PHD

Asymptotic and numerical methods for high-frequency scattering problems

Kim, Tatiana

Award date:
2012

Awarding institution:
University of Bath

Asymptotic and Numerical methods for high-frequency scattering problems

submitted by

Tatiana Kim

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Mathematical Sciences

March 2012

Signature of Author ..... Tatiana Kim

Abstract

This thesis is concerned with the development, analysis and implementation of efficient and accurate numerical methods for solving high-frequency acoustic scattering problems.

Classical boundary or finite element methods that are based on approximating the solution by polynomials can be effective for small and moderate frequencies. However, as the frequency increases, the solution to the scattering problem becomes more oscillatory and classical numerical methods cope very badly with high oscillation. For example, for two-dimensional scattering problems, classical numerical methods require their number of degrees of freedom to grow at least linearly with frequency to capture the oscillatory behaviour of the solution accurately. Therefore, at large frequencies, classical numerical methods become essentially numerically intractable.

In order to overcome the limitations of classical methods, one can seek to incorporate the known asymptotic behaviour of the solution in the numerical method. This involves using asymptotic theory to determine the oscillatory part of the solution and then using classical numerical methods to approximate the slowly varying remainder. Such methods are often referred to as *hybrid numerical-asymptotic methods*.

Determining the high frequency asymptotics of acoustic scattering problems is a classic problem in applied mathematics, with methods such as geometrical optics or the geometrical theory of diffraction providing asymptotic expansions of the solutions. Considerable amount of research has been directed towards both constructing these asymptotic expansions and proving error bounds for truncated asymptotic series of the solution, notably by Buslaev [23], Morawetz and Ludwig [78], and Melrose and Taylor [75], among others. Often, *the oscillatory component of the solution can be determined explicitly* from these asymptotic expansions. This can then be used in designing efficient hybrid methods. Furthermore, from the asymptotic expansions, frequency-dependent bounds on the slowly-varying remainder and its derivatives can be obtained (in some cases these follow directly from classical results, in other cases some additional work is required). The frequency-dependent bounds are the key results used in the frequency-explicit numerical error analysis of the approximation of the slowly-varying remainder. This thesis presents a rigorous justification of one of the key result using only elementary techniques.

Hybrid numerical-asymptotic methods have been shown in theory to be substantially more efficient than classical numerical methods alone. For example, [40] presented a hybrid numerical-asymptotic method in the context of boundary integral equations (BIEs) for solving the problem of high-frequency scattering by smooth, convex obstacles in two dimensions. It was proved in [40] that in order to maintain the accuracy as the frequency increases, the hybrid BIE method requires the number of degrees of freedom to grow slightly faster than $k^{1/9}$, where k is a parameter proportional to the frequency. This is a substantial improvement from the classical boundary integral methods that require $O(k)$ number of degrees of freedom to achieve the same accuracy for this problem.

Despite this slow growth in the number of degrees of freedom, hybrid numerical-asymptotic methods lead to stiffness matrices with entries that are highly-oscillatory singular integrals that can not be computed exactly. Thus, without efficient and accurate numerical treatment of these integrals, the hybrid numerical-asymptotic methods, regardless of their attractive theoretical accuracy, can not be efficiently implemented in practice.

In order to resolve this difficulty, this thesis develops a methodology for approximating the integrals arising from hybrid methods in the context of BIEs. The integrals are transformed under a change of variables into integrals amenable to Filon-type quadratures. Filon-type quadratures are designed to cope well with high oscillations in the integrands. Then, graded meshes are used to capture the singularities accurately.

Along with k-explicit error bounds for the integration methods, this thesis derives k-explicit error bounds for the hybrid BIE methods that incorporate the error of the inexact approximation of the entries of the stiffness matrix. The error bounds suggest that, with an appropriate choice of parameters of Filon quadrature and mesh grading, the overall error of the hybrid method does not deteriorate due to inexact approximation of the stiffness matrix, therefore preserving its attractive theoretical convergence properties.

Acknowledgements

It is a great pleasure to thank many people who have helped me get to this point.

First and foremost and the biggest thanks go to my supervisors Ivan Graham and Valery Smyshlyaev. Ivan's pursuit for clarity and perfection and Valery's ability to explain complex ideas with simplicity taught me a great deal about what it takes to be a good mathematician. Without their assistance, patience and encouragement this thesis would not have been possible. It was a tremendous privilege to work under their guidance and I am deeply grateful for this opportunity.

This thesis have also greatly benefited from many discussions with Victor Dominguez who has closely followed my progress over the last four years. Victor's many great ideas have helped me to construct some nice results in several chapters of this thesis.

Furthermore, I would also like to thank Vasiliy Babich for a number of fruitful comments on my work and for inviting me to the Days of Diffraction conference in St Petersburg, it was a great experience; and Alastair Spence for many helpful suggestions on how to improve the exposition of my work.

I thank Bath/Reading group for many interesting discussions and virtual talks on the topic of this thesis and the Numerical analysis group at Bath for creating a fantastic atmosphere for work.

I thank my examiners Arieh Iserles and Ilia Kamotski for their careful reading of this thesis and useful comments and suggestions on how to improve it.

I thank the University of Bath for the financial support for this project and travels to many conferences during my PhD. Additionally I was fortunate enough to be awarded the Ede and Ravenscroft prize during my time as a PhD student and I thank the prize committee for this honour.

I have been incredibly lucky to have many special friends around me in the last four years. Fynn Scheben, Jane Temple and Euan Spence, in your own ways, you have helped me a great deal to celebrate my thesis today; Andrea, Jane, Marion, and Melina, thank you for all the laughs and giggles, Sundays would not have been so much fun without you; and

all my friends in Bath and beyond who have made life during my PhD so enjoyable and unforgettable, Thank you very much!

Finally, I would like to thank my parents, Alexander and Rosa Kim, who have encouraged and supported me in all my endeavours throughout my life. This thesis is dedicated to you.

Contents

List of Figures **iv**

List of Tables **viii**

1 Introduction **1**

1.1 High-frequency acoustic scattering problem 1

1.2 The main aims of the thesis 9

1.3 The main achievements of the thesis 10

1.4 Outline of the thesis 11

2 Mathematical framework **13**

2.1 Introduction 13

2.2 Boundary integral representation 15

2.2.1 Standard combined-potential boundary integral equation 15

2.2.2 Star-combined potential boundary integral equation 17

2.3 Galerkin method and its error analysis 19

2.3.1 Semi-discrete Galerkin method 23

2.3.2 Fully-discrete Galerkin method 26

3 Asymptotic methods for high-frequency acoustic scattering problems **29**

3.1 Introduction 29

3.1.1 Main results of the asymptotic theory for high-frequency scattering problems 30

3.1.2 Outline of the chapter 33

3.2 Preliminaries 34

3.2.1 Bessel functions 34
3.2.2 Basic asymptotic expansions of Bessel functions at large values of the order or the argument 35
3.2.3 Airy functions and their basic asymptotic expansions 37
3.2.4 Debye's, Olver's and Cherry's asymptotic expansions of Hankel function 41
3.2.5 Zeroes of Airy and Hankel functions 45
3.3 Model problem: scattering by a circle 48
3.3.1 Evaluating the normal derivative of the solution 51
3.3.2 Asymptotic behaviour of the solution on the illuminated part of the boundary 56
3.3.3 Asymptotic behaviour of the solution on the transitional parts of the boundary 63
3.4 Model Problem Method 70
3.4.1 Asymptotic representation of the solution in the transition regions for general convex domains 72
3.4.2 Determining the eikonal analogue 74
3.4.3 The normal derivative of the solution. 78

4 Computation of highly-oscillatory double integrals 81
4.1 Introduction 81
4.1.1 Motivation of the chapter 83
4.1.2 The structure of the double integrals 85
4.1.3 Outline of the remainder of the chapter 93
4.2 Methodology for the computation of highly-oscillatory integrals 94
4.2.1 Abstract methodology for single integrals 94
4.2.2 The idea for the computation of the double integrals 101
4.3 Application to the integrals arising from scattering problem 107
4.4 Stationary points for the phase function 115
4.4.1 Integration over the transition domains 121

5 Filon-Clenshaw-Curtis quadrature **124**

5.1 Introduction 124

5.1.1 Outline of the chapter 125

5.1.2 Survey of existing methods 126

5.1.3 Motivation for the chapter 135

5.2 Filon-Clenshaw-Curtis quadrature 138

5.2.1 Chebyshev interpolating polynomial 138

5.2.2 Trigonometric interpolating polynomials 140

5.2.3 The moments of the Filon-Clenshaw-Curtis quadrature 142

5.3 Error of the Filon-Clenshaw-Curtis quadrature 146

5.3.1 Error estimate in terms of the periodic Sobolev norm of f_c 146

5.3.2 Error estimate in terms of Chebyshev weighted norm of $f^{(m)}$ 150

5.3.3 Extension to the integrals over $[a, b]$ 156

5.4 The application of the Filon-Clenshaw-Curtis rule to integrals with algebraic singularities 159

5.4.1 Error estimate of the composite Filon-Clenshaw-Curtis quadrature . 166

5.5 Numerical examples 168

6 Error of the fully-discrete Galerkin method **176**

6.1 Introduction 176

6.2 Error of the semi-discrete Galerkin approximation 177

6.3 Inexact approximation of the entries of the stiffness matrix 181

6.3.1 Standard formulation 181

6.3.2 Star-combined formulation 183

6.3.3 The error in the inexact approximation of the entries 187

6.4 Strang's Lemma and the error of the fully-discrete Galerkin method 189

6.4.1 Properties of the Chebyshev-weighted norm 189

6.4.2 Error of the fully-discrete Galerkin method 192

6.5 Numerical examples 202

Bibliography **211**

Appendices **221**

A Eikonal equation **221**

A.1 The Eikonal equation 222

A.2 Reflected rays 223

A.3 Amplitude of the wave fronts 223

B Method of characteristics **225**

C Locations of stationary points **227**

D Chebyshev-weighted L_2-norms. **235**

E Behaviour of the integrand in the transformed integrals. **238**

E.1 Log-singularity 239

E.2 Smoothness of the integrand function away from singularities 241

F Proof of Lemma 4.8 **246**

G Proof of Proposition 3.22 **248**

List of Figures

1.1 Example illustration of scattering in 3D. The plot illustrates the the incident wavefield propagating in the direction of the vector $\boldsymbol{a}$. The boundary of the three dimensional scatterer Ω is denoted by Γ. 3

2.1 Parameterised boundary is split into four domains, where the solution has different asymptotic expansions as $k \to \infty$. 24

3.1 Contour of integration c. 32

3.2 The plot of $\cos 3\theta$ is illustrated on the left with the shaded regions indicating where the values of $\cos 3\theta$ are negative. The function $s(x)$ decays exponentially in the shaded regions illustrated on the right. In particular, $s(x)$ decays exponentially as $|x|$ tends to infinity along a contour plotted in red. 39

3.3 The plot illustrates the domains where Hankel function is represented by series S_1 and S_2 defiend in (3.33). We denote the circular neighborhoods of points $\xi = \pm 1$, as $\mathcal{B}_{\pm 1}$, where the Debye's representation of Hankel function is not valid. 42

3.4 Zeroes of Airy function, Ai, denoted as a_s, are plotted as blue starts located strictly on the negative real axis. The ν-zeroes (3.47) of the Hankel function $H_\nu^{(1)}(z)$ are plotted as red stars. 46

3.5 The contour of integration $\mathcal{L}$ in (3.52). In this figure, zeroes of the Hankel function, $H_{k\xi}^{(1)}(ka)$, are illustrated (as red stars) in relation to the contour of integration $\mathcal{L}$. In green stars, the zeroes of the sine function are plotted in relation to $\mathcal{L}$. The contour $\mathcal{L}$ does not contain inside its loop any zeroes of Hankel function provided (3.53) is satisfied. 50

3.6 The domain D in the complex ξ-plane represents the domain where the Debye's asymptotic representation of Hankel function is valid [16] and of the form (3.73). Domain D excludes the point $\xi = a$ and the circular domain around the point of a fixed radius denoted as $\mathcal{B}_1$. The path of integration, $\xi \in [i\beta, i\beta + \infty)$, is also illustrated 57

3.7 Deformed contour of integration $\mathcal{L}'$ 59

3.8 The subcontour of the deformed contour of integration $\mathcal{L}'$ containing the stationary point ξ_0 60

3.9 The deformed contour of integration. 64

3.10 Within an accuracy of terms which are exponentially small as $k \to \infty$, we can replace the contour of integration $\mathcal{L}_1$ with $\mathcal{L}_2$. 65

3.11 Integration path in the new variable z. Due to the definition of z in (3.95), the contour of integration in z extends to infinity with exponentially small error. The contour $\mathcal{L}_3$ is of the same form as a contour c in Figure 3.1. . . . 66

3.12 The exterior to the general convex scatterer domain is separated into illuminated (I), transitional (II and III) and shadow (IV) zones with respect to the incident wave. 70

3.13 The illuminated (I), transitional (II) zones overlap in the domain $\mathcal{P}$. 71

3.14 The contour of integration Λ_3 in the complex plane. $|\gamma_1| = |\gamma_2| = \gamma_0$, where γ_0 is a small positive number. 72

3.15 The eikonal $\tau(\boldsymbol{x})$ equals to $s + t$ 75

4.1 Original (left) and transformed (right) domains of integration under transformation (4.79). Note the upper and lower curves bounding $\tilde{D}$ are monotone due to assumption (4.82). 103

4.2 Transformed domain $\tilde{D}$ is subdivided into three subdomains. 103

4.3 Original and transformed domains of integration. Note the upper curve bounding $\tilde{D}$ contains a turning point and the lower boundary is $\tau = 0$. . . . 105

4.4 Transformed domain $\tilde{T}$ is subdivided into two subdomains. 106

4.5 The original rectangular domains of integration which include the diagonal $s = t$ (where M is singular). The rectangular domains containing the diagonal are separated into the upper triangular and lower triangular subdomains. 108

4.6 On the left, the transformed domains of integration, corresponding to the above-the-diagonal part of the original domains are illustrated; on the right, the transformed domains corresponding to the below-the-diagonal part of the original domains are illustrated. The transformation is governed by (4.79). We consider the integration over the triangular domain AFG in more detail later, see Figure 4.16. 108

4.7 Plot of function $F_1(\tau)$ and $F_2(\tau)$. The function $F_1(\tau)$ has a logarithmic singularity at $\tau = 0$ and the function $F_2(\tau)$ has the square-root singularity at $\tau = \tau_{\max}$. 110

4.8 The original domain of integration, $(\Lambda_2 \times \Lambda_3)$ and the transformed domain of integration. 113

4.9 The transformed domain is subdivided by the dashed lines into five subdomains. 114

4.10 Plot of functions F_j, $j = 1, \ldots, 5$. The function has the square-root singularity at $\tau = \tau_3$ and $\tau = \tau_{\max}$, and a log-singularity at $\tau = \tau_{\min} = 0$. 114

4.11 The curves plotted in blue represent points $(s, t) \in ([0, 2\pi] \times [0, 2\pi])$ where Hypothesis A is not satisfied for the case of an elliptical scatterer $(a = 3, b = 1)$. The points where the integrand is singular are located on the diagonal $s = t$ which is plotted in red. 115

4.12 The boundaries of the illuminated $\Lambda_2 := [b, c]$ and transition domains $\Lambda_1 := [a, b]$ and $\Lambda_3 := [c, d]$ are plotted along with the diagonal $s = t$ in red and blue curves representing points where Hypothesis A is not satisfied within the domain contained in the rectangles. 115

4.13 The boundaries of the triangular domains A_1 and A_2 are plotted in green. . 117

4.14 The subdomain A_1 represents the domain where the integrand is slowly-oscillatory and is defined in (4.100). 117

4.15 The behaviour of the integrand in the domain $[t_1 - \delta, t_1 + \delta] \times [t_1 - \delta, t_1 + \delta]$. Above the diagonal, the integrand of (4.11) is slowly-varying, while below the diagonal it is highly-oscillatory. 117

4.16 On the left, is the plot of the original domain $(\Lambda_1 \times \Lambda_1)^+$ separated into two subdomains by a line PQ: the triangle APQ and the remaining trapezium FGPQ; the right plot corresponds to the transformed domain of the subdomain $FGPQ$. 121

5.1 The deformed contour of integration in the complex plane. 133

6.1 In the Example 1, we consider scattering by an ellipse with major axis greater than the minor axis. The incident wavefield, u^I, is propagating in the direction of the vector $(1,0)^T$. 202

6.2 The plot of the computed approximation of the function $V(\cdot, 600)$. The real part of the solution is plotted in blue and the imaginary part is plotted in red. 204

6.3 The plot of the computed approximation of the function $V(\cdot, 6000)$. The real part of the solution is plotted in blue and the imaginary part is plotted in red. 204

6.4 The function u^I denotes the incident wavefield propagating in the direction of the vector $(1,0)^T$ and scattering off the ellipse with the major axis smaller that the minor axis. 205

6.5 The plot of the computed approximation of the function $V(\cdot, 600)$. The real part of the solution is plotted in blue and the imaginary part is plotted in red. 207

6.6 The plot of the computed approximation of the function $V(\cdot, 6000)$. The real part of the solution is plotted in blue and the imaginary part is plotted in red. 207

6.7 The plot of absolute values of the entries of the stiffness matrix for the case when $k = 600$ corresponding to the standard combined formulation. 210

6.8 The plot of absolute values of the entries of the stiffness matrix for the case when $k = 600$ corresponding to the star-combined formulation. 210

6.9 The plot of absolute values of the entries of the stiffness matrix for the case when $k = 6000$ corresponding to the standard combined formulation. 210

6.10 The plot of absolute values of the entries of the stiffness matrix for the case when $k = 6000$ corresponding to the stan-combined formulation. 210

A.1 Directions of the incident and reflected waves. 224

C.1 Stationary s and t are opposite. 228

C.2 $\alpha(s,t)$ as the angle between $\boldsymbol{T}(t)$ and $\boldsymbol{\rho}(s,t)$, measured anti-clockwise from $\boldsymbol{T}(t)$. 230

C.3 $\beta(t)$ as the angle between $\boldsymbol{T}(t)$ and $\boldsymbol{a}$, measured anti-clockwise from $\boldsymbol{T}(t)$. . 230

C.4 The point t_L that represent the limit of how far the transition zone Λ_1 can extend into the shadow is illustrated. There are similar restrictions for the domain Λ_3 that can be derived similarly to Λ_1. 233

List of Tables

4.1 Geometries of the original and transformed under (4.79) domains respectively. 111

4.2 Description of the transformed domains. In the first column we display the original domain of integration in J_k that is either a rectangular or a triangular domain. See Table 4.1 for plots of these domains. The remaining four columns of the table describe the corresponding transformed domain and contain a sufficient information to enable us to write the double integral J_k in the form (4.93). The intervals $[\tau_j, \tau_{j+1}]$, $j = 0, \ldots, J-1$, in (4.93) are displayed in the third column of the table with τ_j defined in the second column. The turning points ξ_b, ξ_c and ξ_d are defined in (4.95). The *lower* and the *upper boundaries* of functions F_j defined in (4.94) are displayed in the fourth and fifth columns respectively. For each $j = 1, \ldots, J$, we emphasize what type of singularity F_j has at τ by adding **(L)** or **(S)** next to τ in the second column, where **(L)** represents a logarithmic singularity and **(S)** represents a square-root singularity. The function F_j is smooth away from these singularities. The entries of this table were obtained by investigating the geometries of the transformed domains and can be easily verified by looking at the geometry of the transformed domains in Table 4.1. 112

4.3 Table of transformed domains 123

5.1 The table illustrates the errors $E_k(N)$ defined in (5.131) and the ratios of errors, $r(k)$, defined in (5.132). 168

5.2 The table illustrates the errors $E_k(M, q)$ defined in (5.135) and the ratios of errors, $r(k)$, defined in (5.136). 170

5.3 The table illustrates the errors $E_N(M, q)$ defined in (5.137). 171

5.4 The table displays the errors $E_k(M, q, N)$ defined in (5.138) for fixed N and q and varying M and k. 172

5.5 The table displays the ratios $r_k(M, q, N)$ of errors defined in (5.139) for $k = 1000$ to determine the convergence rate of the composite FCC with respect to M. 172

5.6 The table displays the errors $E_k(M, q, N)$ and their ratios $q_k(M, q, N)$ that are defined in (5.140) to determine the convergence rate with respect to k. . 173

5.7 The absolute values of the integrals in (5.141) 174

5.8 The table displays the values of the parameter M required for the composite FCC to achieve an absolute accuracy of $TOL = 1.0e - 012$ for varying k as well as the error of the FCC $E(k, M, q, N, f)$, defined in (5.122). The parameters N and q are fixed at $N = 16$, $q = 12$. 174

5.9 The table of values M required for the composite FCC to achieve an accuracy of TOL with fixed $N = 16$, $q = 12$ and $\beta = -1/4$. In the bottom line of the table, the ratios defined in (5.142), averaged over k, are displayed. . . 175

5.10 The table of values N required for the composite FCC to achieve an accuracy of TOL with fixed $M = 10$, $q = 12$, $\beta = -1/4$. 175

6.1 The errors e_d defined in (6.91) of the fully-discrete Galerkin method for solving the standard and star-combined integral equation for low/moderate k. 203

6.2 The rate of growth in cpu time with growing d. 203

6.3 The ratios of the errors of the fully-discrete Galerkin method for solving the standard and star-combined integral equations. 204

6.4 The errors e_d defined in (6.92) of the fully-discrete Galerkin method for solving the standard and star-combined integral equations for large k. . . . 205

6.5 The errors in the approximation of $V(\cdot, k)$ for two different type of ellipses: on the left, the results are displayed for ellipse with curvature $\kappa(t) = 9$ at the transition points (right columns) and the results on the right correspond to the ellipse with curvature $\kappa(t) = 1/3$. 206

6.6 The errors e_d defined in (6.91) of the fully-discrete Galerkin method for solving the standard and star-combined integral equation for low/moderate k. 207

6.7 The ratios of the errors of the fully-discrete Galerkin method for solving the standard and star-combined integral equations. 208

6.8 Condition numbers for star-combined integral equation. 208

6.9 Ratios $\text{cond}_d(600)/\text{cond}_d(6000)$, between the condition numbers for $k = 600$ and $k = 6000$. 209

Chapter 1

Introduction

1.1 High-frequency acoustic scattering problem

Motivation

High frequency acoustic waves are utilised in a huge number of applications that we benefit from in our daily lives. For example, medical ultrasound detects changes in biological tissue by sending an acoustic wave into the tissue and determining its properties from the scattered waves. More recently, medical high-intensity focused ultrasound (HIFU) has been the subject of intense research and development. The idea behind HIFU is to focus ultrasonic acoustic waves at pathogenic tissue to rapidly heat and destroy it. This is a highly promising technology that can potentially deposit energy in biological tissue via high-frequency acoustic waves without damaging the tissue through which the ultrasound propagates prior to its focal point. In addition to medical applications, ultrasound can be used in industrial applications to detect cracks and flaws in construction material - an application known as ultrasonic testing. Furthermore, seismic inversion uses scattered acoustic wavefield data to determine the properties of subsurface rock structure. Oil companies use seismic inversion for exploration of oil and gas fields.

