$$

$$=\frac{1}{\sqrt{g_{11}g_{22}}}\left[\frac{\partial}{\partial u^1}\left(\sqrt{\frac{g_{22}}{g_{11}}}\frac{\partial f}{\partial u^1}\right)+\frac{\partial}{\partial u^2}\left(\sqrt{\frac{g_{11}}{g_{22}}}\frac{\partial f}{\partial u^2}\right)\right] \qquad \text{(B.24)}$$

Combining (B.18), (B.19) and (B.24), we have

$$\iint_{\sigma} f\Delta^{*}h\,d\sigma + \iint_{\sigma}\left(\frac{1}{g_{11}}\frac{\partial h}{\partial u^{1}}\frac{\partial f}{\partial u^{1}} + \frac{1}{g_{22}}\frac{\partial h}{\partial u^{2}}\frac{\partial f}{\partial u^{2}}\right)d\sigma = \oint_{C} f\frac{\partial h}{\partial n}\,ds \tag{B.25}$$

Interchanging f and h in (B.25), we get

$$\iint_{\sigma} h\Delta^{*}f\,d\sigma + \iint_{\sigma}\left(\frac{1}{g_{11}}\frac{\partial h}{\partial u^{1}}\frac{\partial f}{\partial u^{1}} + \frac{1}{g_{22}}\frac{\partial h}{\partial u^{2}}\frac{\partial f}{\partial u^{2}}\right)d\sigma = \oint_{C} h\frac{\partial f}{\partial n}\,ds \tag{B.26}$$

Subtracting (B.26) from (B.25), we finally obtain the Green's second identity on a curved surface:

$$\iint_{\sigma}\left(f\Delta^{*}h - h\Delta^{*}f\right)d\sigma = \oint_{C}\left(f\frac{\partial h}{\partial n} - h\frac{\partial f}{\partial n}\right)ds \tag{B.27}$$

Assume now σ is on the unit sphere and $u^1 = \theta$, $u^2 = \lambda$, it can be shown that (Kay, 1988, p. 147)

$$g_{11}=1$$

$$g_{22}=\sin^{2}\theta$$

Thus on the unit sphere, the Laplace surface operator is

$$\Delta^{*}f = \frac{1}{\sin\theta}\left[\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{\sin\theta}\frac{\partial^{2}f}{\partial\lambda^{2}}\right] \tag{B.28}$$

From the above derivations, we know that (B.27) is not only valid for a sphere, but also valid for other surfaces (e.g., ellipsoid) which are parameterized by two orthogonal curvilinear coordinates. The author could not find the existing derivations for this identity in literature, thus the detailed derivations are given here.

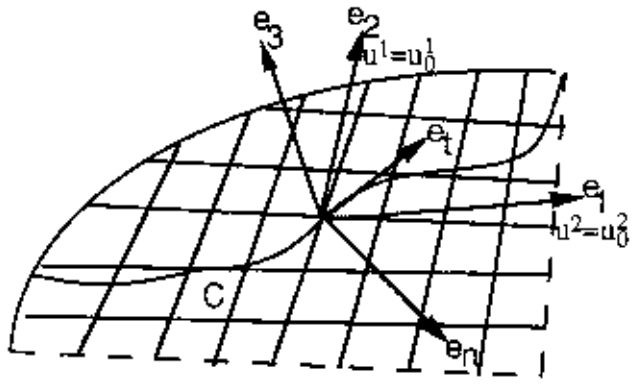


Figure B.1 Coordinate System and Vectors on a Surface Parameterized by u^1 and u^2.

Appendix C

Bessel Functions and Orthogonality Relationship

Here we will be proving the orthogonality of two Bessel functions of two distinct frequency (zeros), but of the same order. The Bessel equation is (Lebedev, 1972)

$$J''(r) + \frac{1}{r}J'(r) + \left(1 - \frac{n^2}{r^2}\right)J(r) = 0 \tag{C.1}$$

where n can be any real or complex number. The argument r can also be any real or complex variable. For this study, we restrict the case to the real n and r. n is the order of the solution J(r). Thus we often denote the solution of (C.1) as $J_n(r)$ to emphasize the order of the Bessel function.