Furthermore, the equations that govern acoustic wave propagation have similar properties to those describing electro-magnetic waves and thus a better understanding of high-frequency acoustic wave propagation is often the first step in a better understanding of electro-magnetic wave propagation. Electro-magnetic waves are used in diverse industrial applications such as radar and telecommunications.

In these applications, the characteristic length scale of the scatterer is much bigger than the wavelength λ of the ultrasound. Hence, it is of great importance that we understand and are able to predict the behaviour of the acoustic wavefield, particularly for the case when acoustic waves have short wavelengths in comparison with the size of the scatterer.

Forward model problem

The standard equation that governs the acoustic wave propagation can be written as

follows,

$$\nabla^2\phi = \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2},$$

where c is a constant that represents the speed at which the wave propagates and ∇^2 is the Laplacian. The equation is a second order linear partial differential equation known as the *wave equation*. If the time dependence of $\phi(\boldsymbol{x}, t)$ is known to be of the form

$$\phi(\boldsymbol{x}, t) := u(\boldsymbol{x})\exp(-i\omega t), \tag{1.1}$$

where $\omega = 2\pi/\lambda$, and λ is the wavelength, then such an acoustic wavefield is called *time-harmonic*. For time-harmonic wavefield, the wave equation reduces to the Helmholtz equation

$$\nabla^2 u + k^2 u = 0,$$

where $k = \omega/c$ is a parameter called the *wavenumber*.

There are two types of scattering problems: the forward problem and the inverse problem. The forward problem is concerned with determining scattered wavefield data given the properties of the incident wavefield and the scatterer.

On the other hand, the inverse problem is concerned with determining the properties of the scatterer given the incident and scattered wavefields are known. An example of the inverse problem can be found in sonar devices: a known wave is emitted and the scattered wavefield data is examined to determine the nature of the object upon which it scattered. In this case, the higher the frequency used, the more details of the geometry can be recovered.

The forward model problem is typically formulated as a boundary value problem. In the case of the Helmholtz equation on an unbounded domain, this is to be solved subject to boundary conditions and a radiation condition.

Let us consider the underlying "simpler" mathematical problem (that we refer to as a model problem) that underpins the modelling, or simulation, of the wave propagation phenomena in applications described earlier.

In the following definition we formulate the model *forward* problem in the unbounded domain.

Definition 1.1. *We define the exterior model scattering problem as follows. We seek the total wavefield u that satisfies the Helmholtz equation in the exterior domain to the scatterer, Ω, with Lipschitz boundary Γ,*

$$\Delta u + k^2 u = 0 \quad \text{in} \quad \mathbb{R}^d \setminus \overline{\Omega}, \tag{1.2}$$

and the Dirichlet boundary condition,

$$u = 0, \qquad on \quad \Gamma, \tag{1.3}$$

with the scattered wavefield, u^S*, satisfying Sommerfeld radiation condition:*

$$\frac{\partial u^S}{\partial r} - iku^S = \mathbf{o}\left(r^{\frac{1-d}{2}}\right), \tag{1.4}$$

as $|r| \to \infty$ *uniformly in* $\hat{x} = x/r$*, where* $u^s = u - u^I$*, where* u^I *represents incident wavefield and* d *denotes the dimension.*

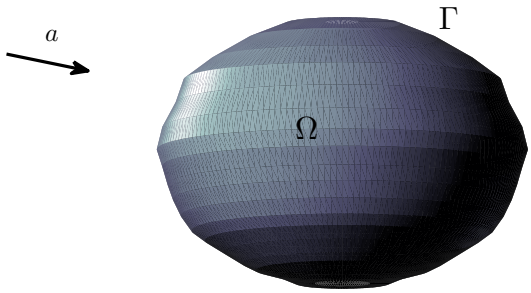


Figure 1.1: *Example illustration of scattering in 3D. The plot illustrates the the incident wavefield propagating in the direction of the vector* $\boldsymbol{a}$*. The boundary of the three dimensional scatterer* Ω *is denoted by* Γ*.*

The exterior scattering problem is known to have a unique solution provided that u and ∇u are locally square integrable [34].

The wavenumber parameter k is typically very large in the applications described earlier. The analytic solution of the Helmholtz equation is possible only for simple obstacles such as a circle. For more general obstacles, asymptotic and numerical methods for finding approximate solutions need to be applied.

The purpose of the thesis

Ultimately, this thesis is devoted to the development, analysis, and implementation of the methods for solving the model exterior scattering problem with controllable accuracy for ***any*** wavenumber k.

Conventional numerical methods, based on polynomial interpolation, have been successfully applied to the model exterior scattering problem, and subsequently to the simulation

of the acoustic wave propagation in applications. However, the applicability of these methods is limited to moderately low wavenumbers k. We explain this fact in more detail shortly.

At much higher frequencies, asymptotic methods can be used to find an approximate solution to the forward problem. However, asymptotic methods are not error-controllable for all k.

Asymptotic methods

Asymptotic methods are typically used to construct the solution to the scattering problem as a series, called an *asymptotic expansion*, with respect to the small parameter $1/k$. For example, well known ray methods are used successfully in many engineering disciplines.

The solution to the Helmholtz equation (1.2) is sought in the form of the ray expansion as follows:

$$u(\boldsymbol{x}) = \left(\sum_{j=0}^{\infty} \frac{A_j(\boldsymbol{x})}{(ik)^j} \right) \exp\left(ik\tau(\boldsymbol{x})\right) . \tag{1.5}$$

The unknown functions in (1.5), are the *Eikonal* $\tau(\boldsymbol{x})$ and the *asymptotes* $A_0(\boldsymbol{x})$, $A_1(\boldsymbol{x})$,..., etc. The equations for $\tau(\boldsymbol{x})$ and $A_j(\boldsymbol{x})$, are obtained by formally substituting the Ray expansion (1.5) into the Helmholtz equation (1.2) and equating terms of order k^2, k^1, k^0, k^{-1} and so on. Namely,

(i) the Eikonal equation (by equating the terms of order k^2):

$$|\nabla\tau(\boldsymbol{x})|^2 = 1, \tag{1.6}$$

for $\boldsymbol{x} \in \mathbb{R}^2 \setminus \Omega$, and

(ii) the Transport equation (by equating the terms of order k^{1-n}):

$$\triangle\tau(\boldsymbol{x})A_n(\boldsymbol{x}) + 2\left(\nabla\tau(\boldsymbol{x}) \cdot \nabla A_n(\boldsymbol{x})\right) = -\triangle A_{n-1}(\boldsymbol{x}), \quad n = 0, 1, 2, ..., \tag{1.7}$$

where, by convention, $A_{-1} = 0$.

Geometrical optics and the geometrical theory of diffraction [65] provide a general set of instructions for the construction of the asymptotic expansion of the solution to a scattering problem.

For the case of high-frequency scattering by a convex obstacle, asymptotic theory suggests that the normal derivative of the solution to the exterior scattering problem can be written as the product of an explicitly known oscillatory function and a relatively slowly-varying function, $V(\boldsymbol{x}, k)$,

$$\frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{x}) = kV(\boldsymbol{x}, k)\exp(ik\boldsymbol{x} \cdot \boldsymbol{a}), \qquad \boldsymbol{x} \in \Gamma, \tag{1.8}$$

where $\boldsymbol{a}$ is a unit vector representing the direction of incidence and the unit vector $\boldsymbol{n}$ is the outward unit normal to the boundary.

In the case of non-convex objects, the function $\partial u/\partial \boldsymbol{n}$ in (1.8) has more complicated asymptotic behaviour, e.g. diffraction from edges or corner points [15, 6], or multiple scattering (non-convex domains) [18, 42, 43]. As we shall see later, results on asymptotic behaviour of the solution are of fundamental importance to the development and analysis of efficient numerical techniques for solving high-frequency scattering problems.

Conventional numerical methods

Both finite element and boundary element methods can be applied to scattering problems. While finite element methods are typically used for solving scattering problems in inhomogeneous bounded media, boundary element methods are a popular choice for solving scattering problems in unbounded domains in predominantly homogeneous media. This is because the exterior scattering problem posed on a homogeneous unbounded domain can be reformulated as a problem on the surface of the scatterer, hence reducing the dimension of the problem posed. Typical features of finite element and boundary element methods in the context of exterior scattering problems are outlined in the table below.

Finite element methods (FEMs)	Boundary element methods (BEMs)
are suitable for propagation in homogeneous and inhomogeneous media	are suitable for predominantly homogeneous media
require the mesh to cover the whole scattering domain	the only require the mesh to cover the boundary of the scatterer
lead to large sparse matrices	lead to smaller dense matrices
applied to problems on unbounded domains, require the radiation condition to be approximated by an artificial boundary condition (e.g. to truncate the domain)	are more suitable for scattering problems on unbounded domains

Conventional finite element and boundary element methods are based on (piecewise) polynomial approximation of the solution and cope very badly with high oscillations. Indeed, it is commonly recommended that with conventional methods, one should carry out a

fixed number of function evaluations per wavelength in order to capture the oscillatory behaviour of the solution to the scattering problem accurately. While classical methods can be effective for small and moderate wavenumbers, as frequency increases, to maintain the accuracy, the number of function evaluations grows with k at least as k^{d-1}, where d denotes the dimension of the scattering domain. Therefore, as $k \to \infty$ conventional numerical methods become essentially numerically intractable.

The finite element approach has even further increase in computational costs due to the "pollution error" contribution [9, 7, 8]. For the case of 1D model problem,

$$\frac{d^2u}{dx^2}(x) + k^2u(x) = -f(x), \quad x \in (0,1) \quad u(0) = 0, \quad \frac{du}{dx}(1) + iku(1) = 0,$$

assuming only $kh < 1$ (where h is the size of the biggest finite element), the relative finite element error in the H^1-norm, satisfies the estimate:

$$e_1 \leq C_1 kh + C_2 k^3 h^2, \tag{1.9}$$

where C_1 and C_2 are constants independent of k and h. The second term is often referred to as a "pollution error" term. The error bound (1.9) indicates that for growing k, even if hk is controlled, the error still grows linearly with k.

Novel FEMs

Several recent developments have enabled finite element methods to be applied to high-frequency scattering problems. The *partition of unity* method proposed by Babuška and Melenk [10] is a general technique in which the approximation space is enriched with functions that represent the oscillatory solution well. For example, sets of plane wave functions and Bessel functions provide complete sets for the approximation of an oscillatory wave-field. Plane waves have been successfully implemented in a finite element (PUFEM), approach e.g. in [85]. In the PUFEM, the finite element space is constructed by multiplying shape functions, which form a partition of unity, with functions that have good local approximation properties. However, the number of degrees of freedom is still required to grow linearly with k as $k \to \infty$ to maintain the accuracy.

Further classes of finite element methods, based on the inclusion of the wave nature of the solution into the formulation, are the *ultra weak variational formulation* (UWVF) [24, 77, 22, 52] and the *Galerkin least squares finite element* methods [77, 98]. The UWVF methods have been shown to be equivalent to a Galerkin least squares method that uses the Trefftz variational formulation, see [49]. Trefftz-type methods are based on approximation spaces made of functions which locally solve the Helmholtz equation. Thus PUFEM and UWVF methods are both Trefftz-type methods. Recently, error analysis of Trefftz-type Galerkin methods for the time-harmonic Maxwell equations has been carried out in [57]

and in particular for UWVF methods in [22].

Novel BEMs

The partition of unity method has also been successfully applied in a boundary element approach (PUBEM). The numerical experiments [58, 87] suggest that such an approach significantly reduces the computational costs of the boundary element method.

Recently, a new class of methods called hybrid asymptotic-numerical boundary element methods has attracted considerable interest. These methods, related to the PUBEM, incorporate the wave nature of the solution into the approximation space [1, 2, 20, 51, 28, 87, 27, 40].

BEMs allow the scattering problem (1.2)-(1.4) to be reformulated for the normal derivative of the solution $v(\boldsymbol{x}) = \partial u(\boldsymbol{x})/\partial \boldsymbol{n}(\boldsymbol{x}) \in L^2(\Gamma)$ as an integral equation on the boundary Γ using Green's integral representation theorem [34, Theorem 2.1]. For example, the function u is the solution of the exterior scattering problem (1.2)-(1.4) if and only if its normal derivative v satisfies the boundary equation (see Theorem 2.2), for $\boldsymbol{x} \in \Gamma$, for a fixed $\eta \in \mathbb{R} \setminus \{0\}$,

$$\frac{1}{2}\frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{x}) + \int_\Gamma \left(\frac{\partial \Phi(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{n}(\boldsymbol{x})} - i\eta\Phi(x,y) \right) \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{y}) dS(\boldsymbol{y}) = \frac{\partial u^I}{\partial \boldsymbol{n}}(\boldsymbol{x}) - i\eta u^I(\boldsymbol{x}), \tag{1.10}$$

where $\Phi(\boldsymbol{x},\boldsymbol{y})$ is the fundamental solution of Helmholtz equation defined later in Theorem 2.1. For the case when the incident wave is planar and the scatterer is convex, the solution is known to be of the form (1.8), i.e. the product of an explicitly known oscillatory term and a relatively slowly-varying function V. In [1, 2] the ansatz (1.8) is utilised so that only the slowly-varying V is approximated. The authors of [1, 2] suggested that in order to maintain the accuracy of the method, the number of degrees of freedom must grow only as $O(k^{1/3})$ as $k \to \infty$.

In [20], the boundary integral formulation (1.10) is modified using the ansatz (1.8) so that the solution of the modified equation is the relatively slowly-varying V. In detail, multiplying the boundary integral equation (1.10) by $\exp(-ik\boldsymbol{x}\cdot\boldsymbol{a})$, we obtain an integral equation for $V(\boldsymbol{x},k)$, $\boldsymbol{x} \in \Gamma$:

$$\frac{1}{2}V(\boldsymbol{x},k) + \int_\Gamma \left(\frac{\partial \Phi(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{n}(\boldsymbol{x})} - i\eta\Phi(x,y) \right) V(\boldsymbol{y},k) \exp(ik(\boldsymbol{y}-\boldsymbol{x})\cdot\boldsymbol{a}) dS(\boldsymbol{y}) = i(\boldsymbol{n}\cdot\boldsymbol{a}k - \eta). \tag{1.11}$$

The function V is slowly-varying away from the shadow boundaries (transition points). However, even when V is slowly-varying, the evaluation of the integral in (1.11) requires the number of quadrature points to grow proportionally with k to maintain the accuracy.

In [20], a novel method called the *localisation principle* tackles the problem of evaluating

the integral using an error-controllable extension of the method of stationary phase. The idea of the localisation method is that the highly-oscillatory integral needs only be approximated around the critical points of its integrand where the only significant contribution to the overall integral is made. The intuitive justification for this phenomenon is that away from critical points the oscillations in the integrand cancel each other out.

The critical points in the integral in (1.11) for a fixed $\boldsymbol{x}$ are

- the kernel singularity, when $\boldsymbol{y} = \boldsymbol{x}$;
- the stationary points of the phase,

$$\phi(\boldsymbol{y}) = |\boldsymbol{y} - \boldsymbol{x}| + (\boldsymbol{y} - \boldsymbol{x}) \cdot \boldsymbol{a}.$$

Expanding the phase function in the Taylor series, one can find that the oscillatory part of the integrand in (1.11) is of the form $\exp(ikx^p)$ with:

- $p = 1$ around the kernel singularity,
- $p = 2$ around the stationary points other than the shadow boundaries,
- $p = 3$ around the shadow boundary stationary points, provided the curvature does not vanish.

The concept of the localised integration in [20] can be understood by considering an example. Integrate

$$I_k^{[-A,A]}[f_A] := \int_{-A}^{A} f_A(x) \exp(ikx^p) dx,$$

where $f_A(x) = S(x, cA, A)(1 - S(x, -A, -cA))$ for $0 < c < 1$ and the smooth cut-off function $S(x, x_0, x_1)$ is defined as:

$$S(x, x_0, x_1) = \begin{cases} 1, & \text{for } \ x \le x_0, \\ \exp\left(\frac{2e^{-1/u}}{u-1}\right), & \text{for } \ x_0 < x < x_1, \ u = \frac{|x|-x_0}{x_1-x_0}, \\ 0, & \text{for } \ x \ge x_1. \end{cases} \tag{1.12}$$

Then, for $0 < \varepsilon < A$ and for $f_\varepsilon(x) = f_A(Ax/\varepsilon)$, see [20, Lemma 3.1],

$$\int_{-A}^{A} f_A(x) \exp(ikx^p) dx = \int_{-\varepsilon}^{\varepsilon} f_\varepsilon(x) \exp(ikx^p) dx + \mathcal{O}\left((k\varepsilon^p)^{-n}\right), \quad n \ge 1. \tag{1.13}$$

Thus, under certain conditions on $k\varepsilon^p$, the integral $I_k^{[-\varepsilon,\varepsilon]}[f_\varepsilon]$ (that can be efficiently computed using classical quadrature methods) is a good approximation to the integral $I_k^{[-A,A]}[f_A]$.

Integration around the shadow boundaries requires special consideration. The function V is not slowly-oscillatory around the shadow boundary. In fact, the n-th derivative of V grows with order $O(k^{n/3})$ as $k \to \infty$. In order to overcome this, a change of variable is employed before the localised integration is applied. Wavenumber-independent convergence was observed experimentally in [20]. The method has been subsequently extended to 3D in [19].

A similar ansatz and ideas have been considered for the problem of scattering by non-convex obstacles in [18]. The theory for this method has been subsequently advanced in [43, 42] for the case of multiple scatterers. However, the frequency-explicit error analysis of the hybrid methods has not been developed in these references.

To summarize, hybrid asymptotic-numerical methods are methods for solving the boundary integral equation (1.10) using Galerkin method with the finite dimensional space that incorporates the ansatz such as (1.8) for the case of planar wave scattering by smooth convex objects. Wavenumber-explicit error estimates of the hybrid asymptotic-numerical methods have been obtained in [40] for the case of a convex obstacle with smooth boundary Γ, and in [27] by a convex polygon.

1.2 The main aims of the thesis

The main aims of this thesis are:

A. The rigorous justification of the asymptotic result (1.8) is of considerable interest to us since it plays a key role in the development and analysis of efficient numerical techniques for solving scattering problems. The often cited paper by Melrose and Taylor [75] provides a very technical proof of this result and will not be reproduced here. It is of benefit, however, to provide a rigorous justification of this result that does not require substantial *a priori* knowledge of short-wave diffraction theory. This potentially provides a better understanding of the behaviour of the solution to the scattering problem to non-specialists in the area of short-wave diffraction. In numerical analysis, deeper understanding of such results may ultimately lead to development of new, more efficient algorithms for solving scattering problems.

B. Another aim of this thesis is to develop, analyse and implement an efficient numerical method for the approximation of the integrals of the form,

$$\int\int_D M(s,t)\exp\left(ik\Psi(s,t)\right)dsdt, \tag{1.14}$$

for any wavenumbers k, where D is either rectangular or triangular domain. The function M in (1.14) may have algebraic singularities and the function Ψ (called the "phase-function") is non-linear in s and t and is smooth.

The Galerkin discretisation of (1.10) leads to a dense system of linear equations, the coefficients of which are of the form (1.14). An accurate computation of such integrals using classical quadratures becomes unfeasible as $k \to \infty$. Such systems are common for hybrid boundary element methods and their efficient assembly remains a substantial hurdle in any practical implementation of the Galerkin method.

Filon-type methods are designed for computing one-dimensional highly-oscillatory integrals. We utilise a Filon-type quadrature method known as Filon-Clenshaw-Curtis quadrature when constructing the numerical scheme for an efficient and accurate computation of (1.14).

C. Once the numerical integration method in **B** is developed, one must make sure that the accuracy of the Galerkin method does not deteriorate due to inexact approximation of integrals (1.14). Thus, robust error estimates for the resulting "fully-discrete" Galerkin method (that incorporates the quadrature) must be derived.

D. This will lead us to an interesting subproblem: a robust error analysis of the Filon-Clenshaw-Curtis quadrature. Filon-type quadratures achieve high accuracy for the approximation of integration of $f(x)\exp(ikx)$ with a relatively small number of quadrature points. In fact, for sufficiently smooth functions f and a fixed number of quadrature points, N, the accuracy increases as the wavenumber k grows. Only recently in [41] have error estimates been obtained for the Filon-Clenshaw-Curtis quadrature that are explicit in the wavenumber k, the number of quadrature points N and Sobolev regularity of the integrand function f. However, error estimates for the Filon-type quadratures applied to highly-oscillatory integrals with algebraic singularities remained an open problem which we also solve in this thesis.

1.3 The main achievements of the thesis

The main achievements of this thesis are:

1. This thesis proves the result (1.8) for the model problem of scattering by a circle using only elementary arguments following the methods of analysis developed in [5] and [78] and making them more explicit. The extension of this proof to more general convex scatterers is also outlined.

2. This thesis develops a numerical integration method for accurate and efficient computation of highly-oscillatory double integrals (1.14) that arise in scattering problems. The numerical method transforms the integrals into a form suitable for Filon-type integration. This has been briefly discussed in our recent publication [39].

3. This thesis makes two novel contributions in the analysis of the Filon-Clenshaw-Curtis quadrature. Firstly, the analysis of the Filon-Clenshaw-Curtis quadrature is extended to obtain novel robust error estimates in terms of the wave parameter k, number of quadrature points and the regularity of the integrand function. Secondly, the error estimates for the composite Filon-Clenshaw-Curtis quadrature applied on graded meshes are derived. This technique can be successfully applied to compute singular highly-oscillatory integrals that often arise when solving scattering problems with error that is independent of k and unaffected by the strength of the singularities.

4. This thesis derives an error analysis for the fully-discrete Galerkin method, i.e. Galerkin method incorporating quadrature for computing (1.14). The convergence properties of the semi-discrete Galerkin method are preserved in the resulting fully-discrete version.

5. This thesis implements the fully-discrete Galerkin scheme in MATLAB for a standard formulation of the hybrid asymptotic numerical method and for a recently derived "star-combined" formulation [96]. Numerical results show even better convergence rates than the theoretical prediction for $k \to \infty$.

1.4 Outline of the thesis

Before presenting the layout of the thesis, some general remarks should be made about the structure of each chapter. The four main chapters of the thesis address its four main aims. Each chapter contains an introduction where the motivation for the chapter and its content are outlined. In addition, a review of the existing literature is presented.

In Chapter 2, we present the mathematical framework for the thesis. We discuss in more detail the hybrid numerical-asymptotic method for two boundary integral equations: with standard combined and star-combined potential operators. We survey the available literature for the analysis of the semi-discrete Galerkin method and present the error estimates for the fully-discrete Galerkin method.

The error estimates for the numerical approximation of the slowly-varying part $V(\boldsymbol{x}, k)$ of the solution $v(\boldsymbol{x})$ in (1.8) are obtained from asymptotic theory. In Chapter 3, we derive these estimes for the case of scattering by a circle. The strategy for solving the scattering problem for the circle underpins the *Model Problem Method* developed in [5]. We outline the extension of the results obtained for the case of the circle to more general convex scatterers.

In Chapter 4, we present a novel technique for computing the highly-oscillatory double integrals with singularities that arise from the the Galerkin discretization. We do this by analysing the behaviour of the double integrals and transforming them into integrals

that are suitable for Filon-type quadrature. We then analyse the behaviour of the resulting integrands by studying the location and type of singularities that are present in the integrand.

In Chapter 5, we provide an extensive literature survey for the Filon-type quadratures. We extend the results of [41] to obtain error estimates for the Filon-Clenshaw-Curtis quadrature in terms of the Chebyshev norm of the slowly-varying part of the integrand, f, instead of the Sobolev norm. This allows us to derive the error bounds for the composite Filon-Clenshaw-Curtis quadrature applied on graded meshes. Such composite Filon-Clenshaw-Curtis methods provide accurate and efficient tools for the approximation of the highly-oscillatory integrals with algebraic singularities.

Finally, in Chapter 6, we derive the k-explicit error estimates of the fully-discrete Galerkin approximation. We conclude the chapter with a discussion on the results of numerical experiments.

Chapter 2

Mathematical framework

2.1 Introduction

Let us consider the forward model problem of the acoustic scattering of an incident plane wavefield $u^I(\boldsymbol{x}) = \exp(ik\boldsymbol{x}\cdot\boldsymbol{a})$ by a bounded sound soft obstacle Ω with a boundary Γ. The unit vector $\boldsymbol{a}$ denotes the direction of incidence. We seek the total wavefield $u := u^I + u^S$ that solves the exterior model scattering problem,

$$\Delta u + k^2 u = 0 \quad \text{in } \mathbb{R}^d \setminus \overline{\Omega}, \tag{2.1}$$

$$u = 0 \quad \text{on } \Gamma, \tag{2.2}$$

$$\frac{\partial u^S}{\partial r} - iku^S = \mathbf{o}\left(r^{\frac{1-d}{2}}\right), \tag{2.3}$$

as $|r| \to \infty$ uniformly in $\hat{x} = x/r$, where u^S represents the scattered wavefield, u^I represents the incident wavefield, and d denotes the dimension.

The boundary value problem (2.1)-(2.3) can be reformulated as an integral equation on the boundary Γ using Green's identities. Green's identities follow from the divergence theorem [1] and are defined as follows:

- *Green's first identity*

  For $u \in C^1(\Omega)$ and $v \in C^2(\Omega \cup \Gamma)$,

$$\int_\Omega u\Delta v \; d\boldsymbol{x} = \int_\Gamma u\frac{\partial v}{\partial \boldsymbol{n}} \; ds(\boldsymbol{x}) - \int_\Omega \nabla u \cdot \nabla v \; d\boldsymbol{x}; \tag{2.4}$$

- *Green's second identity*

[1] **Divergence Theorem**
Let $\boldsymbol{F} : \Omega \cup \Gamma \to \mathbb{R}^d$, $d = 2, 3$, with each component of $\boldsymbol{F}$ in $C^1(\Omega \cup \Gamma)$, then

$$\int_\Omega \nabla \cdot \boldsymbol{F}(\boldsymbol{x}) \; d\boldsymbol{x} = \int_\Gamma \boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{F}(\boldsymbol{x}) \; ds(\boldsymbol{x}).$$

If additionally $u \in C^2(\Omega \cup \Gamma)$, then

$$\int_\Omega (u\Delta v - v\Delta u)\, d\boldsymbol{x} = \int_\Gamma \left(u\frac{\partial v}{\partial \boldsymbol{n}} - v\frac{\partial u}{\partial \boldsymbol{n}} \right)\, ds(\boldsymbol{x}). \tag{2.5}$$

A useful integral representation for the solution u of (2.1) relating u to its values on the boundary Γ is given in the Green's theorem presented below.

Theorem 2.1. ***Green's Theorem*** *[33, Theorem 3.3] Let $u \in C^2(\mathbb{R}^d \setminus \overline{\Omega})$ be a solution to the Helmholtz equation (2.1) satisfying the Sommerfeld radiation condition (2.3). Then,*

$$u(\boldsymbol{x}) = \int_\Gamma \left(u(\boldsymbol{y})\frac{\partial \Phi_k(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{n}(\boldsymbol{y})} - \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{y})\Phi_k(\boldsymbol{x},\boldsymbol{y}) \right) ds(\boldsymbol{y}), \qquad \boldsymbol{x} \in \mathbb{R}^d \setminus \overline{\Omega} \tag{2.6}$$

where the unit vector $\boldsymbol{n}(\boldsymbol{x})$ is an outward normal to Γ and $\partial/\partial\boldsymbol{n}$ denotes a derivative in the normal direction. The function $\Phi_k(\boldsymbol{x},\boldsymbol{y})$ defined as

$$\Phi_k(\boldsymbol{x},\boldsymbol{y}) = \frac{i}{4}H_0^{(1)}(k|\boldsymbol{x}-\boldsymbol{y}|), \ \ in \ \ \mathbb{R}^2,$$

where $H_0^{(1)}(t)$ is the Hankel and

$$\Phi_k(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{4\pi}\frac{e^{ik|\boldsymbol{x}-\boldsymbol{y}|}}{|\boldsymbol{x}-\boldsymbol{y}|}, \quad in \ \ \mathbb{R}^3.$$

represents the fundamental solution of the Helmholtz equation.

The plan for this chapter is as follows. In Section 2.2, we introduce the standard combined and star-combined potential integral equations for determining $v = \partial u/\partial \boldsymbol{n}$. We will derive the boundary equations from Green's integral representation theorem. In Section 2.3.1 we outline the construction of an *optimal* approximation space $\mathbb{X}$ for the semi-discrete Galerkin method that solves the boundary integral equations. In Section 2.3.2, we outline the strategy for construction and analysis of the fully-discrete Galerkin method.

2.2 Boundary integral representation

In this section, we derive two boundary integral equation formulations for the scattering problem (2.1)-(2.3) in Section 2.2.1 and Section 2.2.2. Using Green's second identity and bearing in mind that u^I is a free space solution to the Helmholtz equation, we obtain,

$$\begin{aligned}
&\int_\Gamma \left(u^I(\boldsymbol{y}) \frac{\partial \Phi_k(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{n}(\boldsymbol{y})} - \frac{\partial u^I}{\partial \boldsymbol{n}}(\boldsymbol{y}) \Phi_k(\boldsymbol{x},\boldsymbol{y}) \right) ds(\boldsymbol{y}) \\
&= \int_\Omega \left(u^I(\boldsymbol{x}) \Delta \Phi_k(\boldsymbol{x},\boldsymbol{y}) - \Phi_k(\boldsymbol{x},\boldsymbol{y}) \Delta u^I(\boldsymbol{x}) \right) d\boldsymbol{x} \\
&= \int_\Omega u^I(\boldsymbol{x}) \left(\Delta \Phi_k(\boldsymbol{x},\boldsymbol{y}) + k^2 \Phi_k(\boldsymbol{x},\boldsymbol{y}) \right) d\boldsymbol{x} = 0, \qquad \boldsymbol{x} \in \mathbb{R}^d \setminus \overline{\Omega}.
\end{aligned} \tag{2.7}$$

On the other hand, we also have, for $\boldsymbol{x} \in \mathbb{R}^d \setminus \overline{\Omega}$,

$$u^S(\boldsymbol{x}) = \int_\Gamma \left(u^S(\boldsymbol{y}) \frac{\partial \Phi_k(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{n}(\boldsymbol{y})} - \frac{\partial u^S}{\partial \boldsymbol{n}}(\boldsymbol{y}) \Phi_k(\boldsymbol{x},\boldsymbol{y}) \right) ds(\boldsymbol{y}). \tag{2.8}$$

Therefore, adding (2.7) and (2.8) and imposing the boundary condition (2.2), we obtain,

$$\begin{aligned}
u^S(\boldsymbol{x}) &= \int_\Gamma \left(u(\boldsymbol{y}) \frac{\partial \Phi_k(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{n}(\boldsymbol{y})} - \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{y}) \Phi_k(\boldsymbol{x},\boldsymbol{y}) \right) ds(\boldsymbol{y}) \\
&= -\int_\Gamma \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{y}) \Phi_k(\boldsymbol{x},\boldsymbol{y}) ds(\boldsymbol{y}), \qquad \boldsymbol{x} \in \mathbb{R}^d \setminus \overline{\Omega}.
\end{aligned}$$

Hence we obtain, see also [34, Theorem 3.12]:

$$u(\boldsymbol{x}) = u^I(\boldsymbol{x}) - \int_\Gamma \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{x}) \Phi_k(\boldsymbol{x},\boldsymbol{y}) \mathrm{d}s(\boldsymbol{y}), \qquad \boldsymbol{x} \in \mathbb{R}^d \setminus \overline{\Omega}. \tag{2.9}$$

2.2.1 Standard combined-potential boundary integral equation

We introduce the following notation for the two operators $\mathcal{S}_k$ and $\mathcal{D}_k'$ - the *single-layer* potential and the adjoint *double-layer* potential respectively. Both operators solve the Helmholtz equation and satisfy the Sommerfeld radiation condition, and are defined as follows:

$$\mathcal{S}_k \phi(\boldsymbol{x}) := \int_\Gamma \Phi_k(\boldsymbol{x},\boldsymbol{y}) \phi(\boldsymbol{y}) \mathrm{d}S(\boldsymbol{y}) \qquad \text{and} \qquad \mathcal{D}_k' \phi(\boldsymbol{x}) := \int_\Gamma \frac{\partial \Phi_k(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{n}(\boldsymbol{x})} \phi(\boldsymbol{y}) \mathrm{d}S(\boldsymbol{y}). \tag{2.10}$$

Taking the trace of (2.9) and using (2.2) we obtain:

$$\mathcal{S}_k \left(\frac{\partial u}{\partial \boldsymbol{n}} \right)(\boldsymbol{x}) = u^I(\boldsymbol{x}), \quad \boldsymbol{x} \in \Gamma. \tag{2.11}$$

This equation is an *integral equation of the first kind*. For an infinite set of parameters

k, called *spurious frequencies*, the integral equation (2.11) is not uniquely solvable. In other words, there exist non-trivial solutions u to the Helmholtz equation in the interior domain Ω satisfying homogeneous Neumann boundary conditions $\partial u/\partial \boldsymbol{n}(\boldsymbol{x}) = 0$, see [33, Theorem 3.30].

A second kind boundary integral equation can be derived by taking the trace of (2.11) (see [66, Theorem 6.13]),

$$\left(\frac{1}{2}I + \mathcal{D}_k'\right)\frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{x}) = \frac{\partial u^I}{\partial \boldsymbol{n}}(\boldsymbol{x}), \quad \boldsymbol{x} \in \Gamma. \tag{2.12}$$

This boundary equation also suffers from spurious frequencies as proved in [33, Theorem 3.32] and therefore is not uniquely solvable either.

To overcome this, the two integral equations are combined with a coupling parameter $\eta \in \mathbb{R} \setminus \{0\}$, to obtain the *combined* potential integral equation:

$$\frac{1}{2}v + \mathcal{D}_k' v - i\eta \mathcal{S}_k v = \frac{\partial u^I}{\partial \boldsymbol{n}} - i\eta u^I, \qquad \text{on } \Gamma, \tag{2.13}$$

where $v(\boldsymbol{x}) := \partial u/\partial \boldsymbol{n}(\boldsymbol{x}) \in L^2(\Gamma)$, for $\boldsymbol{x} \in \Gamma$. The combined potential integral equation is often attributed to Burton and Miller [34].

Theorem 2.2. *If $u \in C^2(\mathbb{R}^2 \setminus \Omega) \cap C(\mathbb{R}^2 \setminus \overline{\Omega})$ satisfies the boundary value problem (2.1)-(2.3) then, for every $\eta \in \mathbb{R} \setminus \{0\}$, the function $v := \partial u/\partial \boldsymbol{n} \in L^2(\Gamma)$ is a unique solution of the integral equation:*

$$R_k v = f_k, \tag{2.14}$$

where the integral operator is defined as

$$R_k v := \frac{1}{2}v + \mathcal{D}_k' v - i\eta \mathcal{S}_k v, \tag{2.15}$$

and

$$f_k = \partial_{\boldsymbol{n}} u^I - i\eta u^I. \tag{2.16}$$

Conversely, if $v \in L^2(\Gamma)$ satisfies (2.15) for some $\eta \in \mathbb{R} \setminus \{0\}$, then $u \in C^2(\mathbb{R}^2 \setminus \Omega) \cap C(\mathbb{R}^2 \setminus \overline{\Omega})$ and satisfies the boundary value problem (2.1)-(2.3).

Proof. See the proof in [27, Theorem 2.5]. □

Remark 2.3. *As shown in [26], the choice of the coupling parameter is essential for minimising the condition number of the boundary operator in (2.13). The recommended choices for η (in terms of minimizing the condition number) are to choose it to be proportional to k for large k, and to $(\log k)^{-1}$ for small k in 2D. These choices have initially been*

justified by theoretical studies on the circle and sphere [66, 68] and on the basis of computational experience [21]. Recently, analysis and numerical experiments [26, 12] have shown that these commonly recommended choices are optimal for a wide variety of domains.

Unlike equations (2.11) and (2.12), the boundary integral equation (2.13) is uniquely solvable in $C(\Gamma)$ when Γ is sufficiently smooth, for all values of k such that Im $k \geq 0$ as proved in [34, Theorem 3.34]. Moreover, the existence and uniqueness of the solution in the Sobolev space $H^s(\Gamma)$, $-1 \leq s \leq 0$ have recently been proved in [27] for the case when Γ is a general Lipschitz boundary for Im $k \geq 0$.

We denote the *standard combined potential integral operator* by $\mathcal{R}_k$:

$$\mathcal{R}_k := \frac{1}{2}I + \mathcal{D}'_k - i\eta\mathcal{S}_k, \tag{2.17}$$

The standard combined potential integral equation (2.13) is closely related to another combined potential integral equation formulation:

$$\mathcal{P}_k v := \frac{1}{2}v + \mathcal{D}_k v - i\eta\mathcal{S}_k v = -u^I, \tag{2.18}$$

independently suggested (in 1965) by Brakhage and Werner [17], Leis [69] and Panic [86]. The operator $\mathcal{D}_k$ is called the double-layer potential. The operators $\mathcal{R}_k$ and $\mathcal{P}_k$ are adjoint operators in $L^2(\Gamma)$. They have the same spectrum, norm and condition number [25]. The solution of the equation (2.13) is known to be the normal derivative of the solution to the scattering boundary value problem. In this thesis, we consider the operator $\mathcal{R}_k$ and the equation (2.13) since the behaviour of its solution is better understood, particularly for the case of convex scatterers.

2.2.2 Star-combined potential boundary integral equation

The star-combined boundary integral equation has been introduced in [96].

Theorem 2.4. *[96, Lemma 4.1] Suppose $u \in C^2(\mathbb{R}^2 \setminus \Omega) \cap C(\mathbb{R}^2 \setminus \overline{\Omega})$ solves the boundary value problem (2.1)-(2.3) and let $\eta \in L^\infty(\Gamma)$. Then, $v := \partial u/\partial \boldsymbol{n} \in L^2(\Gamma)$ satisfies the integral equation*

$$\mathcal{A}_k v = f, \tag{2.19}$$

where the integral operator A_k is defined as

$$\mathcal{A}_k := (\boldsymbol{x} \cdot \boldsymbol{n}(\boldsymbol{x}))\left(\frac{1}{2}I + \mathcal{D}'_k\right) + \boldsymbol{x} \cdot \nabla_\Gamma \mathcal{S}_k - i\eta\mathcal{S}_k, \tag{2.20}$$

and ∇_Γ *denotes the surface gradient,*

$$\nabla_\Gamma \mathcal{S}_k \psi(\boldsymbol{x}) := \int_\Gamma \left(\nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}, \boldsymbol{y}) - \boldsymbol{n}(\boldsymbol{x}) \frac{\partial \Phi(\boldsymbol{x}, \boldsymbol{y})}{\partial \boldsymbol{n}(\boldsymbol{x})} \right) \psi(\boldsymbol{y}) d\boldsymbol{y},$$

and the known function f *is defined in terms of* u^I *by:*

$$f := \boldsymbol{x} \cdot \nabla u^I - i\eta u^I.$$

Proof. The proof follows the proof of Lemma 4.1 in [96]. Take the surface gradient of (2.11) (valid since $\mathcal{S}_k : L^2(\Gamma) \to H^1(\Gamma)$ and $\partial u / \partial \boldsymbol{n} \in L^2(\Gamma)$) to obtain,

$$\nabla_\Gamma \mathcal{S}_k \frac{\partial u}{\partial \boldsymbol{n}} = \nabla_\Gamma u^I.$$

Then, equation (2.19) follows by taking the scalar product of the last equation with $\boldsymbol{x}$ [2], adding $(\boldsymbol{x} \cdot \boldsymbol{n})$ times (2.12) and subtracting $i\eta$ times (2.11). □

The star-combined operator (2.20) is coercive for all Lipschitz star-shaped domains, uniformly in k, see Theorem 2.5 below. The uniform coercivity of the operator leads to the stability analysis and the k-explicit error analysis of the corresponding hybrid Galerkin method for any choice of the approximation space. The uniform coercivity property of the star-combined operator (2.20) is advantageous over the standard combined operator (2.15) because the standard operator has only been proven to be coercive for a circle and a sphere, in [40]. The proof of the uniform coercivity of the star-combined operator is obtained in [96] by a novel application of the Morawetz and Ludwig identity [78].

Theorem 2.5. *[96, Theorem 1.1] Suppose that* Ω *is a bounded star-shaped Lipschitz domain and* $\boldsymbol{x}$ *is the position vector relative to an origin from which* Ω *is star-shaped. Then, for all* $\phi \in L^2(\Gamma)$*,*

$$Re(\mathcal{A}_k \phi, \phi)_{L^2(\Gamma)} \geq \gamma \|\phi\|^2_{L^2(\Gamma)}, \tag{2.21}$$

where $\mathcal{A}_k$ *is defined as in (2.20) with*

$$\eta(\boldsymbol{x}) = kr + i\frac{d-2}{2}, \tag{2.22}$$

where $r = |\boldsymbol{x}|$ *and*

$$\gamma := \frac{1}{2} ess \inf_{\boldsymbol{x} \in \Gamma} (\boldsymbol{x} \cdot \boldsymbol{n}(\boldsymbol{x})) > 0. \tag{2.23}$$

If a vector $\boldsymbol{n}$ is well-defined everywhere on the boundary $ess\inf$ in (2.23) can be replaced by inf.

[2] in [96] the combined equation is presented with a suitable vector field $\boldsymbol{Z}$ which we replaced with $\boldsymbol{x}$

When Γ is a unit circle or sphere, the star-combined operator $\mathcal{A}_k$ reduces to the standard combined operator $\mathcal{R}_k$ with the coupling parameter η given by (2.22). Therefore, Theorem 2.5 provides alternative proofs of the coercivity results as $k \to \infty$ of [40], and also shows that coercivity holds uniformly for all k on the circle and the sphere provided we make the choice of coupling constant as in (2.22).

The Galerkin method for solving the star-combined equation (2.19) is essentially no more difficult to implement than the standard-combined operator (2.15). We will discuss the implementation of both standard and star-combined formulations in more detail in Chapter 6.

2.3 Galerkin method and its error analysis

Let us consider in more detail the structure for the analysis of boundary element methods and survey the literature for developments in these areas.