By changing the variable $r = kx$ and denoting the new function as $Z_k(x) = J_n(kx)$, we obtain a modified form of the Bessel equation (of the same order):

$$Z_k''(x) + \frac{1}{x}Z_k'(x) + \left(k^2 - \frac{n^2}{x^2}\right)Z_k(x) = 0 \tag{C.2}$$

By another substitution $r = \ell x$ and notation $Z_\ell(x) = J_n(\ell x)$, we arrive at another equation:

$$Z_\ell''(x) + \frac{1}{x}Z_\ell'(x) + \left(\ell^2 - \frac{n^2}{x^2}\right)Z_\ell(x) = 0 \tag{C.3}$$

Multiplying (C.2) by xZ_ℓ and (C.3) by xZ_k and subtracting one equation from the other yield:

$$\left(\ell^2 - k^2\right)xZ_\ell Z_k = x\left(Z_\ell Z_k'' - Z_k Z_\ell''\right) + \left(Z_\ell Z_k' - Z_k Z_\ell'\right) \tag{C.4}$$

Without loss of generality, we assume $0 \le x \le 1$. Taking integrals on both sides of (C.4) and applying integration by part, we have

$$\left(\ell^2 - k^2\right)\int_0^1 xZ_\ell Z_k dx = x\left(Z_\ell Z_k' - Z_k Z_\ell'\right)\Big|_0^1 = \left(Z_\ell Z_k' - Z_k Z_\ell'\right)\Big|_{x=1} \tag{C.5}$$

Recalling that $Z_\ell = J_n(\ell x)$, $Z_k = J_n(kx)$ and $Z_\ell' = \ell J_n'(\ell x)$, $Z_k' = kJ_n'(kx)$, we write (C.5) as

$$\left(\ell^2 - k^2\right)\int_0^1 xJ_n(\ell x)J_n(kx)dx = kJ_n(\ell)J_n'(k) - \ell J_n(k)J_n'(\ell) \tag{C.6}$$

The solution of (C.1), i.e., $J_n(r)$, for any n is well-known in literature. Normally $J_n(r)$ contains an infinite power series in r. For example, for a non-negative integer n, $J_n(r)$ is the Bessel function (of the first kind)

$$J_n(r) = \sum_{i=0}^{\infty} \frac{(-1)^i}{(i)!(i+n)!}\left(\frac{r}{2}\right)^{n+2i} \tag{C.7}$$

Using such an analytical form of $J_n(r)$, we are able to find the zeros (roots). Let the zeros be r_{nm}, m = 1, ... ,∞, and if k and ℓ in (C.6) happen to be two of r_{nm}, then we see that

$$(r_{n\ell}^2 - r_{nk}^2)\int_0^1 xJ_n(r_{n\ell}x)J_n(r_{nk}x)dx = 0 \tag{C.8}$$

Since $r_{n\ell} \neq r_{nk}$, we conclude that $J_n(r_{n\ell}x)$ and $J_n(r_{nk}x)$ are orthogonal with respect to weight x in the domain $0 \leq x \leq 1$.

For the normalizing factor, we apply L'Hôpital's rule in (C.6):

$$\begin{aligned}\int_0^1 xJ_n^2(kx)dx &= \lim_{\ell \to k} \frac{1}{\ell^2 - k^2}\left[kJ_n(\ell)J_n'(k) - \ell J_n(k)J_n'(\ell)\right] \\ &= \frac{1}{2k}\left[k\left(J_n'(k)\right)^2 - J_n(k)J_n'(k) - kJ_n(k)J_n''(k)\right]\end{aligned} \tag{C.9}$$

If we regard k as a variable and hence $J_n(k)$ satisfies (C.1), we have

$$J_n''(k) + \frac{1}{k}J_n'(k) + \left(1 - \frac{n^2}{k^2}\right)J_n(k) = 0 \tag{C.10}$$