Recall, for example, the standard combined boundary integral equation (2.15),

$$\mathcal{R}_k v = f_k.$$

The variational formulation for (2.15) seeks $v \in L^2(\Gamma)$ such that,

$$a_k(v, w) := (\mathcal{R}_k v, w) = (f_k, w), \qquad \text{for all} \quad w \in L^2(\Gamma). \tag{2.24}$$

Using standard arguments in the abstract theory of variational methods we deduce that the solution, $v \in L^2(\Gamma)$, of (2.24) is unique if the sesquilinear form, a_k, is continuous and coercive (Lax-Milgram), i.e. there exist constants $B_k > 0$ and $\alpha_k > 0$ that satisfy the inequalities

$$\begin{aligned} a_k(v, v) &\geq \alpha_k \|v\|^2 \qquad \forall v \in L^2(\Gamma), \qquad \textit{coercivity}, \\ |a_k(u, v)| &\leq B_k \|u\| \|v\| \qquad \forall u, v \in L^2(\Gamma), \qquad \textit{continuity}. \end{aligned} \tag{2.25}$$

Here the notation $\|\cdot\|$ represents the L^2 norm on Γ, $\|\cdot\|_{L^2(\Gamma)}$.

<u>I. Error estimates for the semi-discrete Galerkin method:</u>

Given a finite dimensional approximation space $\mathbb{X}$, the uniquely determined semi-discrete Galerkin solution $\tilde{v} \in \mathbb{X}$ of the system of equations

$$a_k(\tilde{v}, \tilde{w}) := (f_k, \tilde{w}), \qquad \text{for all} \quad \tilde{w} \in \mathbb{X}, \tag{2.26}$$

satisfies the quasi-optimal error estimate (Cea's Lemma)

$$\|v - \tilde{v}\| \leq \frac{B_k}{\alpha_k} \inf_{\tilde{w} \in \mathbb{X}} \|v - \tilde{w}\|. \tag{2.27}$$

The convergence analysis of the semi-discrete Galerkin scheme for the discretisation of the boundary integral equations consists of two components. The first component is the proof of stability of the Galerkin method. This is determined by the k-explicit bounds on the coercivity and continuity constants α_k and B_k in (2.25). The second component is the construction of the approximation space $\mathbb{X}$ which ensures the best approximation error is small for large k.

(a) Continuity and coercivity: The stability of the Galerkin approximation as $k \to \infty$ may be determined by the k-explicit estimates of the continuity and coercivity constants B_k and α_k. Note that the obvious choice for the continuity constant is $B_k = \|\mathcal{R}_k\|$, [25]:

$$|a_k(u,v)| = |(\mathcal{R}_k u, v)| \leq \|\mathcal{R}_k\| \|u\| \|v\|.$$

Moreover, the ratio of B_k/α_k is bounded below by the *condition number* of the operator $\mathcal{R}_k$, see e.g. [25]:

$$\frac{B_k}{\alpha_k} \geq \text{cond } \mathcal{R}_k := \|\mathcal{R}_k\| \|\mathcal{R}_k^{-1}\|. \tag{2.28}$$

<u>Estimates for the standard combined operator $\mathcal{R}_k$ defined in (2.17)</u>

The estimates for these constants have only recently been derived for the boundary integral equation (2.17). In [40] the upper bounds for $\|\mathcal{R}_k\|$ have been derived for 2D and 3D smooth curves and surfaces Γ, with the coupling parameter $\eta = k$ to be bounded as follows, for sufficiently large k,

$$\begin{aligned} \|\mathcal{R}_k\| &\leq C_1 k^{1/3}, \quad \text{for a circle in 2D or a sphere in 3D,} \\ \|\mathcal{R}_k\| &\leq C_2 k^{1/2}, \quad \text{for a Lipschitz scatterer in 2D} \\ \|\mathcal{R}_k\| &\leq C_3 k, \quad \text{for a Lipschitz scatterer in } 3D. \end{aligned}$$

where the constants C_1, C_2 and C_3 are independent of k. Furthermore, in [40] a lower bound on the coercivity constant α_k has been proved for the case when Γ is a circle to be of the form,

$$\alpha_k \geq \frac{1}{2}, \quad \text{for} \quad k \geq k_0. \tag{2.29}$$

Also in [40], the operator $\mathcal{R}_k$ has been shown to be uniformly coercive for the case of the sphere, i.e.

$$\alpha_k \geq c, \quad \text{for} \quad k \geq k_0,$$

where c is independent of k.

These results have been further extended in [26] and then in the follow up paper [12] and experimentally verified in [13]. In [26, 12] the upper and lower bounds on the condition number of the operator $\mathcal{R}_k$ have been derived for various geometries of the

scatterer. The bounds are explicit in the wavenumber k, the coupling parameter η and the dimension $d = 2, 3$. The results illustrate strong dependence of the bounds for the condition number on the geometry of the domains: for example, in 2D, with $\eta = k$,

$$\begin{aligned} \text{cond } \mathcal{R}_k &\backsim k^{1/3} && \text{for a circle,} \\ \text{cond } \mathcal{R}_k &\backsim k^{1/2} && \text{for a star-like polygon.} \end{aligned}$$

In [13] a numerical method for the estimation of the coercivity constant α_k in (2.25) has been proposed for general boundary integral operators in acoustic scattering by exploiting properties of their *numerical range* (or *field of values*). The numerical range $W(T)$ of a bounded linear operator T in Hilbert space $\mathcal{H}$ is defined as

$$W(T) := \{(Tu, u), u \in \mathcal{H}, \|u\| = 1\}. \tag{2.30}$$

The operator T is coercive if and only if 0 is not in the closure of its numerical range as shown in [13, Proposition 3.3]. Numerical experiments conducted in [13] suggest that the standard combined operator $\mathcal{R}_k$ is coercive for a wide range of domains by showing that the numerical range of the operator $\mathcal{R}_k$,

$$W(\mathcal{R}_k) = \{a_k(u, u), u \in L^2(\Gamma), \|u\| = 1\},$$

does not contain 0.

<u>Estimates for the star-combined operator $\mathcal{A}_k$</u>

For the star-combined operator, the coercivity constant α_k that does not depend on k is given in (2.23). Moreover, if Γ is Lipschitz, the operator $\mathcal{A}_k$ is a bounded operator on $L^2(\Gamma)$ and the continuity constant B_k satisfies:

$$\|B_k\| \leq Ck^{(d-1)/2} \left(1 + \frac{\|\eta\|_\infty}{k}\right), \tag{2.31}$$

for all $k \geq k_0 > 0$, where d denotes the dimension and where C is independent of k and η [96, Theorem 4.2]. Note that with η given by (2.22), $\|\eta\|_\infty = O(k)$ as $k \to \infty$, therefore, $\|B_k\|_\infty \leq Ck^{(d-1)/2}$. As we will see later, the implementation of the star-combined equation (2.19) requires some additional work to "set up" compared to the standard combined equation (2.19). However, the star-combined operator has the advantage of being uniformly coercive for all Lipschitz star-shaped domains.

(b) Approximation spaces: Constructing an *optimal* approximation space and obtaining explicit best approximation error estimates is a difficult task. It requires detailed results from the asymptotic theory on the behaviour of the exact solution of the cor-

responding boundary integral equation. Such a robust analysis has been carried out in [40] and [27] for smooth and convex obstacles and convex polygons respectively.

In order to ensure that the best approximation error is small, an ansatz such as (1.8) is incorporated in the approximating space $\mathbb{X}$ so that only the slowly-varying function V is approximated rather than the highly-oscillatory function v. For example, for the case of the scattering by convex obstacles,

$$v(\boldsymbol{x}) = kV(\boldsymbol{x}, k)\exp(ik\boldsymbol{x}\cdot\boldsymbol{a}), \quad \boldsymbol{x}\in\Gamma. \tag{2.32}$$

for planar wave incidence. Essentially, the approximation space is then constructed so that the basis functions are a product of a suitable polynomial basis and the known oscillatory function $\exp(ik\boldsymbol{x}\cdot\boldsymbol{a})$.

From the asymptotic theory (see also Chapter 3), for the case of smooth convex obstacles, the slowly-varying function V is known to have a different asymptotic behaviour in the illuminated, shadow and transition domains (or "Fock" domains) of the boundary Γ. In [40], a partition of unity is introduced to separate the boundary Γ into these regions. The slowly-varying function V is then approximated by polynomials separately on each region of the boundary. Then, the approximation space is of the form,

$$\mathbb{X} := \mathbb{X}_{Illuminated}\times\mathbb{X}_{Transition}\times\mathbb{X}_{Shadow}.$$

The wavenumber-explicit, best approximation error can then be found by utilising the information on the asymptotic behaviour of V. We discuss this in more detail in Chapter 3.

II. Error estimates for the fully-discrete Galerkin method:

The Galerkin discretisation (2.26) leads to a dense system of linear equations that requires the computation of highly-oscillatory double integrals with singularities: $a_k(\phi,\psi)$, where ϕ, $\psi\in\mathbb{X}$. Using conventional numerical quadratures to assemble the discretisation matrix is computationally expensive and would destroy the attractive theoretical computational cost of the hybrid method [40] for high-frequency problems.

In practice, we seek a *fully-discrete Galerkin solution*, $\tilde{\tilde{v}}\in\mathbb{X}$, that satisfies the following system of equations:

$$\tilde{a}_k(\tilde{\tilde{v}},\phi) = (f_k,\phi), \quad \text{for all } \phi\in\mathbb{X}, \tag{2.33}$$

where $\tilde{a}_k(\psi,\phi)$ denotes an approximation to $a_k(\psi,\phi)$ obtained by an efficient numerical quadrature suitable for highly-oscillatory integrals.

Cea's Lemma (2.27) determines the error bounds of the semi-discrete solution that is obtained assuming that integrals $a_k(\psi,\phi)$, $\phi,\psi\in\mathbb{X}$ are computed exactly.

The following theorem, known as the Strang Lemma, determines the error bounds of the fully-discrete Galerkin method. The error bounds incorporate the error contribution of the inexact approximation of the integrals $a_k(\psi, \phi)$ by $\tilde{a}_k(\psi, \phi)$ for all $\phi, \psi \in \mathbb{X}$.

Theorem 2.6. ***Strang Lemma*** *[30, Theorem 4.2.2]*

Consider the family of discrete problems (2.33) and assume that the associated sesquilinear form $\tilde{a}_k$ is coercive and continuous:

1) $\tilde{a}_k(z, z) \geq \tilde{\alpha}_k ||z||^2 \quad \forall z \in \mathbb{X}$,

2) $|\tilde{a}_k(z, w)| \leq \tilde{B}_k ||z|| ||w|| \quad \forall z, w \in \mathbb{X}$.

Then the following holds:

$$\|v - \tilde{\tilde{v}}\| \leq \inf_{z \in \mathbb{X}} \left\{ \left(1 + \frac{\tilde{B}_k}{\tilde{\alpha}_k} \right) \|v - z\| + \frac{1}{\tilde{\alpha}_k} \sup_{w \in \mathbb{X}} \frac{|a_k(z, w) - \tilde{a}_k(z, w)|}{\|w\|} \right\}, \tag{2.34}$$

where $\tilde{\tilde{v}} \in \mathbb{X}$ is the fully-discrete Galerkin solution that satisfies the following system of equations (2.33).

We will use Theorem 2.6 in Section 6.4 of Chapter 6, where we derive the error bounds of the fully-discrete Galerkin solution.

2.3.1 Semi-discrete Galerkin method

In this section, we give an overview of the semi-discrete Galerkin method applied to the exterior scattering problem for 2D scatterers with smooth and convex obstacles. We present the error estimate for the semi-discrete Galerkin method obtained in [40].

We parametrise the convex boundary Γ by 2π-periodic parametrisation:

$$\Gamma = \{\boldsymbol{\gamma}(s) : s \in [0, 2\pi]\},$$

where $|\boldsymbol{\gamma}'(s)| > 0$ for all $s \in [0, 2\pi]$. We define the illuminated part of the boundary as an interval in $[0, 2\pi]$ where $\boldsymbol{n}(\boldsymbol{\gamma}(s)) \cdot \boldsymbol{a} < 0$, where $\boldsymbol{n}$ is the outward normal to the boundary Γ and $\boldsymbol{a}$ is the direction of incidence. Similarly, in the shadow part we have $\boldsymbol{n}(\boldsymbol{\gamma}(s)) \cdot \boldsymbol{a} > 0$. Finally, we define the two transition regions as intervals which contain the *transition points* $s = t_1$ and $s = t_2$, satisfying: $\boldsymbol{n}(\boldsymbol{\gamma}(s)) \cdot \boldsymbol{a} = 0$, see Figure 2.1.

We denote four intervals in $[0, 2\pi]$ representing the illuminated, two transition and shadow parts of the boundary as Λ_1, Λ_2, Λ_3 and Λ_4 respectively. The solution, v, of the boundary integral equation, e.g. (2.15), is known to be of the form (2.32). The relatively slowly-varying part, V, of the solution has different asymptotic behaviour in the illuminated, shadow and transition regions of the boundary Γ.

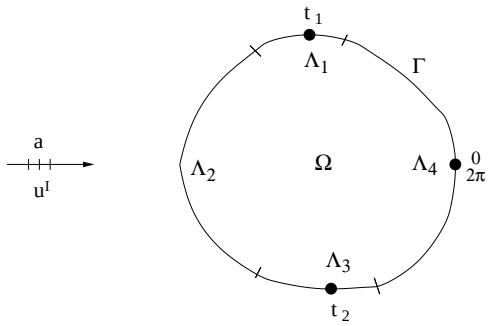


Figure 2.1: *Parameterised boundary is split into four domains, where the solution has different asymptotic expansions as $k \to \infty$.*

Approximation space In order to reflect the structure of the solution, we introduce a partition of unity: for all $j = 1, 2, 3, 4$ and for $\chi_j \in L_\infty[0, 2\pi]$,

$$\text{supp } \chi_j = \Lambda_j, \quad 0 \le \chi_j(s) \le 1, \quad \text{and} \quad \left(\sum_{j=1}^{4} \chi_j \right)(s) = 1, \qquad s \in [0, 2\pi]. \tag{2.35}$$

Then we write the solution in each zone of the boundary as follows:

$$\chi_j(s) v(s) = k V_j(s, k) \exp(ik\boldsymbol{\gamma}(s) \cdot \boldsymbol{a}), \tag{2.36}$$

where $V_j(s, k) := \chi_j(s) V(s, k)$, $j = 1, 2, 3, 4$, denotes the unknown "amplitudes" of v in the illuminated ($l = 2$), transition ($l = 1, 3$), and shadow ($l = 4$) parts of the boundary. The solution v is known to be exponentially small in the shadow region. Thus, we approximate the solution in the shadow part of the boundary Λ_4 by zero.

In this thesis, we consider a partition of unity χ_j, $j = 1, 2, 3, 4$ with χ_j defined as,

$$\chi_j(s) = \begin{cases} 1, & \text{if } s \in \Lambda_j, \\ 0 & \text{if } s \notin \Lambda_j. \end{cases}$$

We can write the solution v as follows,

$$v(s) = k \sum_{j=1}^{3} V_j(s, k) \exp(ik\boldsymbol{\gamma}(s) \cdot \boldsymbol{a}) + \chi_4(s) v(s). \tag{2.37}$$

The asymptotic expansions of V_l, $l = 1, 2, 3$ are presented in Theorem 3.1, Theorem 3.3 and Theorem 3.2 in Chapter 3. Hence, we construct the approximation space $\mathbb{V}_k^{\boldsymbol{d}}$ for the Galerkin discretization,

$$\mathbb{V}_k^{\boldsymbol{d}} = \oplus_{j=1}^{3} \mathbb{V}_k^j, \tag{2.38}$$

where $\mathbb{V}_k^2$ is an approximation subspace for the illuminated part Λ_2 and $\mathbb{V}_k^1$ and $\mathbb{V}_k^3$ for

transition parts, Λ_1, and Λ_3 respectively,

$$\mathbb{V}_k^j = \text{span}\left\{\phi_{j,m}, \quad m = 0, ..., d_j : \quad \phi_{j,m}(s) = kP_m(s)\exp(ik\boldsymbol{\gamma}(s)\cdot\boldsymbol{a}) \ , \ s \in \Lambda_j\right\},$$

where $\{P_m\}_{m=0}^{d_j}$ is a suitable basis for the polynomials of degree d_j.

We seek a semi-discrete Galerkin solution $\tilde{v} \in \mathbb{V}_k^{\boldsymbol{d}}$ which satisfies the following system of equations:

$$a_k(\tilde{v}, w) = (f_k, w), \quad \text{for all } w \in \mathbb{V}_k^d. \tag{2.39}$$

The error of the semi-discrete Galerkin scheme is bounded as follows [40]:

Theorem 2.7. *Let $\tilde{v} \in \mathbb{V}_k^{\boldsymbol{d}}$ be the semi-discrete Galerkin solution with Λ_j defined as follows,*

$$\Lambda_1 = [t_1 - c_1k^{-1/3}, t_1 + c_2k^{-2/9}], \ \Lambda_2 = [t_1 + c_2k^{-2/9}, t_2 - c_2k^{-2/9}],$$
$$\Lambda_3 = [t_2 - c_2k^{-2/9}, t_2 + c_1k^{-1/3}], \ \Lambda_4 = [t_2 + c_1k^{-1/3}, 2\pi] \cup [0, t_1 - c_1k^{-1/3}]. \tag{2.40}$$

Let the vector $\boldsymbol{d}$ in $\mathbb{V}_k^{\boldsymbol{d}}$ be equal to $\boldsymbol{d} = (d_T, d_I, d_T)$. Then for all $6 \le n \le \min\{d_I, d_T\} + 1$, there exists a constant C_n such that

$$\|v - \tilde{v}\| \le C_n \left(\frac{B_k}{\alpha_k}\right)\left[k\left\{k^{-2/3}\left(\frac{k^{1/9}}{d_I}\right)^n + k^{-4/3}\left(\frac{k^{1/9}}{d_T}\right)^n\right\} + \exp\left(-c_0k^\delta\right)\right], \tag{2.41}$$

where C_n is independent of k, and d_I and d_T are the degrees of the polynomial approximations in the illuminated and transition zones respectively.

We will discuss the derivation of this result in more details in Chapter 6.

The three terms in (2.41) in curly brackets represent the error bounds of the best polynomial approximations in the illuminated, transition and shadow parts of the boundary respectively.

Taking the polynomial basis of the same degree in the illuminated and transition domains: $d_I = d_T = d$, the estimate (2.41) simplifies to the following,

$$\|v - \tilde{v}\| \le C_n \left(\frac{B_k}{\alpha_k}\right) k\left\{k^{-4/3}\left(\frac{k^{1/9}}{d}\right)^n + \exp\left(-c_0k^\delta\right)\right\}. \tag{2.42}$$

The error bound (2.42) suggests that in order to maintain accuracy as $k \to \infty$, the polynomial degrees need only grow very mildly with k, more precisely, as $k^{1/9}$. This is a substantial reduction in complexity from their conventional analogues.

2.3.2 Fully-discrete Galerkin method

We can write the equation (2.39) in a matrix form:

$$\boldsymbol{RV} := \begin{pmatrix} R_{[1,1]} & R_{[1,2]} & R_{[1,3]} \\ R_{[2,1]} & R_{[2,2]} & R_{[2,3]} \\ R_{[3,1]} & R_{[3,2]} & R_{[3,3]} \end{pmatrix} \begin{pmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \\ \boldsymbol{V}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{F}_1 \\ \boldsymbol{F}_2 \\ \boldsymbol{F}_3 \end{pmatrix} =: \boldsymbol{F}. \tag{2.43}$$

In the equation (2.43), the vectors $\boldsymbol{V}$ and $\boldsymbol{F}$ and the matrix $\boldsymbol{R}$ are defined as follows.

(i) The right hand side vector, $\boldsymbol{F}$, is defined as follows:

$$\boldsymbol{F} = [\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3]^T, \quad \text{where} \quad \boldsymbol{F}_j(m) = (f_k, \phi_{j,m}),$$

where $f_k(s) = \partial_{\boldsymbol{n}(s)} u^I(s) - i\eta u^I(s)$. Note that $\boldsymbol{F}_j(m)$ is a one-dimensional integral with a non-oscillatory, well-behaved integrand. In fact, since

$$\partial_{\boldsymbol{n}} u^I(s) := \nabla u^I(\boldsymbol{\gamma}(s)) \cdot \boldsymbol{n} = ik(\boldsymbol{a} \cdot \boldsymbol{n}) \exp\left(ik\boldsymbol{\gamma}(s) \cdot \boldsymbol{a}\right),$$

we have

$$\begin{aligned} \boldsymbol{F}_j(m) &:= (f_k, \phi_{j,m}) \\ &= k^2 \int_{\Lambda_j} \left(\partial_{\boldsymbol{n}} u^I - iku^I\right)(s) \overline{\phi_{j,m}(s)} ds = ik^2 \int_{\Lambda_j} \left(\boldsymbol{a} \cdot \boldsymbol{n} - 1\right) P_m(s) ds. \end{aligned} \tag{2.44}$$

Therefore, the entries of the vector $\boldsymbol{F}$ are one-dimensional, smooth and non-oscillatory integrals that can be computed accurately and efficiently using conventional numerical integration methods.

(ii) The vector $\boldsymbol{V} := [\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3]^T$ is a vector of unknown coefficients of the approximating polynomials for $V(\boldsymbol{\gamma}(s), k)$ (that we simply denote by $V(s, k)$ in the remainder of this thesis) in the transition domains and the illuminated zon:

$$V(s,k) \approx \sum_{m=0}^{d_1} \boldsymbol{V}_1(m)\phi_{1,m}(s) + \sum_{m=0}^{d_2} \boldsymbol{V}_2(m)\phi_{2,m}(s) + \sum_{m=0}^{d_3} \boldsymbol{V}_3(m)\phi_{3,m}(s).$$

(iii) The matrix $\boldsymbol{R}$ is a dense matrix, with the (n, m)-th entry of the subblock $R_{[l,j]}$ given

by

$$
\begin{aligned}
R_{[l,j]}(n,m) &= a_k(\phi_{j,n},\phi_{l,m}) \\
&:= (\mathcal{R}_k\phi_{j,n},\phi_{l,m}) \\
&= \left(\frac{1}{2}\phi_{j,n} + \mathcal{D}_k\phi_{j,n} - ik\mathcal{S}_k\phi_{j,n},\phi_{l,m}\right) \\
&= k^2\frac{1}{2}\left(\int_{\Lambda_j} P_n(s)P_m(s)ds\right)\delta_{j,l} \qquad (2.45a)\\
&+ k^2\int_{\Lambda_l}\int_{\Lambda_j} P_n(t)P_m(s)\left[\partial_{\boldsymbol{n}(s)}\Phi_k(\boldsymbol{\gamma}(s),\boldsymbol{\gamma}(t)) - ik\Phi_k(\boldsymbol{\gamma}(s),\boldsymbol{\gamma}(t))\right]\times \\
&\qquad\qquad \times\exp\left(ik\boldsymbol{a}\cdot[\boldsymbol{\gamma}(t)-\boldsymbol{\gamma}(s)]\right)|\boldsymbol{\gamma}'(t)|dtds, \qquad (2.45b)
\end{aligned}
$$

where $\Phi_k(s,t)$ denotes the fundamental solution of the Helmholtz equation in two dimensions:

$$\Phi_k(\boldsymbol{\gamma}(s),\boldsymbol{\gamma}(t)) = \frac{i}{4}H_0^{(1)}\left(k\left|\boldsymbol{\gamma}(t)-\boldsymbol{\gamma}(s)\right|\right). \qquad (2.46)$$

Remark 2.8. *Notice from (2.44) that both the left-hand side and the right-hand side of (2.43) contain a coefficient k^2. In practice, we divide both sides of the equation (2.43) by k^2 and compute the integrals (2.44), (2.45a) and (2.45b) without the coefficient k^2.*

The first term (2.45a) vanishes except when $j = l$, i.e. both s and t belong to the same interval Λ_j. In this case, the integrand of this single integral is a product of two polynomials and therefore the integral can be computed exactly. On the other hand, the second term (2.45b) is a double integral that cannot be integrated exactly. As we show in Lemma 4.3, Chapter 4, these highly-oscillatory double integrals can be written in the form:

$$a_k(\phi_{j,n},\phi_{l,m}) := k^2\int_{\Lambda_j}\int_{\Lambda_l} M(s,t)\exp\left(ik\Psi(s,t)\right)dtds, \qquad (2.47)$$

where $M(s,t)$ is a relatively slowly-varying function with singularities, $\Psi(\boldsymbol{\gamma}(s),\boldsymbol{\gamma}(t))$ is the phase-function, and $\phi_{j,n}$ and $\phi_{l,m}$ are basis elements of the approximation space $\mathbb{V}_k^{\boldsymbol{d}}$.

We denote by $\tilde{a}_k(\phi_{j,n},\phi_{l,m})$ the numerical approximation of the integral $a_k(\phi_{j,n},\phi_{l,m})$ leading us to the fully-discrete system (2.33).

The strategy for obtaining the fully discrete system for solving the exterior model problem is as follows:

- Develop a numerical integration method for accurate and efficient approximation of $a_k(\phi_{j,n},\phi_{l,m})$ in (2.47) (Chapter 4).
- Analyse the wavenumber-explicit error estimates of the numerical integration method (Chapter 5) and derive the wavenumber-explicit error estimates of the fully-discrete Galerkin method (Chapter 6).

- Implement the fully-discrete Galerkin method and carry out the numerical experiments (Chapter 6).

Chapter 3

Asymptotic methods for high-frequency acoustic scattering problems

3.1 Introduction

Asymptotic methods are typically used to construct an approximate solution to the scattering problem as a series, called an *asymptotic expansion*, with respect to a small parameter $1/k$. Geometrical optics and the geometrical theory of diffraction [65] provide a general set of recipes on how to construct the asymptotic expansion of the solution to a scattering problem. Considerable amount of research has been directed towards both constructing these asymptotic expansions and proving error bounds for truncated asymptotic series of the solution, notably by Buslaev [23], Morawetz and Ludwig [78], and Melrose and Taylor [75], among others, see further references in [5].

In this chapter, we will state the main results of the asymptotic theory which provide an asymptotic expansion of the solution to the high-frequency scattering problem (2.1) - (2.3) with respect to a small parameter $1/k$.

These important results are used for the construction and analysis of efficient numerical methods for the computation of high-frequency scattering problems such as the hybrid boundary integral method described in Chapter 2. Rigorous proofs of these results however are very technical.

Our motivation for this chapter is to provide an insight into asymptotic techniques used to justify these results without substantial a priori knowledge of asymptotic high-frequency diffraction theory. This in turn may prove beneficial for further development of new hybrid methods and techniques for their analysis for high frequency scattering problems.

In Chapter 2, we have described the two-dimensional problem of acoustic scattering of the time-harmonic incident wave-field $u^I(\boldsymbol{x}) = \exp(ik\boldsymbol{x}\cdot\boldsymbol{a})$, where the unit vector $\boldsymbol{a}$ denotes the direction of incidence, by a bounded sound-soft obstacle Ω with a smooth and convex boundary Γ. We seek the total wavefield $u(\boldsymbol{x}) = u^I(\boldsymbol{x}) + u^S(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^2 \setminus \overline{\Omega}$, which

satisfies the boundary value problem (2.1)- (2.3). This problem can be reformulated as a boundary integral equation with solution as the normal derivative of the total wavefield, $v(\boldsymbol{x}) := (\partial u/\partial \boldsymbol{n})(\boldsymbol{x})$. The function v solves the following boundary integral equation:

$$\mathcal{R}_k v(\boldsymbol{\gamma}(s)) = f_k(\boldsymbol{\gamma}(s)) \qquad \text{for} \quad \boldsymbol{\gamma}(s) \in \Gamma, \quad s \in [0, 2\pi], \tag{3.1}$$

where $\boldsymbol{\gamma}$ is a 2π-periodic parametrisation of the boundary Γ and $\mathcal{R}_k$ and f_k are defined in (2.13). We denote $\boldsymbol{\gamma}(t_1)$ and $\boldsymbol{\gamma}(t_2)$ as shadow boundary transition points on the boundary Γ, i.e. the normal to the boundary at these points is perpendicular to the direction of incidence $\boldsymbol{a}$. Thus $\boldsymbol{n}(t_j) \cdot \boldsymbol{a} = 0$, $j = 1, 2$, see Figure 1.1.

The solution $v \in L^2(\Gamma)$ of the boundary integral equation (3.1) is shown in this chapter to be *the product of an explicitly known highly-oscillatory function and a relatively slowly-varying function* $V(s, k)$:

$$v(s) = kV(s, k) \exp(ik\boldsymbol{\gamma}(s) \cdot \boldsymbol{a}). \tag{3.2}$$

This information was used in [40] in construction of the approximation space for the Galerkin method and its analysis as described in Chapter 2. Let us now describe the three main results of the asymptotic theory describing the behaviour of the slowly-varying part of the solution, i.e. the function V.

3.1.1 Main results of the asymptotic theory for high-frequency scattering problems

The asymptotic behaviour of the function V, for large k, in (3.2) is different in the illuminated, transition and shadow parts of the boundary Γ. There are three theorems, presented below, that describe these behaviours: the first theorem is a commonly well known "geometrical optics" expansion of V for s in the illuminated part of the boundary. The second theorem that describes the uniform asymptotic behaviour of V for s in the transition regions (also known as "Fock" domains) is also a well known expansion. The third theorem describes the behaviour of the solution v on the shadow part of the boundary [46].

Theorem 3.1 (**Illuminated part of the boundary**, Λ_2)**.** *For any small and positive* Δ*, for all* $s \in (t_1 + \Delta, t_2 - \Delta)$ *and all non-negative integers* N*:*

$$V(s, k) = \sum_{j=0}^{N} k^{-j} d_j(s) + r_N(s, k), \tag{3.3}$$

where $d_j(s) \in C^\infty(t_1+\Delta, t_2-\Delta)$*, and for all* $n \geq 0$*,*

$$|D_s^n r_N(s,k)| \leq c_{N,n}(1+k)^{-N-1}, \tag{3.4}$$

where $c_{N,n} = c_{N,n}(\Delta)$ *is a constant independent of* k*.*

Theorem 3.2 (**Transition part of the boundary**, Λ_1, Λ_3)**.** *There exist* $\Delta > 0$ *such that for* $s \in I_\Delta := (t_1-\Delta, t_1+\Delta) \cup (t_2-\Delta, t_2+\Delta)$*:*

$$V(s,k) = \sum_{l=0}^{L}\sum_{m=0}^{M} k^{-1/3-2l/3-m} b_{l,m}(s)\Psi^{(l)}\left(k^{1/3}Z(s)\right) + R_{L,M}(s,k)\ , \tag{3.5}$$

where the remainder satisfies

$$|D_s^n R_{L,M}(s,k)| \leq C_{L,M,n}(1+k)^{\mu+n/3}, \tag{3.6}$$

where $\mu := -\min\{2(L+1)/3, (M+1)\}$ *and* $C_{L,M,n}$ *are independent of* k*. Here* $b_{l,m}$ *are* $\mathcal{C}^\infty$ *complex-valued functions on* I_Δ*,* Z *is a* $\mathcal{C}^\infty$ *real-valued function on* I_Δ*, with simple zeros at* t_1 *and* t_2*, positive-valued in* $(t_1,t_2)\cap I_\Delta$ *and negative-valued in* $(t_2-2\pi, t_1)\cap I_\Delta$*, and* $\Psi : \mathbb{C} \to \mathbb{C}$*, the so-called "Fock's integral", is defined as follows:*

$$\Psi(\tau) := \exp\left(-i\tau^3/3\right)\int_c \frac{e^{-iz\tau}}{\mathrm{Ai}(e^{2\pi i/3}z)}dz \tag{3.7}$$

where Ai *is the Airy function and the contour of integration* c *is as in Figure 3.1.*

The function $\Psi^{(l)}$ denotes the l-th derivative of Ψ defined in (3.7) and this integral converges exponentially for any small positive θ due to asymptotic properties of Airy function Ai described later in (3.31). Further, there exist $\boldsymbol{\gamma} > 0$ and $c_0 \neq 0$ such that for any $l \in \mathbb{N}\cup\{0\}$,

$$D_\tau^{(l)}\Psi(\tau) = c_0 D_\tau^{(l)}\left\{\exp(-i\tau^3/3 - i\tau\alpha_1)\right\}\left(1+O(\exp(-|\tau|\boldsymbol{\gamma}))\right), \quad \text{as} \quad \tau \to -\infty, \tag{3.8}$$

where $\alpha_1 = \exp(-2i\pi/3)a_1$, where $a_1 < 0$ is the rightmost root of the Airy function Ai, see Section 3.2.5. From (3.8), we deduce that Ψ as well as its derivatives decay exponentially as $\tau \to -\infty$.

Theorem 3.3 (**Shadow part of the boundary**, Λ_4)**.** *For any small* $\Delta > 0$*, there exist positive constants* δ*,* C *and* c *such that for* $s \in S_\Delta := [0, t_1-\Delta] \cup [t_2+\Delta, 2\pi]$*,*

$$\|v\|_{L_2(S_\Delta)} \leq C\exp\left(-ck^\delta\right). \tag{3.9}$$

When designing numerical methods, in order to reflect the structure of the solution, we introduce a partition of unity $\chi_j \in L_\infty[0,2\pi]$, possibly depending on k, defined in (2.35).

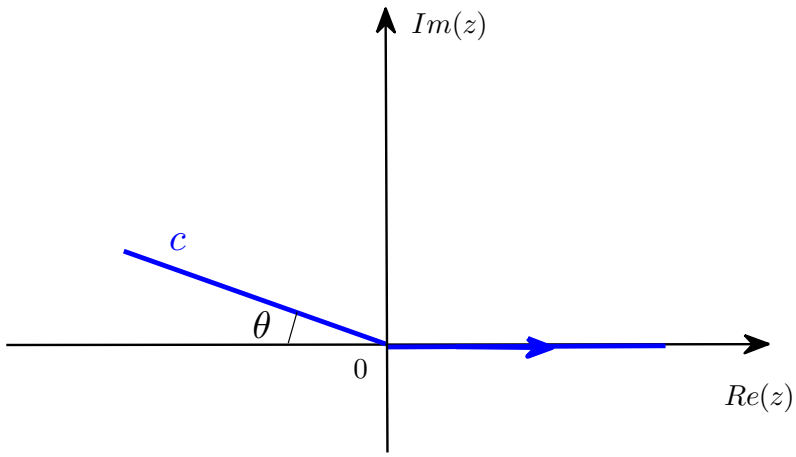


Figure 3.1: *Contour of integration c.*

Then we write the solution in each zone of the boundary as follows:

$$\chi_j(s)v(s) = kV_j(s,k)\exp(ik\boldsymbol{\gamma}(s)\cdot\boldsymbol{a}), \tag{3.10}$$

where $V_j(s,k) := \chi_j(s)V(s,k)$, $j = 1,2,3,4$, denotes the unknown "amplitudes" of v in the illuminated ($l = 2$), transition ($l = 1,3$), and shadow ($l = 4$) parts of the boundary. The solution v is known to be exponentially small in the shadow region, see Theorem 3.3. Thus, we approximate the solution in the shadow part of the boundary Λ_4 by zero.

Therefore, we can write

$$v(s) = k\sum_{j=1}^{3} V_j(s,k)\exp(ik\boldsymbol{\gamma}(s)\cdot\boldsymbol{a}) + \chi_4(s)v(s).$$

Remark 3.4. *The asymptotic expansion for $V_j(s,k)$, $j = 1,2,3,4$, in the transition zones, given in (3.5), adopts the form (3.3) by replacing the coefficients $b_{l,m}(s)$ by an appropriate smooth, 2π-periodic extensions and replacing $\Psi^{(l)}\left(k^{1/3}Z(s)\right)$ in (3.5) by its asymptotics as $k\to\infty$. Then, the global asymptotic behaviour of $V(s,k)$, for $s\in[0,2\pi]$ follows as we state in the corollary below.*

Corollary 3.5. *[40, Corollary 5.3] In the same notation as Theorem 3.2, the functions $b_{l,m}(s)$ can be non-uniquely extended to 2π-periodic C^∞ functions such that, for all L, $M\in\mathbb{N}\cup 0$, the decomposition*

$$V(s,k) = \sum_{l=0}^{L}\sum_{m=0}^{M} k^{-1/3-2l/3-m} b_{l,m}(s)\Psi^{(l)}\left(k^{1/3}Z(s)\right) + R_{L,M}(s,k), \tag{3.11}$$

holds for all $s \in [0, 2\pi]$*, with remainder term satisfying for all* $n \in \mathbb{N} \cup 0$*,*

$$|D_s^n R_{L,M}(s,k)| \leq C_{L,M,n}(1+k)^{\mu+n/3}, \quad \text{where} \quad \mu := -\min\{2(L+1)/3, (M+1)\}, \tag{3.12}$$

where constants $C_{L,M,n}$ *are independent of* k*.*

The corollary then leads to the following k-dependent estimates of the derivatives of the function V.

Theorem 3.6. *[25, Theorem 2.2] For all* $n \in \mathbb{N} \cup 0$ *there exist constants* $C_n > 0$ *independent of* k *and* s *such that for all* k *sufficiently large,*

$$|D_s^n V(s,k)| \leq C_n \begin{cases} 1, & n = 0, 1 \\ k^{-1}(k^{1/3} + |w(s)|)^{-n-2}, & n \geq 2, \end{cases}, \tag{3.13}$$

where $w(s) = (s - t_1)(t_2 - s)$*. The estimates (3.13) are uniform in* $s \in [0, 2\pi]$*.*

The estimates (3.13) indicate that the derivatives of V remain bounded as $k \to \infty$ for s in the illuminated zone bounded away from t_1 and t_2 (so that $|w(s)|$ is bounded away from zero). On the other hand, if s is in the region of width $O(k^{-1/3})$ around the transition points, then $|w(s)| \leq ck^{-1/3}$, with c is independent of k. Hence $|D_s^n V(s,k)|$ may blow up with $O(k^{(n-1)/3})$.

In each of the zones Λ_1, Λ_2, and Λ_3, the error in best approximation by polynomials can be estimated using (3.13).

Then the error bound for the semi-discrete Galerkin scheme for the discretization of the boundary integral equations such as the standard combined potential boundary integral equation (1.10).

3.1.2 Outline of the chapter

In Section 3.2, we describe the properties of Hankel and Airy functions. Asymptotic expansions of Hankel function in Debye's, Olver's and Cherry's form will also be presented as needed for high frequency asymptotic analysis of exact solution of scattering by a circle. In Section 3.3, we prove Theorems 3.1 and 3.2 for the model problem of scattering by a circle essentially following ideas from [5]. In Section 3.4 we outline the Model Problem Method [5] for constructing asymptotic expansion of the solution to the scattering problem for cases of general smooth and convex domains.

3.2 Preliminaries

In this section, we introduce Bessel functions and present their known asymptotic behaviour for certain large parameters they depend upon. We also introduce Airy functions and their basic properties and asymptotic expansions.

3.2.1 Bessel functions

Bessel functions are solutions of the Bessel's differential equation:

$$z^2\frac{d^2w}{dz^2} + z\frac{dw}{dz} + (z^2 - \nu^2)w = 0. \tag{3.14}$$

Here ν is generally a complex "order" parameter, $\nu \in \mathbb{C}$. For any complex ν, Bessel equation has two independent solutions: the Bessel function of the first kind $J_\nu(z)$ and the Bessel function of the second kind, $Y_\nu(z)$ of order ν, also known as Neumann function. When ν is not an integer, the Bessel function $J_\nu(z)$ can be represented as an absolutely convergent infinite series solution of (3.14) as follows, e.g. [16, Section I.3.1],

$$J_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{m=0}^{\infty} \frac{\left(-\frac{1}{2}z^2\right)^m}{m!\Gamma(\nu+m+1)},$$

where Γ is a Gamma function [3, Chapter 6]. For $\nu = n \in \mathbb{Z}$,

$$J_n(z) := \lim_{\nu\to n} J_\nu(z).$$

The function $J_\nu(z)$ is an entire function of ν for z in a complex plane cut along the negative real axis from 0 to $-\infty$, [16]. Moreover, for fixed ν, the function $z^{-\nu}J_\nu(z)$ is an entire function of $z \in \mathbb{C} \setminus (-\infty, 0)$, [16].

The Neumann function $Y_\nu(z)$ is defined as follows for all ν,

$$\begin{aligned} Y_\nu(z) &= \frac{J_\nu(z)\cos\nu\pi - J_{-\nu}(z)}{\sin\nu\pi}, \qquad \nu \neq n \in \mathbb{N}, \\ Y_\nu(z) &= \lim_{\nu\to n} Y_\nu(z), \qquad n \in \mathbb{N}. \end{aligned}$$

In particular, for $\nu = n \in \mathbb{N}$,

$$\begin{aligned} Y_n(z) = &-\frac{1}{\pi}\left(\frac{1}{2}z\right)^{-n} \sum_{m=0}^{\nu-1} \frac{(n-m-1)!}{m!}\left(\frac{1}{2}z\right)^{2m} + \frac{2}{\pi}\ln\left(\frac{1}{2}z\right) J_n(z) \\ &-\frac{1}{\pi}\left(\frac{1}{2}z\right)^{n} \sum_{m=0}^{\infty} \frac{\eta(m+1)+\eta(n+m+1)}{(-1)^m m!(n+m)!}\left(\frac{1}{2}z\right)^{2m}, \end{aligned}$$

with

$$\eta(1) = -E, \quad \eta(n) = -E + \sum_{m=1}^{n-1} \frac{1}{m},$$

where $E = 0.5772\ldots$ is the Euler's constant. The Neumann function $Y_\nu(z)$ is an entire function of ν for z in a complex plane cut along the negative real axis from 0 to $-\infty$, [16].

Other solutions of Bessel equation are Bessel functions of the third kind, also called Hankel functions, which are defined as a linear combination of the Bessel functions of the first and second kind as follows:

$$H_\nu^{(1)}(z) = J_\nu(z) + iY_\nu(z), \qquad H_\nu^{(2)}(z) = J_\nu(z) - iY_\nu(z).$$

Functions $H_\nu^{(1)}(z)$ and $H_\nu^{(2)}(z)$ are entire functions of their index $\nu \in \mathbb{C}$ for $z \in \mathbb{C}\backslash(-\infty, 0)$.

3.2.2 Basic asymptotic expansions of Bessel functions at large values of the order or the argument

Asymptotic expansion of Bessel functions at large values of the order

The following estimates are obtained from [16, (I.210) and (I.212)]. For a fixed z and $|\nu| \to \infty$ and $|\arg \nu| < \pi$,

$$J_\nu(z) = \sqrt{\frac{1}{2\pi\nu}} \left(\frac{ez}{2\nu}\right)^\nu \left[1 + O(\nu)^{-1}\right], \tag{3.15}$$

$$Y_\nu(z) = -\sqrt{\frac{2}{\pi\nu}} \left(\frac{ez}{2\nu}\right)^{-\nu} \left[1 + O(\nu)^{-1}\right]. \tag{3.16}$$

Asymptotic expansion of Bessel functions at large values of the argument

The following estimates are obtained from [3] formulae (9.2.5), (9.2.7), (9.2.11) and (9.2.13). For a fixed order $\nu \in \mathbb{C}$, as $|z| \to \infty$, Bessel and Hankel function's asymptotic expansions are

$$\begin{aligned} J_\nu(z) &= \sqrt{2/(\pi z)} \cos\left(z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi\right) \left(1 + O(|z|^{-1})\right), \\ J_\nu'(z) &= -\sqrt{2/(\pi z)} \sin\left(z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi\right) \left(1 + O(|z|^{-1})\right), \end{aligned} \tag{3.17}$$

for $|\arg z| < \pi$ and

$$\begin{aligned} H_\nu^{(1)}(z) &= \sqrt{2/(\pi z)} \exp\left\{i\left(z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi\right)\right\} \left(1 + O(|z|^{-1})\right), \\ H_\nu^{(1)\prime}(z) &= i\sqrt{2/(\pi z)} \exp\left\{i\left(z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi\right)\right\} \left(1 + O(|z|^{-1})\right), \end{aligned} \tag{3.18}$$

for $-\pi < \arg z < 2\pi$.