Thus

$$kJ_n(k)J_n''(k) + J_n(k)J_n'(k) = -\left(1 - \frac{n^2}{k^2}\right)kJ_n^2(k) \tag{C.11}$$

Substituting (C.11) into (C.9), we get

$$\int_0^1 xJ_n^2(kx)dx = \frac{1}{2}\left[\left(J_n'(k)\right)^2 + \left(1 - \frac{n^2}{k^2}\right)J_n^2(k)\right] \tag{C.12}$$

If k is one of the zeros for $J_n(r) = 0$, then from the recursive formula (Lebedev, 1972, p. 100)

$$J_n'(k) - \frac{n}{k}J_n(k) = -J_{n+1}(k) \tag{C.13}$$

We get (since $J_n(k) = 0$)

$$J_n'(k) = -J_{n+1}(k) \tag{C.14}$$

Therefore, from (C.12) and (C.14), we obtain the normalizing factor for the Bessel function:

$$\int_0^1 xJ_n^2(kx)dx = \frac{1}{2} J_{n+1}^2(k) \tag{C.15}$$

where k must be a zero of $J_n(r)$.

List of References

Banerji, S., Note on the Product of any Number of Legendre Functions of Different Degrees, Bull. Calcutta Math. Soc., V. 10(1918-19), pp. 179-185, 1920.

Bath, M., Spectral Analysis in Geophysics, Elsevier Sci. Pub. Co., New York, 1974.

Beyer, W.H., CRC Standard Mathematical Tables, 28th Ed., CRC Press, Inc., Boca Raton, 1987.

Bieberbach, L., Conformal Mapping, Chelsea Pub. Co., New York, 1953.

Brouwer, D., Solution of the Problem of Artificial Satellite Theory Without Drag, Astron. J., V. 64, pp. 378-397, 1959.

Brouwer, D. and G.M. Clemence, Methods of Celestial Mechanics, Academic Press, New York, 1961.

Cheney, R.E., B.C. Douglas, R.W. Agreen, L. Miller and N.S. Doyle, The NOAA Geosat Geophysical Data Records—User Handbook, NOAA Technical Memorandum NOS NGS-46, Rockville, Maryland, 1987.

Churchill, R.V., and J.W. Brown, Complex Variables and Applications, McGraw-Hill Book Co., New York, 1984.

Colombo, O., Numerical Methods for Harmonic Analysis on the Sphere, Dept. of Geodetic Science and Surveying, Rep. No. 310, The Ohio State University, Columbus, 1981.

Colombo, O., Altimetry, Orbits and Tides, NASA Technical Memorandum 86180, Greenbelt, Maryland, 1984.

Cook, G.E., Perturbations of Near-Circular Orbits by the Earth's Gravitational Field, Planet. and Space Sci., V. 14, pp. 433-444, 1966.

Courant, R. and D. Hilbert, Methods of Mathematical Physics, Vol. I, Interscience Pub., Inc., New York, 1953.

Cui, C.H., Die Bewegung künstlicher Satelliten im anisotropen Gravitationsfelde einer gleichmäßig Rotierenden starren Modellerde. Deutsche Geod. Komm., Reihe C, No. 357, Munich, 1990.

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

Denker, H., Radial Orbit Error Reduction and Sea Surface Topography Determination Using One Year of Geosat Altimeter Data, Dept. of Geodetic Science and Surveying, Rep. No. 404, The Ohio State University, Columbus, 1990.

Denker, H. and R.H. Rapp, Geodetic and Oceanographic Results from the Analysis of One Year of Geosat Data, J. Geophys. Res., V. 95, No. C8, pp. 13151-13168, 1990.

Dettman, J.W., Applied Complex Variables, The MacMillan Co., New York, 1965.

Dettman, J.W., Mathematical Methods in Physics and Engineering, McGraw-Hill, Inc., New York, 1988.

Dongarra, J.J., E.B. Moler, J.R. Bunch and G.W. Stewart, Linpack User's Guide, The Society for Industrial and Applied Math. (SIAM), Philadelphia, 1979.