Let us introduce the function $W_\nu(z)$, also known as Wronskian, for $\nu \in \mathbb{C}$ and $z \in \mathbb{C}$ as follows,

$$W_\nu(z) := J_\nu^{'}(z)H_\nu^{(1)}(z) - J_\nu(z)H_\nu^{(1)'}(z). \tag{3.19}$$

Then the following property holds.

Proposition 3.7. *For all $\nu \in \mathbb{C}$, $z \in \mathbb{C}$,*

$$W_\nu(z) = -\frac{2i}{\pi z}. \tag{3.20}$$

Proof. We begin by introducing the following notation,

$$y_1(z) := J_\nu(z) \quad \text{and} \quad y_2(z) := H_\nu^{(1)}(z).$$

Then $y_1(z)$ and $y_2(z)$ solve the Bessel's differential equation (3.14), i.e.

$$z^2 y_j''(z) + z y_j'(z) + \left[z^2 - \nu^2\right] y_j(z) = 0, \qquad j = 1, 2.$$

Dividing through by z:

$$z y_j''(z) + y_j'(z) + \left[z - \frac{\nu^2}{z}\right] y_j(z) = 0, \qquad j = 1, 2.$$

Hence,

$$\left(z y_j'(z)\right)' + q_\nu(z) y_j(z) = 0, \quad \text{where} \quad q_\nu(z) = z - \frac{\nu^2}{z}. \tag{3.21}$$

Therefore,

$$\begin{aligned}
(zW_\nu(z))' &= \left(z y_1'(z) y_2(z) - z y_2'(z) y_1(z)\right)' \\
&= \left(z y_1'(z)\right)' y_2(z) + z y_1'(z) y_2'(z) - z y_1'(z) y_2'(z) - \left(z y_2'(z)\right)' y_1(z) \\
&= \left(z y_1'(z)\right)' y_2(z) - \left(z y_2'(z)\right)' y_1(z) \\
&= -q_\nu(z) y_1(z) y_2(z) - \left(-q_\nu(z) y_2(z) y_1(z)\right) \\
&= 0.
\end{aligned}$$

Hence

$$z W_\nu(z) = \text{const}, \quad \text{for all} \quad z \in \mathbb{C}. \tag{3.22}$$

In order to find the value of the constant, we use the asymptotic expansions (3.17) and (3.18) to determine the limit value of $zW_\nu(z)$ for $\operatorname{Im} z = 0$ and $|z| \to \infty$. Denoting

$\phi = z - \frac{1}{2}\nu\pi - \frac{1}{4}\pi$, we obtain

$$\begin{aligned} W_\nu(z) &= \frac{2}{\pi z}\left[-\sin(\phi)\exp(i\phi) - i\cos(\phi)\exp(i\phi)\right]\left(1 + O(|z|^{-1})\right) \\ &= -\frac{2i}{\pi z}\left[\exp(i\phi)\exp(-i\phi)+\right]\left(1 + O(|z|^{-1})\right) \\ &= -\frac{2i}{\pi z}\left(1 + O(|z|^{-1})\right). \end{aligned}$$

Then,

$$\lim_{|z|\to\infty} zW_\nu(z) = -\frac{2i}{\pi}.$$

Hence, from equation (3.22), we deduce that $zW_\nu(z) = -2i/\pi$ for all z. □

3.2.3 Airy functions and their basic asymptotic expansions

The Airy functions $\mathrm{Ai}(z)$ and $\mathrm{Bi}(z)$ are linearly independent solutions to the Airy equation

$$w''(z) - zw(z) = 0. \tag{3.23}$$

All solutions of Airy equation are entire functions of z and can be expanded in powers of z in series that converge for any $z \in \mathbb{C}$ [5] as we show in the following proposition.

Proposition 3.8. *Any solution of the Airy equation takes the form:*

$$\begin{aligned} w(z) &= w(0)\left(1 + \frac{z^3}{2\cdot 3} + \frac{z^6}{2\cdot 3\cdot 5\cdot 6} + \dots\right) \\ &+ w'(0)\left(z + \frac{z^4}{3\cdot 4} + \frac{z^7}{3\cdot 4\cdot 6\cdot 7} + \dots\right). \end{aligned} \tag{3.24}$$

Proof. Substitute the power series

$$w(z) = \sum_{j=0}^{\infty} c_j z^j,$$

into the Airy equation to obtain:

$$\begin{aligned} w''(z) - zw(z) &:= \sum_{j=2}^{\infty} j(j-1)c_j z^j - \sum_{j=0}^{\infty} c_j z^{j+1} \\ &= c_2 + \sum_{j=0}^{\infty}(j+3)(j+2)c_{j+3}z^{j+1} - \sum_{j=0}^{\infty} c_j z^{j+1} \\ &= c_2 + \sum_{j=0}^{\infty}\left[(j+3)(j+2)c_{j+3} - c_j\right]z^{j+1}. \end{aligned} \tag{3.25}$$

Setting the coefficients of t^j, $j = 0, 1, 2, \ldots$ to zero in (3.25), we deduce:

$$
\begin{aligned}
& & c_0 = w(0), \quad c_1 = w'(0), \quad c_2 = 0, \\
& j = 1 & 3 \cdot 2\ c_3 = c_0 \\
& j = 2 & 4 \cdot 3\ c_4 = c_1 \\
& j = 3 & 5 \cdot 4\ c_5 = c_2 = 0 \\
& & \ldots \\
& j = n & (n+3)(n+2)c_{n+3} - c_n = 0, \quad n \geq 3.
\end{aligned}
$$

The two series in (3.24) both converge absolutely for any z. Moreover, the two solutions are linearly independent. Hence any other solution is their linear combination. Thus, (3.24) follows. □

Proposition 3.9. *[16, Formulae (I.256) and (I.257)] The functions* $\mathrm{Ai}(z)$ *and* $\mathrm{Bi}(z)$*, represented in an integral form as follows: for* $z \in \mathbb{C}$*,*

$$\mathrm{Ai}(z) = \frac{1}{2\pi i} \int_{\infty\ e^{-i\pi/3}}^{\infty\ e^{i\pi/3}} \exp\left(\frac{x^3}{3} - zx\right) dx, \tag{3.26}$$

and

$$\mathrm{Bi}(z) = \frac{1}{2\pi} \left(\int_{-\infty}^{\infty\ e^{i\pi/3}} \exp\left(\frac{x^3}{3} - zx\right) dx + \int_{-\infty}^{\infty\ e^{-i\pi/3}} \exp\left(\frac{x^3}{3} - zx\right) dx \right), \tag{3.27}$$

are solutions of Airy equation (3.23). The contour of integration in (3.26) is illustrated in Figure 3.2 on the right plot in red.

Proof. Consider an entire function

$$s(x) := \exp\left(\frac{1}{3}x^3\right).$$

The function s is exponentially small as $x \to \infty e^{\pi i/3}$ (i.e. $x = te^{\pi i/3}$, $t \to +\infty$). To see this we observe that for $x \in \mathbb{C}$, $x^3 = |x|^3 \exp(i3\theta)$, where $\theta = \arg(x)$. Therefore, for θ satisfying the inequality

$$\cos(3\theta) < 0,$$

the following holds: $s(x) \to 0$ as $|x| \to \infty$. In Figure 3.2, the corresponding subintervals in $[0, 2\pi]$ for θ where $\cos\theta < 0$ are shaded in the left plot. Also in Figure 3.2, the domains in the complex plane for x where $s(x) \to 0$ as $x \to \infty$ are illustrated on the right plot as

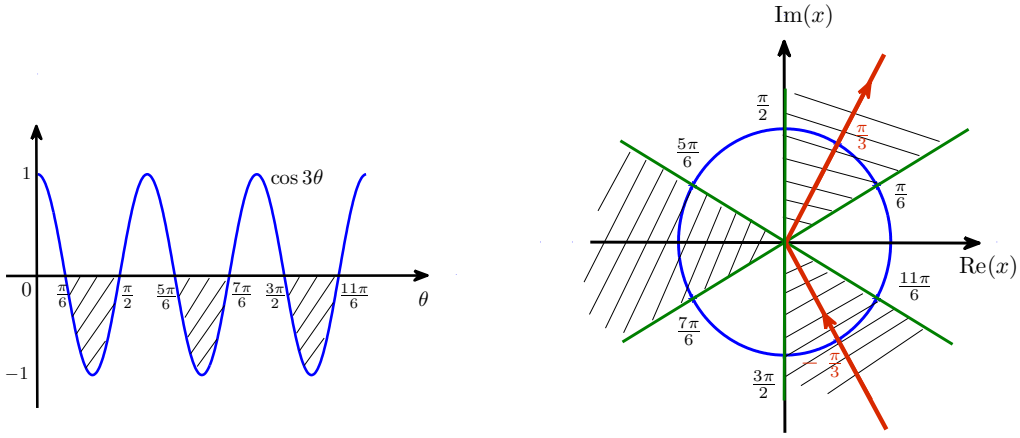


Figure 3.2: *The plot of* $\cos 3\theta$ *is illustrated on the left with the shaded regions indicating where the values of* $\cos 3\theta$ *are negative. The function* $s(x)$ *decays exponentially in the shaded regions illustrated on the right. In particular,* $s(x)$ *decays exponentially as* $|x|$ *tends to infinity along a contour plotted in red.*

shaded domains. In particular,

$$\lim_{x\to\infty e^{\pm\pi i/3}} s(x) = 0. \tag{3.28}$$

Differentiating Airy function $\mathrm{Ai}(z)$ twice, we obtain,

$$\mathrm{Ai}''(z) = \frac{1}{2\pi i}\int_{\infty\ e^{-i\pi/3}}^{\infty\ e^{i\pi/3}} x^2 \exp\left(\frac{x^3}{3} - zx\right) dx. \tag{3.29}$$

On the other hand, integrating $z\mathrm{Ai}(z)$ by parts, we deduce,

$$\begin{aligned}
z\mathrm{Ai}(z) &:= \int_{\infty\ e^{-i\pi/3}}^{\infty\ e^{i\pi/3}} z\exp\left(\frac{x^3}{3} - zx\right) dx \\
&= -\int_{\infty\ e^{-i\pi/3}}^{\infty\ e^{i\pi/3}} \exp\left(\frac{x^3}{3}\right)\frac{d}{dx}\left(\exp(-zx)\right) dx \\
&= -\left[\exp\left(\frac{x^3}{3}\right)\exp(-zx)\right]_{x=\infty\ e^{-i\pi/3}}^{x=\infty\ e^{i\pi/3}} + \int_{\infty\ e^{-i\pi/3}}^{\infty\ e^{i\pi/3}} x^2\exp\left(\frac{x^3}{3}\right)\exp(-zx) dx \\
&= \mathrm{Ai}''(z),
\end{aligned}$$

due to (3.28) and (3.29). Hence $\mathrm{Ai}(z)$ defined in (3.26) is the solution of Airy equation (3.23). Similarly we can prove that $\mathrm{Bi}(z)$ is also the solution of Airy equation (3.23). □

Similarly to Proposition 3.7, we can show that Wronskian for Airy functions is constant for all z.

Proposition 3.10. *For any z,*

$$W(z) := w_1(z)\frac{dv}{dz}(z) - v(z)\frac{dw_1}{dz}(z) = i, \tag{3.30}$$

where $v(z) = \mathrm{Ai}(z)$ and $w_1(z) = 2e^{2\pi i/3}\mathrm{Ai}(ze^{2\pi i/3})$. The notation for the functions v and w_1 is consistent with notation in Babich and Buldyrev [5]; notice that $w_1(z)$ is also a solution of (3.23) which can be verified by a direct substitution.

Proof. Differentiating the Wronskian we obtain:

$$\begin{aligned} W'(z) &= \frac{dw_1}{dz}\frac{dv}{dz} + w_1\frac{d^2v}{dz^2} - \frac{dv}{dz}\frac{dw_1}{dz} - v\frac{d^2w_1}{dz^2} \\ &= w_1\frac{d^2v}{dz^2} - v\frac{d^2w_1}{dz^2} \quad \text{applying (3.23)} \\ &= zw_1v - zvw_1 = 0. \end{aligned}$$

Therefore, for all z, $W(z) = C$, where C is a constant. In order to find C, we evaluate $W(z)$ at a point $z = 0$ using known values of v and w_1 at 0 which can be found from (3.26) and are given for example in [3, (10.4.4) and (10.4.5)]:

$$\begin{aligned} w_1(0) &= \frac{2\sqrt{\pi}e^{i\pi/6}}{3^{2/3}\Gamma(2/3)}, \qquad & w_1'(0) &= \frac{2\sqrt{\pi}e^{-i\pi/6}}{3^{4/3}\Gamma(4/3)}, \\ v(0) &= \frac{\sqrt{\pi}}{3^{2/3}\Gamma(2/3)}, \qquad & v'(0) &= -\frac{\sqrt{\pi}}{3^{4/3}\Gamma(4/3)}. \end{aligned}$$

Therefore,

$$W(0) = -\frac{2\pi}{3^2\Gamma\left(\frac{2}{3}\right)\Gamma\left(\frac{4}{3}\right)}\left(e^{i\pi/6} + e^{-i\pi/6}\right) = i,$$

where $\Gamma(a)$ is the Gamma function. Here we have used the Triplication formula for the Gamma functions [3, (6.1.19)]:

$$\Gamma(2) = \frac{3^{3/2}}{2\pi}\Gamma\left(\frac{2}{3}\right)\Gamma(1)\Gamma\left(\frac{4}{3}\right),$$

and that $\Gamma(1) = 1$ and $\Gamma(2) = 1$. Hence (3.30) follows. □

Asymptotic expansion of Airy function at large values of the argument [5]

The asymptotic expansion of Airy function $\mathrm{Ai}(z)$ for complex z as $|z| \to \infty$ can be obtained from their integral representation via the method of steepest descent cf [5, Appendix A.1,

(A.1.2.5), (A.1.3.1)]. As a result, for $-2\pi/3 \leq \arg z \leq 2\pi/3$, and $|z| \to \infty$,

$$\mathrm{Ai}(z) = \frac{1}{\pi}\frac{\exp\left(-\frac{2z^{3/2}}{3}\right)}{2z^{1/4}}\sum_{n=0}^{\infty}\frac{\Gamma\left(3n+\frac{1}{2}\right)}{(2n)!}\left(-9z^{3/2}\right)^{-n}. \tag{3.31}$$

On the other hand, for $2\pi/3 \leq \arg z \leq 4\pi/3$, and $|z| \to \infty$,

$$\begin{aligned}\mathrm{Ai}(z) := &\frac{1}{\pi}\frac{\exp\left(-\frac{2z^{3/2}}{3}\right)}{2z^{1/4}}\sum_{n=0}^{\infty}\frac{\Gamma\left(3n+\frac{1}{2}\right)}{(2n)!}\left(-9z^{3/2}\right)^{-n}\\ &-\frac{1}{\pi}\frac{\exp\left(\frac{2z^{3/2}}{3}\right)}{2iz^{1/4}}\sum_{n=0}^{\infty}\frac{\Gamma\left(3n+\frac{1}{2}\right)}{(2n)!}\left(9z^{3/2}\right)^{-n}.\end{aligned} \tag{3.32}$$

3.2.4 Debye's, Olver's and Cherry's asymptotic expansions of Hankel function

The asymptotic expansion of the Hankel function for large values of the argument as well as the order are given by Debye's [36], Olver's [80] and Cherry's [29] asymptotic expansion. We will use these expansions later in the chapter.

Debye's asymptotic expansion of Hankel function

Debye's asymptotic representation describes the behaviour of Hankel function $H^{(1)}_{N\xi}(N)$ for large values of N. This a WKB-type expansion for Bessel's ODE (3.14). We denote by S_1 and S_2 the following series, e.g. [4],

$$\begin{aligned}S_1 := &\sqrt{\frac{2}{\pi N}}\frac{1}{(\xi^2-1)^{1/4}}\exp\left\{N\left[-\xi\ln\left(\xi+\sqrt{\xi^2-1}\right)\right.\right.\\ &\left.\left.+\sqrt{\xi^2-1}\right]\right\}\left(1+\sum_{m=1}^{\infty}\frac{A_m(\xi)}{N^m}\right),\\ S_2 := &\sqrt{\frac{2}{\pi N}}\frac{i}{(\xi^2-1)^{1/4}}\exp\left\{N\left[-\xi\ln\left(\xi+\sqrt{\xi^2-1}\right)\right.\right.\\ &\left.\left.-\sqrt{\xi^2-1}\right]\right\}\left(1+\sum_{m=1}^{\infty}\frac{B_m(\xi)}{N^m}\right),\end{aligned} \tag{3.33}$$

where function $A_m(\xi)$ and $B_m(\xi)$ are regular functions of ξ in the complex plane with cuts along the real line from $-\infty$ to -1 and from 1 to $+\infty$. Moreover, for $|\xi| \to +\infty$, $A_m(\xi) = O(|\xi|^{-m})$ and $B_m(\xi) = O(|\xi|^{-m})$. Functions A_m and B_m are known and can be found for example in [3].

In Figure 3.3 we illustrate how series S_1 and S_2 in (3.33) represent $H^{(1)}_{N\xi}(N)$. In small circular neighborhoods of points $\xi = \pm 1$, the Debye's asymptotic representation of Hankel

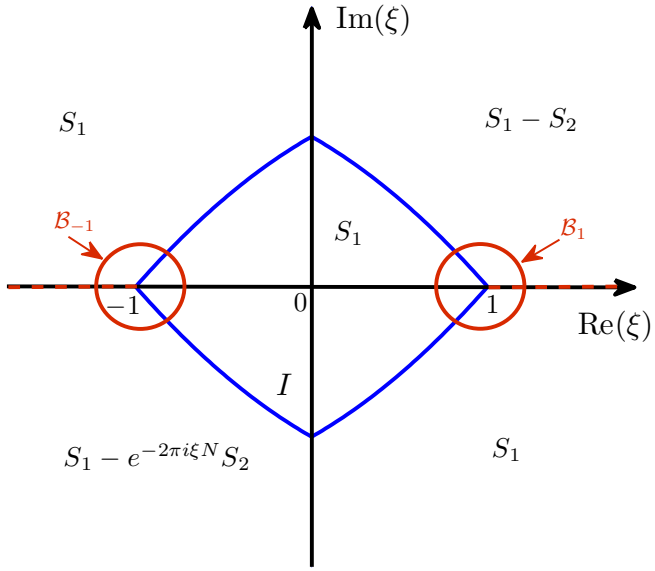


Figure 3.3: *The plot illustrates the domains where Hankel function is represented by series S_1 and S_2 defiend in (3.33). We denote the circular neighborhoods of points $\xi = \pm 1$, as $\mathcal{B}_{\pm 1}$, where the Debye's representation of Hankel function is not valid.*

function is not valid. Here, the Olver's or Cherry's expansion can be used instead, see below. We denote the circular neighborhoods of points $\xi = \pm 1$, as $\mathcal{B}_{\pm 1}$.

In Figure 3.3, we denote the region inside the rhombus-like area (plotted in blue) by I. The boundary of I in the first quadrant is a curve described by the following equation:

$$\pi = \arg\{\xi \log\left(\xi + \sqrt{\xi^2 - 1}\right) - \sqrt{\xi^2 - 1}\}.$$

The boundaries of I in the second, third and fourth quadrants are obtained by reflection with respect to the coordinate axes.

In the domain I, as well as in the fourth quadrant (where $\operatorname{Im} \xi < 0$ and $\operatorname{Re} \xi > 0$), it is often more convenient to use the following formula equivalent to (3.33),

$$H^{(1)}_{N\xi}(N) := \sqrt{\frac{2}{\pi N}} \frac{e^{-\pi i/4}}{(\xi^2 - 1)^{1/4}} \exp\left\{Ni\left[-\xi \cos^{-1}\xi + \sqrt{1 - \xi^2}\right]\right\} \times \\ \times \left(1 + \sum_{m=1}^{M} \frac{A_m(\xi)}{N^m} + R_{M,N}(\xi)\right). \tag{3.34}$$

Notice that then $0 < \operatorname{Re}(\cos^{-1}(\xi)) < \pi$.

In the domain I, and in the fourth quadrant (where $\operatorname{Im}(\xi) < 0$ and $\operatorname{Re}(\xi) > 0$), excluding

$\mathcal{B}_{\pm 1}$, the remainder term $R_{M,N}(\xi)$ in (3.34) is uniform and is of order

$$R_{M,N}(\xi) = O(N^{-M-1}), \quad \xi \in I \setminus \mathcal{B}_{\pm 1}.$$

Outside of the domain I, the remainder term is of order $R_{M,N}(\xi) = O(N^{-M-1}|\xi|^{-M-1})$.

Olvers's asymptotic expansion of Hankel function

In $\mathcal{B}_{\pm 1}$, the asymptotic expansion of Hankel function $H^{(1)}_{N\xi}(N)$ for large N is more complicated (the asymptotic expansion corresponds to the *turning points* for the WKB asymptotics). Here we use Olver's uniform asymptotic representation, [4]

$$\begin{aligned} H^{(1)}_{N\xi}(N) = & \frac{\mathrm{Ai}\left(e^{2\pi i/3} N^{2/3}\zeta(\xi)\right)}{N^{1/3}} \left(\sum_{m=0}^{M-1} \frac{C_m(\xi)}{N^{2m}} + R^{I}_{N,M}(\xi) \right) \\ & + \frac{\mathrm{Ai}'\left(e^{2\pi i/3} N^{2/3}\zeta(\xi)\right)}{N^{5/3}} \left(\sum_{m=0}^{M-1} \frac{D_m(\xi)}{N^{2m}} + R^{II}_{N,M}(\xi) \right), \end{aligned} \tag{3.35}$$

where $C_m(\xi)$ and $D_m(\xi)$ are regular functions in $\mathcal{B}_{\pm 1}$, and $C_0(\xi) \neq 0$. Functions C_m and D_m are known and can be found for example in [3]. The function $\zeta(\xi)$ is defined as follows,

$$\zeta(\xi) = \left(\frac{3}{2}\right)^{2/3} \left(\xi \log\left(\xi + \sqrt{\xi^2 - 1}\right) - \sqrt{\xi^2 - 1} \right)^{2/3}. \tag{3.36}$$

The function $\zeta(\xi)$ is regular at $\xi = 1$ because, by direct inspection, (3.36) admits a *regular* Taylor expansion at $\xi = 1$.