Engelis, T., Radial Orbit Error Reduction and Sea Surface Topography Determination Using Satellite Altimetry, Dept. of Geodetic Science and Surveying, Rep. No. 377, The Ohio State University, Columbus, 1987a.

Engelis, T., Spherical Harmonic Expansion of the Levitus Sea Surface Topography, Dept. of Geodetic Science and Surveying, Rep. No. 385, The Ohio State University, Columbus, 1987b.

Engelis, T., On the Simultaneous Improvement of a Satellite Orbit and Determination of Sea Surface Topography Using Altimeter Data, manuscripta geodaetica, V. 13, pp. 180-190, 1988.

Engelis, T.and P. Knudsen, Orbit Improvement and Determination of the Ocean Geoid and Topography from 17 Days of Seasat Data, manuscripta geodaetica, V. 14, pp. 193-201, 1989.

Freeden, W., On Spherical Spline Interpolation and Approximation, Math. Meth. in the Appl. Sci., V. 3, No. 4, pp. 551-575, B.G. Teubner Stuttgart, 1981.

Fu, L.L., E.J. Christensen and M. Lefebvre, TOPEX/POSEIDON: The Ocean Topography Experiment, EOS, V. 72, No. 35, American Geophysical Union, 1991.

Gaier, D., Lectures on Complex Approximation, Birkhäuser Boston, Inc., 1987.

Garfinkel, B., The Orbit of a Satellite of an Oblate Planet, Astron. J., V. 64, pp. 353-367, 1959.

Gelfand, I.M. and S.V. Fomin, Calculus of Variations, Prentice-Hall, Englewood Cliffs, 1962.

Gerstl, M., On the Recursive Computation of the Integrals of the Associated Legendre Functions, manuscripta geodetica, V. 5, pp. 181-199, 1980.

Giacaglia, G.E.D., Transformations of Spherical Harmonics and Applications to Geodesy and Satellite Theory, studia geogh. et geod., V. 24, pp. 1-11, 1980.

Goad, C.C., An Efficient Algorithm for the Evaluation of Inclination and Eccentricity Functions, manuscripta geodaetica, V. 12, pp. 11-15, 1987.

Goad, C.C. and A. Mueller, An Automated Procedure for Generating an Optimum Set of Independent Double Difference Observables Using Global Positioning System Carrier Phase Measurements, manuscripta geodetica, V. 13, pp. 365-369, 1988.

Goldstein, H., Classical Mechanics, Addison-Wesley Publishing Company, Reading, Massachusetts, 1950.

Gooding, R.H., and D.G. King-Hele, Explicit Forms of Some Functions Arising in the Analysis of Resonant Satellite Orbits, Royal Aero. Est., Technical Report 88035, 1988.

Haines, G.V., Spherical Cap Harmonic Analysis, J. Geophys. Res., V. 90, No. B3, pp. 2583-2591, 1985a.

Haines, G.V., Magsat Vertical Field Anomalies Above 40°N from Spherical Cap Harmonic Analysis, J. Geophys. Res., V. 90, No. B3, pp. 2593-2598, 1985b.

Haines, B.J., G.H. Born, J.G. Marsh, R.G. Williamson, Precise Geosat Orbits for the Exact Repeat Mission, J. Geophys. Res., V. 95, No. C3, pp. 2871-2886, 1990.

Heiskanen, W.A. and H. Moritz (H/M), Physical Geodesy, W.H. Freeman, New York, 1967.

Hobson, E.W., The Theory of Spherical and Ellipsoidal Harmonics, Chelsea Publishing Co., New York, Second Reprint, 1965.

Kaplan, W., Advanced Mathematics for Engineers, Addison-Wesley Publishing Co., 1981.

Katsambalos, K.E., The Effect of the Smoothing Operator on Potential Coefficient Determinations, Dept. of Geodetic Science, Rep. No. 287, The Ohio State University, Columbus, 1979.

Kaula, W.M., Theory of Satellite Geodesy, Blaisdell Publishing Co., 1966.

Kay, D.C., Tensor Calculus, Schaum's Outline Series, McGraw-Hill, Inc., 1988.

Koblinsky, C.J., L.E. Braatz, T.L. Engelis, S.M. Klosko and R.G. Williamson, Geocenter Definition in the Determination of Dynamic Height Using Geosat Satellite Altimetry, paper presented at the Fall meeting of the AGU, San Francisco, December, 1989.

Koblinsky, C., J.G. Marsh, B. Beckley, A. Brenner, and R. Williamson, Geosat Orbit Replacement Software for Altimeter Geophysical Data Records, GEM-T2 Ephemerides for November 1986 to November 1988, Geodynamics Branch, Goddard Space Flight Center, Greenbelt, MD 20771, 1990.

Kolmogorov, A.N. and S.V. Fomin, Introductory Real Analysis, Dover Publications, Inc., New York, 1970.

Kosteleck, J., J. Klocôvík, Z. Kalina, Computation of Normalized Inclination Function to High Degree for Satellites in Resonances, manuscripta geodaetica, V. 11, pp. 293-304, 1986.

Kozai, Y., The Motion of a Close Earth Satellite, Astron. J., V. 64, pp. 367-377, 1959.

Lebedev, N.N., Special Functions and Their Applications, Dover Publications, Inc., New York, 1972.

Lerch, F.J., Error Spectrum of Goddard Satellite Models for the Gravity Field, NASA Technical Memorandum 86223, Greenbelt, Maryland, 1985.

Lerch, F.J., Optimum Data Weighting and Error Calibration for Estimation of Gravitational Parameters, Bulletin Géodésique, V. 65, pp. 44-52, 1991.

Levitus, S., Climatological Atlas of the World Ocean, NOAA Professional Paper 13, U.S. Dept. of Commerce, Rockville, 1982.

Lipschutz, M.M., Differential Geometry, McGraw-Hill, Inc., New York, 1969.

Lisitzin, E., Sea Level Changes, Elsevier Oceanography Series, Amsterdam, 1974.

Mainville, A., The Altimetery-Gravimetry Problem Using Orthonormal Base Functions, Dept. of Geodetic Science and Surveying, Rep. No. 373, The Ohio State University, Columbus, 1987.

Marsh, J.G., et al., The GEM-T2 Gravitational Model, NASA Technical Memorandum 100746, Washington, DC, 1989.

Marsh, J.G., C.J. Koblinsky, F. Lerch, S.M. Klosko, J.W. Robbins, R.G. Williamson, and G.B. Patel, Dynamic Sea Surface Topography, Gravity, and Improved Orbit Accuracies From the Direct Evaluation of Seasat Altimeter Data, J. Geophys. Res., V. 95, No. C8, pp. 13129-13150, 1990.

Marshall, J.C., Determining the Ocean Circulation and Improving the Geoid from Satellite Altimetry, J. Phys. Ocean., V. 15, pp. 330-349, 1985.

Mayhan, R.J., Discrete-Time and Continuous-Time Linear Systems, Addison-Wesley Publishing Co., 1984.

Meissel, P., On the Linearization of the Geodetic Boundary Value Problem, Dept. of Geodetic Science, Rep. No. 152, The Ohio State University, Columbus, 1971.

Mikhail, E.M., Observations and Least Squares, IEP-Dun-Donnelley, Harper & Row, New York, 1976.

Miller, W., Symmetry and Separation of Variables, Addison-Wesley Publishing Co., 1977.

Moritz, H., Advanced Physical Geodesy, Herbert Wechmann Verlag Karlsrule, FRG, 1980.

Morse, P.M. and H. Fesbach, Methods of Theoretical Physics, McGraw-Hill Inc., New York, 1953.

Müller, C., Lectures on the Theory of Spherical Harmonics, Boeing Sci. Res. Lab., Math. Note No. 276, 1962.

Nerem, R.S., Determination of the General Ocean Circulation Using Satellite Altimetry from a Simultaneous Solution for the Earth's Gravity Field, Ph.D. dissertation, Univ. of Tex. at Austin, May 1989.

Nerem, R.S., B.D. Tapley and C.K. Shum, Determination of the Ocean Circulation Using Geosat Altimetry, J. Geophys. Res., V. 95, No. B3, pp. 3163-3179, 1990.