The remainder terms $R^{I}_{N,M}(\xi)$ and $R^{II}_{N,M}(\xi)$ are uniform in $\mathcal{B}_{\pm 1}$ and are of order:

$$R^{I}_{N,M}(\xi) = O\left(\frac{1}{N^{2M}}\right), \qquad R^{II}_{N,M}(\xi) = O\left(\frac{1}{N^{2M}}\right). \tag{3.37}$$

It will also be useful to write down the Olver's expansion for $H^{(1)}_{\nu}(\nu t)$ for $0 \leq \arg \nu \leq \pi$, $|\arg t| < \pi$, $|t-1| \leq \text{const} < 1$ and $|\nu| \to \infty$, [4]

$$\begin{aligned} H^{(1)}_{\nu}(\nu t) = & \frac{\mathrm{Ai}\left(e^{2\pi i/3} \nu^{2/3}\tilde{\zeta}(t)\right)}{\nu^{1/3} t^{1/3}} \left(\sum_{m=0}^{M-1} \frac{C_m(1/t)}{(\nu t)^{2m}} + O\left(\frac{1}{\nu^{2M}}\right) \right) + \\ & + \frac{\mathrm{Ai}'\left(e^{2\pi i/3} \nu^{2/3}\tilde{\zeta}(t)\right)}{(\nu t)^{5/3}} \left(\sum_{m=0}^{M-1} \frac{D_m(1/t)}{(\nu t)^{2m}} + O\left(\frac{1}{\nu^{2M}}\right) \right), \end{aligned} \tag{3.38}$$

where

$$\tilde{\zeta}(t) := t^{2/3} \zeta\left(\frac{1}{t}\right). \tag{3.39}$$

In (3.38), the functions $C_m(1/t)$ and $D_m(1/t)$ are regular functions in the circular neigh-

borhood of $t = 1$.

Cherry's form of the uniform asymptotic expansion

Uniform asymptotic expansion of the Hankel functions were derived by Cherry in 1950 [29] and then, in a different form, by Olver in 1954 [82]. The Cherry's asymptotic representation of Hankel function is given as follows, see [5, Section 7.2, (7.2.4) and (13.1.1)] for $|\arg \nu| \leq \pi/2$, $0 < \varepsilon < |t| < E < \infty$, $|\arg t| \leq \pi/2$, ($\varepsilon > 0$ and $E > 0$ are fixed constants), for $|\nu t| \to \infty$,

$$H_\nu^{(1)}(\nu t) = e^{i\pi/3}(\nu t)^{1/3}\left(\frac{T_\nu(t)}{1-t^2}\right)^{1/4} C_0(\zeta(t))\mathrm{A}_+\left(T_\nu(t) + \sum_{m=1}^{M}\frac{p_m(z(t))}{(\nu t)^{2m-2/3}} + R_M(\nu,t)\right), \tag{3.40}$$

where

$$\mathrm{A}_+(z) := \mathrm{Ai}\left(e^{\frac{2\pi i}{3}} z\right),$$

is the Airy function. In (3.40), $T_\nu(t)$ and $z(t)$ are defined as follows (see [5, (7.2.4)]),

$$T_\nu(t) := \nu^{2/3} z(t) = \left[\frac{3}{2} i\nu t \int_1^{1/t} \cos^{-1}(x)dx\right]^{2/3} = \nu^{2/3}\tilde{\zeta}(t) \tag{3.41}$$

$$= 2\left(\frac{\nu t}{2}\right)^{2/3}\left(\frac{1}{t}-1\right)\left[1 - \frac{1}{30}\left(\frac{1}{t}-1\right) + \ldots\right], \tag{3.42}$$

and where $p_1(z), p_2(z), \ldots, p_N(z)$ are functions that are analytic at $z = 0$, i.e. at $t = 1$. Hence $T_\nu(t)$ is analytic at $t = 1$. The function $\tilde{\zeta}$ in (3.41) also appears in (3.38).

The function $C_0(\zeta)$ in (3.40), defined as

$$C_0(\zeta) := \frac{d}{d\zeta}\left(\zeta + \sum_{m=1}^{M}\frac{p_m(\zeta)}{(\nu)^{2m}}\right), \tag{3.43}$$

is regular at $\zeta = 0$. Furthermore, by (3.42), the function $\left(T_\nu(t)/(1-t^2)\right)^{1/4}$ is analytic near $t = \pm 1$.

Remark 3.11. *The Olver's expansion (3.38) can be shown to be equivalent to Cherry's asymptotic expansion (3.40) via the decomposition of the Airy function A_+ in (3.40) into Taylor series about $\nu^{2/3}z$ and using Airy differential equation (3.23) for A_+. Thus, uniform error bounds for the Cherry's asymptotic expansion are implied by the uniform error bounds (3.37) for Olver's asymptotic expansion. In (3.40), the remainder term $R_M(\nu,t)$ is as a result bounded as follows,*

$$R_M(\nu,t) = O\left(\frac{1}{(\nu t)^{2(M+1)}}\right),$$

uniformly in $t \in \mathcal{B}_1$.

3.2.5 Zeroes of Airy and Hankel functions

Proposition 3.12. *Denote a_s, $s = 1, 2, 3, \ldots$ as zeroes of Airy function:*

$$\mathrm{Ai}(a_s) = 0.$$

Then a_s, $s = 1, 2, 3, \ldots$ are real and negative.

Proof. Let $-z_s$ denote a root of $\mathrm{Ai}(z)$, i.e. $\mathrm{Ai}(-z_s) = 0$. Denote $y_s(z) := \mathrm{Ai}(z - z_s)$, then,

$$-y_s''(z) + zy_s(z) = z_s y_s(z), \quad y_s(0) = 0, \quad \lim_{\substack{t\to\infty,\\ \mathrm{Im}(t)=0}} y_s(t) = 0, \tag{3.44}$$

i.e. the z_s can be considered as eigenvalues of the boundary value problem (3.44). Multiplying the equation (3.44) by the complex conjugate functions $\overline{y_s}$, we obtain:

$$-y_s''(z)\overline{y_s}(z) + z|y_s(z)|^2 = z_s|y_s(z)|^2. \tag{3.45}$$

Moreover,

$$\begin{aligned}\int_0^\infty y_s''(z)\overline{y_s}(z) &= \left[y_s'(z)\overline{y_s}(z)\right]_{z=0}^{z=\infty} - \int_0^\infty |y_s'(z)|^2 dz,\\ &= -\int_0^\infty |y_s'(z)|^2 dz,\end{aligned}$$

due to the boundary conditions in (3.44). Therefore, (3.45) yields,

$$\int_0^\infty \left(|y_s'(z)|^2 + z|y_s(z)|^2\right) dz = z_s \int_0^\infty |y_s(z)|^2 dz.$$

Hence, all z_s are real and positive. We conclude that all zeroes a_s of Airy function $\mathrm{Ai}(z)$ are real and negative since $a_s = -z_s$. □

The zeroes a_s of the Airy function $\mathrm{Ai}(z)$ have been thoroughly studied [82, pp. 364-367] and may be expressed asymptotically as

$$a_s = -\left[\frac{3}{2}\pi\left(s - \frac{1}{4}\right) + O\left(\frac{1}{s}\right)\right]^{2/3}, \tag{3.46}$$

for large values of s. Higher order terms of the asymptotic expansion for a_s as $s \to \infty$ can also be found in (10.4.94) and (10.4.105) in [3].

The ν-zeros of the Hankel function, $H_\nu^{(1)}(z)$, for large $|z|$, [32]

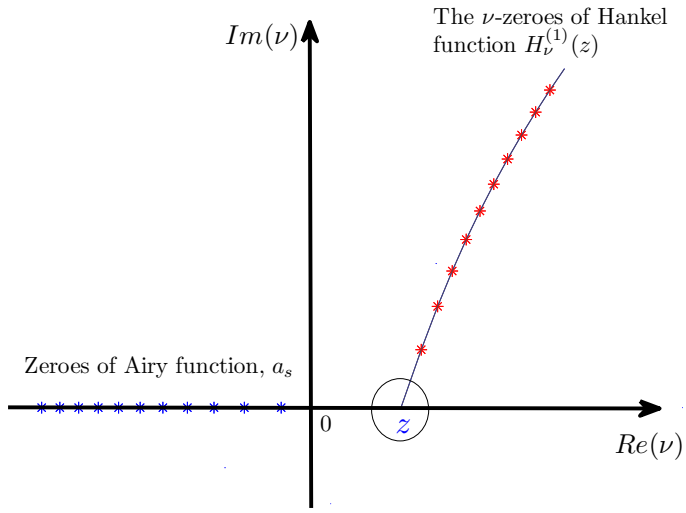


Figure 3.4: *Zeroes of Airy function,* Ai*, denoted as a_s, are plotted as blue starts located strictly on the negative real axis. The ν-zeroes (3.47) of the Hankel function $H_\nu^{(1)}(z)$ are plotted as red stars.*

Proposition 3.13. *Given z with $|arg\ z| < \pi$ and $|z| >> 1$, for $s = 1, 2, 3, \ldots$,*

$$H_{\nu_s}^{(1)}(z) = 0, \qquad \nu_s = z + 2^{-1/3} a_s e^{-2\pi i/3} z^{1/3} + O\left(|z|^{-1/3}\right), \tag{3.47}$$

where a_s, $s = 1, 2, 3, \ldots$, are zeroes of Airy function: $\mathrm{Ai}(a_s) = 0$.

Sketch of proof. See also [32, Section 4]. From (3.38), we deduce that asymptotically for large $|\nu|$, $H_\nu^{(1)}(\nu t) = 0$ when $\mathrm{Ai}\left(e^{2\pi i/3}\nu^{2/3}\tilde{\zeta}\right) = 0$. Thus,

$$a_s = e^{2\pi i/3}\nu^{2/3}\tilde{\zeta}.$$

Hence we obtain the asymptotic relation between zeroes of the Airy function $\mathrm{Ai}(a_s) = 0$ and ν-zeroes of Hankel function, $H_\nu^{(1)}(z)$,

$$\nu_s = \frac{z}{t(\zeta)} = e^{-\pi i}\left(\frac{a_s}{\tilde{\zeta}}\right)^{3/2}, \tag{3.48}$$

for $|\nu| \to \infty$, where ζ and t are related by, cf (3.35),

$$\zeta(t) = \left(\frac{3}{2}\right)^{2/3}\left(t\log\left(t + \sqrt{t^2 - 1}\right) - \sqrt{t^2 - 1}\right)^{3/2}. \tag{3.49}$$

Then, from (3.39), we deduce

$$\tilde{\zeta}(t) = \left(\frac{3}{2}\right)^{2/3}\left(\log\left(\frac{1}{t} + \sqrt{\frac{1}{t^2} - 1}\right) - \sqrt{\frac{1}{t^2} - 1}\right)^{3/2}. \tag{3.50}$$

From the second equality in (3.48), we deduce that the limiting case that gives rise to large values of ν_s is $|\zeta| \to 0$, which yields from (3.50),

$$t(\tilde{\zeta}) = 1 - 2^{-\frac{1}{3}}\tilde{\zeta} + \frac{3}{10}2^{-\frac{2}{3}}\tilde{\zeta}^2 + O\left(\tilde{\zeta}^3\right).$$

Thus, for $|z| \to \infty$,

$$\begin{aligned}
\nu_s = \frac{z}{t(\tilde{\zeta})} &= \frac{z}{1 - 2^{-\frac{1}{3}}\tilde{\zeta} + O\left(\tilde{\zeta}^2\right)} \\
&= z\left(1 + 2^{-\frac{1}{3}}\tilde{\zeta} + O\left(\tilde{\zeta}^2\right)\right) \\
&= z + 2^{-\frac{1}{3}}a_s e^{-\frac{2}{3}\pi i} z^{\frac{1}{3}} + O\left(z^{-\frac{1}{3}}\right)
\end{aligned}$$

Hence the result (3.47) follows. Notice that the Debye asymptotic expansions ensure the absence of other zeroes for $H_\nu^{(1)}(z) = 0$, for large $|\nu|$. □

In Figure 3.4, the zeroes a_s of Airy function and the corresponding leading order terms in (3.47) describing zeroes ν_s of the Hankel function, $H_\nu^{(1)}(z)$, are illustrated.

3.3 Model problem: scattering by a circle

In this section, we consider the model problem of high-frequency scattering by a circular domain in homogeneous media exploiting the ideas presented in [5]. The exact solution to this problem is well known and can be obtained by standard separation of variables method, see for example [95, 16]. The exact solution is given as a series involving Hankel and Bessel functions. The series can be re-formulated as a contour integral using the Cauchy Residue Theorem, the general procedure known as Watson's transformation [100]. Then the function $v(s,k) = (\partial u/\partial \boldsymbol{n})(\boldsymbol{\gamma}(s))$, can be simplified to be only expressed as a contour integral with the integrand containing the Hankel function. By replacing the Hankel function by its asymptotics and with an appropriate change of variables for the resulting integral, Theorem 3.1 and Theorem 3.2, as well as Theorem 3.3, can be proved for the case of a circle. We do this for Theorem 3.1 and Theorem 3.2 in Section 3.3.2 and Section 3.3.3 respectively.

In the following proposition, we present the exact solution of the scattering problem (2.1) - (2.3) in terms of the Bessel and Hankel functions.

Proposition 3.14. *In the case of scattering by a circle of radius a centered at the origin, the solution, u, of the scattering problem (2.1) - (2.3) with the direction of incidence along the vector* $\boldsymbol{a} = (1,0)^T$ *in polar coordinates* (r,s) *can be written as the following infinite sum [95, 16]:*

$$u(r,s) := (u^I + u^S)(r,s) := \sum_{n=0}^{\infty} 2\delta_n \left[J_n(kr) - \frac{H_n^{(1)}(kr)J_n(ka)}{H_n^{(1)}(ka)} \right] e^{in\pi/2} \cos ns, \tag{3.51}$$

where $\delta_n = 1,\ n > 0$ *and* $\delta_0 = 1/2$.

Proof. This is a classical result obtained by the standard separation of variables, see [95, 16]. Namely,

$$u^I(\boldsymbol{x}) = e^{ik\boldsymbol{x}\cdot\boldsymbol{a}} \quad \Longleftrightarrow \quad u^I(r,s) = e^{ikr\cos s},$$

is decomposed into Fourier series in s,

$$u^I(r,s) = \sum_{n=0}^{\infty} 2\delta_n J_n(kr) e^{in\pi/2} \cos ns.$$

The scattered wave is an outgoing wave. Thus, it can be represented as the following series:

$$u^S(r,s) = \sum_{n=0}^{\infty} 2\delta_n b_n H_n^{(1)}(kr) e^{in\pi/2} \cos ns,$$

for some coefficients b_n. Then we can write the total wavefield u as follows,

$$u(r,s) = \sum_{n=0}^{\infty} 2\delta_n \left[J_n(kr) + b_n H_n^{(1)}(kr) \right] e^{in\pi/2} \cos ns.$$

Then, coefficients b_n are found from the boundary conditions (2.2) for the total wavefield $b_n = -J_n(ka)/H_n^{(1)}(ka)$. □

This infinite sum can be expressed as a contour integral via the "reverse" application of the Cauchy Residue theorem.

Proposition 3.15. *The solution, u, of the scattering problem (2.1) - (2.3) in polar coordinates (r,s) can be written in an integral form,*

$$u(r,s) := ik \int_{\mathcal{L}} \left[J_{k\xi}(kr) - \frac{H_{k\xi}^{(1)}(kr) J_{k\xi}(ka)}{H_{k\xi}^{(1)}(ka)} \right] e^{-ik\xi\pi/2} \frac{\cos k\xi s}{\sin k\xi\pi} \mathrm{d}\xi, \tag{3.52}$$

using the analyticity of the Bessel and Hankel functions with respect to their index. Here, for $\beta = O\left(k^{-2/3-\gamma}\right)$ with small $\gamma > 0$, the contour of integration $\mathcal{L}$ consists of the lines $(+\infty - i\beta, -i\beta]$ and $[i\beta, i\beta + \infty)$ joined by the segment $[-i\beta, i\beta]$ in the complex ξ-plane, see Figure 3.5.

Remark 3.16. *In order to ensure that the contour of integrations $\mathfrak{L}$ does not contain zeroes of the Hankel function $H_{k\xi}^{(1)}(ka)$ in (3.52), we take, for large enough k,*

$$\beta \leq Ck^{-2/3-\gamma} \quad \textit{for some small} \;\; \gamma > 0. \tag{3.53}$$

This condition arises from relation (3.47) between zeroes, $\xi = h_s$, $s = 0, 1, 2, ...$ of the Hankel function $H_{k\xi}^{(1)}(ka)$ and zeros $z = a_s$ of the Airy functions $\mathrm{Ai}(z)$ for large values of k, see (3.47),

$$h_s = a + 2^{-1/3} a_s e^{-2\pi i/3} k^{-2/3} a^{1/3} + O\left(k^{-4/3} a^{-1/3}\right).$$

In Figure 3.5 we illustrate the contour of integration $\mathcal{L}$ in relation to the zeroes of Hankel function and zeroes of sine function where the integrand of (3.52) blows up. Provided β is chosen so that (3.53) is satisfied, the contour $\mathcal{L}$ does not contain inside its loop any zeroes of the Hankel function, and therefore the application of Cauchy Residue Theorem to (3.52) yields (3.51).

Remark 3.17. *The integration over the interval $[-i\beta, i\beta]$ has to be understood in the sense of Cauchy principal value:*

$$\int_{-i\beta}^{i\beta} = \lim_{\varepsilon \to 0} \left(\int_{-i\beta}^{-i\varepsilon} + \int_{i\varepsilon}^{i\beta} \right).$$

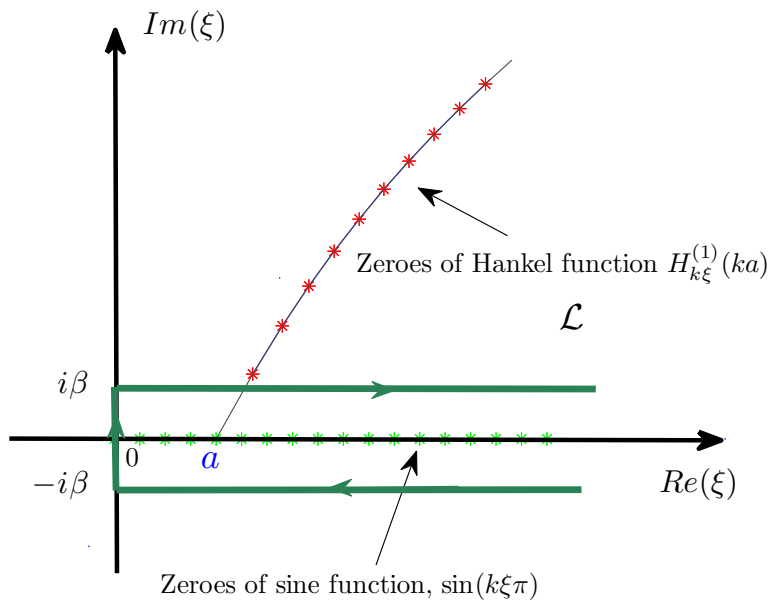


Figure 3.5: *The contour of integration $\mathcal{L}$ in (3.52). In this figure, zeroes of the Hankel function, $H^{(1)}_{k\xi}(ka)$, are illustrated (as red stars) in relation to the contour of integration $\mathcal{L}$. In green stars, the zeroes of the sine function are plotted in relation to $\mathcal{L}$. The contour $\mathcal{L}$ does not contain inside its loop any zeroes of Hankel function provided (3.53) is satisfied.*

Proof of Proposition 3.15.

Proof. Deforming the contour $\mathcal{L}$ to the right yields residue contributions at

$$\xi_n = \frac{n}{k}.$$

We deduce from the Residue Theorem that

$$u(s,k) = 2\pi i \sum_{n=0}^{N} \delta_n \mathrm{Res}\left\{F_{k\xi}(kr),\ \xi_n\right\} + \int_{\mathcal{L}_N} F_{k\xi}(kr)d\xi, \tag{3.54}$$

where the contour of integration $\mathcal{L}_N$ is the contour $\mathcal{L}$ shifted to the right by $(N+1/2)/k$,

$$\mathcal{L}_N := \left\{\tilde{\xi} \in \mathbb{C} : \tilde{\xi} = \xi + \frac{1}{k}\left(N + \frac{1}{2}\right), \quad \text{where} \quad \xi \in \mathcal{L}\right\}.$$

The function $F_n(kr)$ in (3.54) denotes the integrand of (3.52),

$$F_{k\xi}(kr) = ik\left[J_{k\xi}(kr) - \frac{H^{(1)}_{k\xi}(kr)J_{k\xi}(ka)}{H^{(1)}_{k\xi}(ka)}\right] e^{-ik\xi\pi/2}\frac{\cos k\xi s}{\sin k\xi\pi}.$$

Recall the definition of $\operatorname{Res}\{F_{k\xi}(kr), \xi_n\}$. We can write $F_{k\xi}(kr)$ as

$$F_{k\xi}(kr) = \frac{f_{k\xi}(kr)}{\sin k\xi\pi},$$

with

$$f_{k\xi}(kr) = ik\left[J_{k\xi}(kr) - \frac{H^{(1)}_{k\xi}(kr)J_{k\xi}(ka)}{H^{(1)}_{k\xi}(ka)}\right]e^{-ik\xi\pi/2}\cos k\xi s.$$

Since both $f_{k\xi}(kr)$ and $\sin k\xi\pi$ are holomorphic inside $\mathcal{L}$,

$$\begin{aligned}\operatorname{Res}\{F_{k\xi}(kr),\ \xi_n\} := \frac{f_{k\xi_n}(kr)}{\left[\frac{d}{d\xi}\sin k\xi\pi\right]_{\xi=\xi_n}} &= (-1)^n\frac{f_n(kr)}{k\pi}\\ &= e^{in\pi}\frac{f_n(kr)}{k\pi}.\end{aligned}$$

Hence, the equivalence of (3.51) and (3.52) follows.