Neumann, G., Ocean Currents, Elsevier Oceanography Series Vol. 4, Elsevier Publishing Co., 1968.

Officer, C. B., Introduction to Theoretical Geophysics, Springer-Verlag, New York, 1974.

Page, C.H., Physical Mathematics, D. Van Nostrand Co., Inc., 1955.

Pal, B., On the Numerical Calculation of the Roots of the Equations $P_n^m(u) = 0$ and $(d/du)P_n^m(u) = 0$ Regarded as Equations in n, Bull. Calcutta Math. Soc., V. 10(1918-19), pp. 85-95, 1920.

Papoulis, A., Signal Analysis, McGraw Hill, Inc., New York, 1977.

Paul, M.K., Recurrence Relations for Integrals of Associated Legendre Functions, Bulletin Géodésique, V. 52, pp. 177-190, 1978.

Pavlis, N.K., Modeling and Estimation of a Low Degree Geopotential Model from Terrestrial Gravity Data, Dept. of Geodetic Science and Surveying, Rep. 386, The Ohio State University, Columbus, 1988.

Pellinen, L.P., A Method for Expanding the Gravity Potential of the Earth in Spherical Functions, Translated from Russian, ACIC-TC-1282, NTIS: AD-661819, Moscow, 1966.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical Recipes, The Art of Scientific Computing, Cambridge University Press, Cambridge, 1989.

Rapp, R.H., Combination of Satellite, Altimetric and Terrestrial Gravity Data, Lecture Notes in Earth Sciences, Vol. 25, pp. 261-284, Springer-Verlag, New York, 1989b.

Rapp, R.H., Geometric Geodesy, Vol. I & II, Dept. of Geodetic Science and Surveying, The Ohio State University, Columbus, 1989a.

Rapp, R.H., Global Geopotential Solutions, Lecture Notes in Earth Sciences, Vol. 7, pp. 365-415, Springer-Verlag, New York, 1986.

Rapp, R.H., The Treatment of Permanent Tidal Effects in the Analysis of Satellite Altimeter Data for Sea Surface Topography, manuscripta geodaetica, V. 4, No. 6, pp. 368-372, 1989c.

Rapp, R.H. and N. K. Pavlis, The Development and Analysis of Geopotential Coefficient Models to Spherical Harmonic Degree 360, J. Geophys. Res., V. 95, No. B3, pp. 21889-21911, 1990.

Rapp, R.H., Y.M. Wang and N. Pavlis, The Ohio State 1991 Geopotential and Sea Surface Topography Harmonic Coefficient Models, Dept. of Geodetic Science and Surveying, Rep. No. 410, The Ohio State University, Columbus, 1991.

Reigber, C., Gravity Field Recovery from Satellite Tracking Data, Lecture Notes in Earth Sciences, Vol. 25, pp. 197-234, Springer-Verlag, New York, 1989.

Ricardi, L.J. and M.L. Burrows, A Recurrence Technique for Expanding a Function in Spherical Harmonics, IEEE Trans. on Computers, June 1972.

Rivlin, T.J., An Introduction to the Approximation of Functions, Dover Publications, Inc., New York, 1981.

Sacerdote, F. and F. Sansò, Overdetermined Boundary Value Problems in Physical Geodesy, manuscripta geodaetica, V. 10, pp. 195-207, 1985.

Sansò, F. and B. Stock, A Numerical Experiment in the Altimetry-Gravimetry Problem II, manuscripta geodaetica, V. 10, pp. 23-31, 1985.

Sansone, G., Orthogonal Functions, Translated from the Italian, Interscience Publishers, Inc., New York, 1959.

Schaffrin, B., E. Heidenreich, and E. Grafarend, A Representation of the Standard Gravity Field, manuscripta geodaetica, V. 2, No. 2, pp. 135-174, 1977.

Schaffrin, B., An Alternative Approach to Robust Collocation, Bulletin Géodésique, V. 63, pp. 395-404, 1989.

Schrama, E.J.O., Estimability of Potential Coefficients from Orbit Perturbations, Rep. No. 86.1, Dept. of Geodesy, Delft University of Technology, Delft, The Netherlands, 1986.

Schrama, E.J.O., The Role of Orbit Errors in Processing of Satellite Altimeter Data, Publ. on Geodesy, New Series No. 33, Netherlands Geod. Comm., Delft, 1989.

Schwarz, K.P., Tesseral Harmonic Coefficients and Station Coordinates from Satellite Observations by Collocation, Dept. of Geodetic Science and Surveying, Rep. No. 217, The Ohio State University, Columbus, 1974.

Smirnov, V.I. and N.A. Lebedev, Functions of a Complex Variable, Translated by Scripta Technica Ltd., The M.I.T Press, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1968.

Sneeuw, N.J., Inclination Functions: Group Theoretical Background and a Recursive Algorithm, Report of the Faculty of Geodetic Engineering, No. 912, Delft University of Technology, Delft, The Netherlands, 1991.

Sommerfeld, A., Partial Differential Equations in Physics, Lectures on Theoretical Physics, Vol. VI, Academic Press, Inc., New York, 1949.

Spiegel, M.R., Vector Analysis, Schaum's Outline Series, McGraw-Hill, Inc., New York, 1959.

Spiegel, M.R., Complex Variables, Schaum's Outline Series, McGraw-Hill, Inc., New York, 1964.

Spiegel, M.R., Theoretical Mechanics, Schaum's Outline Series, McGraw-Hill, Inc., New York, 1959.

Stewart, G.W., Introduction to Matrix Computations, Academic Press, Inc., New York, 1973.

Synge, J.L. and A. Schild, Tensor Calculus, Dover Publications, Inc., New York, 1978.

Taylor, A.E., General Theory of Functions and Integration, Blaisdell Publishing Co., New York, 1965.

Thomson, S.W. and P.G. Tait, Treatise on Natural Philosophy, Vol. I, Cambridge: At the University Press, 1879.

Thong, N.C., and E.W. Grafarend, A Spheroidal Harmonic Model of the Terrestrial Gravitational Field, manuscripta geodaetica, V. 14, No. 5, pp. 285-304, 1989.

Tolstov, G.P., Fourier Series, Dover Publications, Inc., New York, 1976.

Tranter, C.J., Bessel Functions with Some Physical Applications, Hart Publishing Company, Inc., New York, 1968.

Uotila, U.A., Notes on Adjustment Computation, Part I, Dept. of Geodetic Science and Surveying, The Ohio State University, Columbus, 1986.

Wagner, C.A., Radial Variations of a Satellite Orbit Due to Gravitational Errors: Implications for Satellite Altimetry, J. Geophys. Res., V. 90, No. B4, pp. 3027-3036, 1985.

Wagner, C.A., Accuracy Estimate of Geoid and Ocean Topography Recovered Jointly from Satellite Altimetry, J. Geophys. Res., V. 91, No. B1, pp. 453-461, 1986.

Wagner, C.A., Lectures on Satellite Altimetry. Lecture Notes in Earth Sciences, Vol. 25, pp. 285-334, Springer-Verlag, New York, 1989.

Wagner, C.A., Private Communication, February 1991.

Widder, D.V., Advanced Calculus, 2nd Ed., Dover Publications, Inc., New York, 1989.

Wieser, M., The Global Digital Terrain Model TUG87, Internal Report on set-up, origin and characteristics, Dept. of Mathematical Geodesy, The Technical University in Graz, 1987.

Wunsch, C. and E.M. Gaposchkin, On Using Satellite Altimetry to Determine the General Circulation of the Oceans with Application to Geoid Improvement, Reviews of Geophys. and Space Phys., V. 18, pp. 725-745, 1980.

Wunsch, C., Global-Scale Sea Surface Variability From Combined Altimetric and Tide Gauge Measurements, J. Geophys. Res., V. 96, No. C8, pp. 15053-15082, 1991.

Zienkiewicz, O. and K. Morgan, Finite Elements and Approximation, John Wiley & Sons, New York, 1983.