Since $J_{k\xi}$ and $H^{(1)}_{k\xi}$ are analytic with respect to ξ and $H^{(1)}_{k\xi}(ka) \neq 0$ and due to asymptotic properties of Bessel functions for $|\nu| = |k\xi| \to \infty$, see (3.15) and (3.16), we ensure that both (3.51) and (3.52) converge as $N \to \infty$ and $|\xi| \to \infty$ respectively.

□

The normal derivative of the function u on the boundary of the circle can also be written as a contour integral by differentiating (3.52) with respect to $\boldsymbol{n}(s)$. We will show that the integrand of this contour integral simplifies on the boundary with only Hankel function remaining in the integrand.

3.3.1 Evaluating the normal derivative of the solution

Differentiating (3.52) with respect to r and then setting $r = a$, we obtain the normal derivative of the solution $u(r, s)$ on the boundary:

$$v(s,k) := \left.\frac{\partial u(r,s)}{\partial r}\right|_{r=a} = ik^2\int_{\mathfrak{L}}\frac{W_{k\xi}(ak)\ e^{-ik\xi\pi/2}}{H^{(1)}_{k\xi}(ka)}\frac{\cos k\xi s}{\sin k\xi\pi}d\xi\ , \tag{3.55}$$

where the Wronskian $W_{k\xi}(rk)$ is defined by (3.19). Then using Proposition 3.7, the normal derivative of the solution $u(r, s)$ on the boundary simplifies to the following form:

$$v(s,k) = \frac{2k}{\pi a}\int_{\mathfrak{L}}\frac{1}{H^{(1)}_{k\xi}(ka)}\frac{e^{-ik\xi\pi/2}\cos k\xi s}{\sin k\xi\pi}d\xi\ . \tag{3.56}$$

Before we begin to evaluate the asymptotics of this integral for large k, we study the behaviour of the integrand for $\xi \in \mathfrak{L}$ and s in the illuminated or in the transition zones, and simplify (3.56) accordingly. In Theorem 3.18 we show how the integrals are simplified, for $s \in [0, \pi - \delta]$, $\delta > 0$.

Theorem 3.18. *When $s \in [\delta, \pi - \delta]$, where $\delta > 0$, the integral (3.56) simplifies asymptotically for large k to the integration only along the upper branch of $\mathfrak{L}$ as follows,*

$$v(s,k) = -\frac{2ik}{\pi a} \int_{i\beta}^{i\beta+\infty} \frac{e^{-ik\xi(s-\pi/2)}}{H_{k\xi}^{(1)}(ka)} d\xi + O\left(e^{-ck^{\gamma}}\right), \tag{3.57}$$

for some $c, \gamma > 0$.

To prove Theorem 3.18 we need to show firstly that the integral along the lower branch of $\mathfrak{L}$ is exponentially small when $k \to \infty$. This result follows from Lemma 3.20 which we state later and prove using Lemma 3.19 below.

Lemma 3.19. *For $\xi \in [-i\beta, -i\beta + \infty]$, and $\beta > 0$ and $a > 0$,*

$$\mathrm{Im}\left(\cos^{-1}\left(\frac{\xi}{a}\right)\right) \geq \gamma > 0. \tag{3.58}$$

Proof. Let $z \in \mathbb{C}$ with $\zeta := \mathrm{Re}(z) \in [0, \pi]$ and $\eta := \mathrm{Im}(z)$, be defined as follows,

$$z := \cos^{-1}\left(\frac{\xi}{a}\right). \tag{3.59}$$

Then, proving (3.58) is equivalent to proving that $\eta > 0$. From (3.59), we deduce,

$$\begin{aligned} \xi &= a\cos(z) = a\left(\frac{e^{iz} + e^{-iz}}{2}\right) = \frac{a}{2}e^{i\zeta}e^{-\eta} + \frac{a}{2}e^{-i\zeta}e^{\eta} \\ &= \frac{a}{2}(\cos\zeta + i\sin\zeta)e^{-\eta} + \frac{a}{2}(\cos\zeta - i\sin\zeta)e^{\eta} \\ &= a\cos\zeta\cosh\eta - ia\sin\zeta\sinh\eta. \end{aligned} \tag{3.60}$$

Therefore,

$$\mathrm{Re}(\xi) = a\cos\zeta\cosh\eta, \qquad \text{and} \qquad \mathrm{Im}(\xi) = -a\sin\zeta\sinh\eta. \tag{3.61}$$

Along the path $\xi \in [-i\beta, -i\beta + \infty]$, the imaginary part of ξ is negative: $\mathrm{Im}(\xi) = -\beta < 0$. Together with (3.61), the latter implies that ζ and η must satisfy the inequality

$$\sin\zeta\sinh\eta = \frac{\beta}{a} > 0.$$

Since $\zeta \in (0, \pi)$, we deduce that $\sinh \eta \geq \beta/a > 0$. Then

$$\eta = \operatorname{Im}\left(\cos^{-1}\left(\frac{\xi}{a}\right)\right) \geq \sinh^{-1}\left(\frac{\beta}{a}\right) =: \gamma,$$

as required. □

To motivate the following lemma, we make one elementary observation. The function $\exp(ikf(x))$ is exponentially decaying as $k \to \infty$ when the imaginary part of a function f is positive at x. Later, we will use this observation to prove Theorem 3.18.

For $\xi \in [-i\beta, -i\beta + \infty) \setminus \mathcal{B}_1$, we can replace the Hankel function in (3.56) by its Debye's asymptotic expansion (3.34). In Lemma 3.20 below, we show the resulting phase function in the integrand in (3.56) has positive imaginary part along the path $\xi \in [-i\beta, -i\beta + \infty]$.

Lemma 3.20. *Consider a function $\phi(\xi)$, for $\xi \in [-i\beta, -i\beta + \infty]$, $\beta > 0$ and $a > 0$ defined as follows,*

$$\phi(\xi) := \xi \cos^{-1}\left(\frac{\xi}{a}\right) - \sqrt{a^2 - \xi^2} + \xi\left(s - \frac{3\pi}{2}\right). \tag{3.62}$$

Then, for $s \in [\delta, \pi - \delta]$, $\delta > 0$

$$\operatorname{Im}(\phi(\xi)) \geq \beta\delta > 0. \tag{3.63}$$

Proof. Let $t := \operatorname{Re}(\xi)$, then we can write $\xi = -i\beta + t$, $t \geq 0$. Then,

$$\begin{aligned} \phi(-i\beta + t) &= (-i\beta + t)\cos^{-1}\left(\frac{-i\beta + t}{a}\right) - \sqrt{a^2 - (-i\beta + t)^2} \\ &+ (-i\beta + t)\left(s - \frac{3\pi}{2}\right). \end{aligned}$$

When $t = 0$, we have,

$$\phi(-i\beta) = -i\beta \cos^{-1}\left(-i\frac{\beta}{a}\right) - \sqrt{a^2 + \beta^2} - i\beta\left(s - \frac{3\pi}{2}\right). \tag{3.64}$$

Note that for $x \in \mathbb{R}$,

$$\cos^{-1}(-ix) = \frac{\pi}{2} - \sin^{-1}(-ix) = \frac{\pi}{2} + i\log\left(x + \sqrt{1 + x^2}\right) = \frac{\pi}{2} + i\sinh^{-1}(x).$$

Then equation (3.64) can be written as follows

$$\phi(-i\beta) = -i\beta\left(\frac{\pi}{2} + i\sinh^{-1}\left(\frac{\beta}{a}\right)\right) - \sqrt{a^2 + \beta^2} - i\beta\left(s - \frac{3\pi}{2}\right).$$

Hence for $s \in [0, \pi - \delta)$,

$$\operatorname{Im}(\phi(-i\beta)) = -\beta\frac{\pi}{2} + \beta\left(\frac{3\pi}{2} - s\right) = \beta(\pi - s) \geq \beta\delta > 0. \tag{3.65}$$

Since $\phi(\xi)$ is analytic in ξ, we have for $\xi = -i\beta + t$,

$$\frac{\partial \phi}{\partial \xi} := \lim_{h \to 0} \frac{\phi(\xi + h) - \phi(\xi)}{h} = \frac{\partial \phi}{\partial t},$$

with h tending to 0 along the real axis. On the other hand, from (3.62),

$$\frac{\partial \phi}{\partial \xi} = \cos^{-1}\left(\frac{\xi}{a}\right) + \left(s - \frac{3\pi}{2}\right).$$

Thus using Lemma 3.19, for $t \geq 0$,

$$\operatorname{Im}\left(\frac{\partial \phi}{\partial t}\right) = \operatorname{Im}\left(\cos^{-1}\left(\frac{i\beta + t}{a}\right)\right) \geq \gamma > 0. \tag{3.66}$$

Therefore, the function $\operatorname{Im}(\phi(-i\beta+t))$ increases monotonically in $t \geq 0$ with $\operatorname{Im}(\phi(-i\beta)) > 0$. Hence, the result follows. □

Proof of Theorem 3.18

Proof. The proof consists of investigating the behaviour of the integrand of (3.56), F, along the three branches of $\mathfrak{L}$. We denote the integrand as follows,

$$F(\xi) := \frac{1}{H^{(1)}_{k\xi}(ka)} \frac{e^{-ik\xi\pi/2} \cos k\xi s}{\sin k\xi\pi}. \tag{3.67}$$

Note that although the right hand side of (3.67) depends on parameters k, a and s, we do not reflect this in the notation for $F(\xi)$ for convenience.

Firstly, let us consider the behaviour of the function F along the vertical branch of $\mathfrak{L}$. The function $F(\xi)$ is an odd function with respect to ξ. To see this, note that $H^{(1)}_{-k\xi}(ka) = e^{i\pi k\xi} H^{(1)}_{k\xi}(ka)$, see e.g. [3, (9.1.6)], and hence

$$F(-\xi) = -\frac{e^{ik\xi\pi/2}}{H^{(1)}_{-k\xi}(ka)} \frac{\cos k\xi s}{\sin k\xi\pi} = -\frac{e^{-ik\xi\pi/2}}{H^{(1)}_{k\xi}(ka)} \frac{\cos k\xi s}{\sin k\xi\pi} = -F(\xi).$$

Therefore, the integration of $F(\xi)$ along the $[-i\beta, i\beta]$ branch of $\mathfrak{L}$ vanishes identically.

Along the lower branch of $\mathfrak{L}$, i.e. $[-i\beta, -i\beta + \infty)$, the integrand $F(\xi)$ turns out to be exponentially small. Indeed $F(\xi)$ is of order $O(e^{-k\beta\delta})$ for $\xi = -i\beta + t$, $t \geq 0$. To see this, first express the trigonometric functions in $F(\xi)$ by complex exponentials:

$$\frac{\cos(k\xi s)}{\sin(k\xi\pi)} = ie^{ik\xi(s-\pi)} \left[1 + O(e^{-2k\beta s}) + O(e^{-2k\beta\pi})\right].$$

Then, for $\xi \in [-i\beta, -i\beta + \infty)$, and $s \in [\delta, \pi - \delta]$,

$$\frac{\cos(k\xi s)}{\sin(k\xi\pi)} e^{-ik\xi\pi/2} = i \exp\left(ik\xi\left(s - \frac{3\pi}{2}\right)\right)\left[1 + O(e^{-2k\beta\delta})\right]. \tag{3.68}$$

Now, we replace the Hankel function $H^{(1)}_{k\xi}(ka)$ in $F(\xi)$ by its Debye's asymptotic expansion given in (3.34) (and also later in (3.73)) for $\xi \in [-i\beta, -i\beta+\infty)\backslash\mathcal{B}_1$. Then we write equation (3.67) in the form

$$F(\xi) = f(\xi, k)\exp(ik\phi(\xi))\left[1 + O(e^{-2k\beta\delta})\right], \tag{3.69}$$

where the phase function $\phi(\xi)$ is defined in (3.62) and the function $f(\xi, k)$ satisfies for all $\xi \in [-i\beta, -i\beta + \infty) \setminus \mathcal{B}_1$ with a fixed $a \in \mathbb{R}$, and $s \in [\delta, \pi - \delta)$, the following inequality,

$$|f(\xi, k)| \leq Ck^{1/2}\left(1 + |\xi|\right)^{1/2}, \tag{3.70}$$

where $C > 0$ is a constant independent of k and ξ. From Lemma 3.20, we deduce for $\xi \in [-i\beta, -i\beta + \infty) \setminus \mathcal{B}_1$,

$$F(\xi) = O\left(k^{1/2}\left(1 + |\xi|\right)^{1/2} e^{-k\beta\delta}\right) \to 0, \quad \text{as} \quad k \to \infty.$$

Now, from (3.66) and (3.63), we deduce that for $t \geq 0$

$$\mathrm{Im}(\phi(-i\beta + t)) \geq \beta\delta + \gamma t.$$

Then, also using (3.70) we obtain,

$$\begin{aligned}
\left|\int_0^\infty F(-i\beta + t)dt\right| &\leq \left|\int_0^\infty f(-i\beta + t, k)\exp(ik\phi(-i\beta + t))dt\right| \\
&\leq C_1 k^{1/2}\int_0^\infty (1 + t)^{1/2}\exp\left(-k\left(\beta\delta + \gamma t\right)\right)dt \\
&= C_1 k^{1/2}\exp(-k\beta\delta)\int_0^\infty (1 + t)^{1/2}\exp(-k\gamma t)dt,
\end{aligned}$$

where $C_1 > 0$ is a constant independent of k. Note that there exist a constant $C_2 > 0$ also independent of k such that $(1 + t)^{1/2} \leq C_2 e^{k\gamma t/2}$ for large enough k. Then,

$$\begin{aligned}
\left|\int_0^\infty F(-i\beta + t)dt\right| &\leq C_3 k^{1/2}\exp(-k\beta\delta)\int_0^\infty \exp(-k\gamma t/2)dt \\
&\leq C_4\frac{k^{1/2}\exp(-k\beta\delta)}{k\gamma} \leq C_5\exp(-k\beta\delta),
\end{aligned} \tag{3.71}$$

where $C_j > 0$, $j = 3, 4, 5$, are constants independent of k. From (3.71) we deduce that the integration of (3.56) along the lower branch of $\mathfrak{L}$, excluding the small part intersecting $\mathcal{B}_1$, yields an exponentially small term that is represented by $O\left(e^{-ck^\gamma}\right)$ in (3.57).

For ξ in the small part of the contour $[-i\beta, -i\beta+\infty) \cap \mathcal{B}_1$, we use Olver's asymptotic expansion of $H^{(1)}_{k\xi}(ka)$ given in (3.35). Using then the asymptotic expansions (3.31), (3.32) for the Airy functions in (3.35), we conclude that the integrand $F(\xi)$ remains exponentially small in k on $[-i\beta, -i\beta+\infty) \cap \mathcal{B}_1$.

Finally, on the upper part of $\mathfrak{L}$: $\xi \in [i\beta, i\beta+\infty]$, again we replace trigonometric functions by their complex exponentials to obtain, for $s \in [\delta, \pi-\delta)$,

$$\frac{\cos(k\xi s)}{\sin(k\xi\pi)} e^{-ik\xi\pi/2} = -ie^{-ik\xi(s-\pi/2)} \left[1 + O(e^{-2k\beta\delta})\right]. \tag{3.72}$$

Combining this with with the asymptotic expansion for $H^{(1)}_{k\xi}(ka)$ from (3.34) and (3.35) for the upper part of $\mathcal{L}$ results in an exponentially small contribution into the integral of the term corresponding in (3.72) to $O(e^{-2k\beta\gamma})$ (the technical details, somewhat analogous to those for the lower contour, are omitted).

Hence, by combining (3.71) and (3.72) the result of the theorem (3.57) follows. □

3.3.2 Asymptotic behaviour of the solution on the illuminated part of the boundary

In this section, we prove Theorem 3.1 in the case of a circle for $t_1 + \Delta < s < \pi - \delta$. We do this by evaluating the asymptotics of the simplified integral representation of $v(s,k)$ in (3.57) from Theorem 3.18. We employ the Debye's asymptotic expansion of Hankel's function reviewed in Section 3.2.4.

Debye's form of the asymptotic expansion From Theorem 3.18, we know that in order to obtain an asymptotic expansion of $v(s,k)$ for $s \in (\pi/2+\Delta, \pi-\delta]$, $\Delta, \delta > 0$, we require Debye's asymptotic expansion of the Hankel function $H^{(1)}_{k\xi}(ka)$ for $\xi \in [i\beta, i\beta+\infty)$.

In Figure 3.6, we illustrate the domain D where the Debye's asymptotic expansion of $H^{(1)}_{k\xi}(ka)$ is of the form (3.34) presented as follows,

$$\begin{aligned} H^{(1)}_{k\xi}(ka) &= \sqrt{\frac{2}{\pi k}} \frac{\exp\left\{ik\left(\sqrt{a^2-\xi^2} - \xi\cos^{-1}\left(\frac{\xi}{a}\right)\right) - \frac{i\pi}{4}\right\}}{(a^2-\xi^2)^{1/4}} \times \\ &\quad \times \left(1 + \sum_{m=1}^{M} \frac{A_m\left(\frac{\xi}{a}\right)}{k^m} + R_{k,M}\left(\frac{\xi}{a}\right)\right), \end{aligned} \tag{3.73}$$

with

$$\left|R_{k,M}\left(\frac{\xi}{a}\right)\right| \le C_M k^{-2(M+1)} (1+|\xi|)^{-M-1}$$

for all ξ in D - the domain in the complex plane illustrated as shaded regions in Figure 3.6. The functions $A_m(\xi/a)$ and $R_{k,M}(\xi/a)$ in (3.73) are regular for all ξ in D that excludes

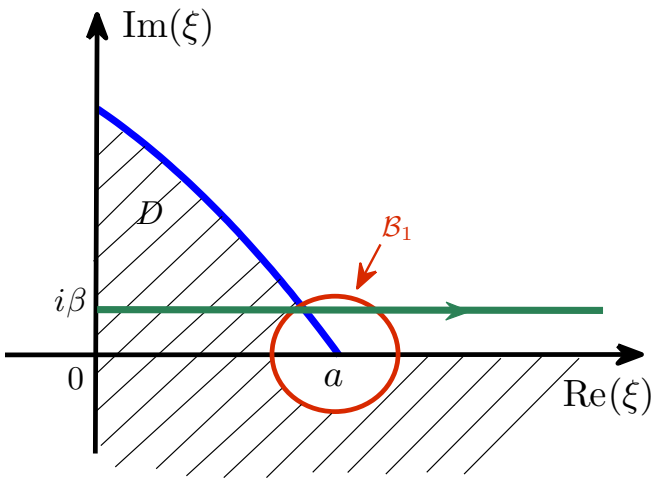


Figure 3.6: *The domain D in the complex ξ-plane represents the domain where the Debye's asymptotic representation of Hankel function is valid [16] and of the form (3.73). Domain D excludes the point $\xi = a$ and the circular domain around the point of a fixed radius denoted as $\mathcal{B}_1$. The path of integration, $\xi \in [i\beta, i\beta + \infty)$, is also illustrated*

the circular neighborhood of $\xi = a$. We know from Section 3.2.4 that for $\mathrm{Re}(\xi) > 0$ and $\mathrm{Im}(\xi) < 0$, the expansion (3.73) is valid.

Now we proceed to the **proof of Theorem 3.1 for** $s \in (\pi/2+\Delta, \pi-\delta)\cup(\pi+\delta, 3\pi/2-\Delta)$.

Remark 3.21. *The proof below does not include the case when $|s-\pi| \leq \delta$. However, the result of Theorem 3.1 still holds for $|s-\pi| \leq \delta$ as could be seen, for example, from the more general theory discussed in Section 3.4, where the Model Problem Method is described that can be used to prove (3.3) and (3.5) for general convex scatterers.*

Proof. We will prove the theorem for $s \in (\pi/2 + \Delta, \pi - \delta)$ for any small $\delta > 0$. The case when $s \in (\pi + \delta, 3\pi/2 - \Delta)$ can be proved analogously.

The main idea of the proof is to replace the Hankel function in the integral (3.57) by its Debye's asymptotic expansion (3.73) and apply the method of steepest descent. Writing $v(s,k)$ as in Theorem 3.18 and using Debye's asymptotic expansion (3.73), we deduce

$$\begin{aligned} v(s,k) &= -\frac{2ik}{\pi a}\int_{i\beta}^{i\beta+\infty} \frac{e^{-ik\xi(s-\pi/2)}}{H^{(1)}_{k\xi}(ka)} d\xi + EST \\ &= -\frac{2ik}{\pi a}\int_{i\beta}^{i\beta+\infty} F(\xi)\exp\left(ikG_s(\xi)\right) d\xi + EST, \end{aligned} \tag{3.74}$$

where EST represents henceforth the exponentially small terms with respect to k,

$$EST := O\left(e^{-ck^{\gamma}}\right), \tag{3.75}$$

with some $c > 0$ and $\gamma > 0$; EST remain exponentially small upon differentiation in s. In (3.74), the function $F(\xi)$ is regular in the shaded domain in Figure 3.6 as follows,

$$F(\xi) \quad = \quad \sqrt{\frac{\pi k}{2}} e^{i\pi/4} \left(a^2 - \xi^2\right)^{1/4} \left(1 + \sum_{m=1}^{M} \frac{\tilde{A}_m(k\xi)}{k^m} + O\left(k^{-M}\right)\right) + EST, \tag{3.76}$$

The phase function $G_s(\xi)$ is defined as follows:

$$G_s(\xi) = -\xi\left(\frac{\pi}{2} - s\right) + \xi \cos^{-1}\left(\frac{\xi}{a}\right) - \sqrt{a^2 - \xi^2}. \tag{3.77}$$

The notation $G_s(\xi)$ means that we think of G as a function of ξ parametrised by s.

The first derivative of $G_s(\xi)$ is given as follows,

$$\begin{aligned} G_s(\xi) \quad &= \quad -\left(\frac{\pi}{2} - s\right) + \cos^{-1}\left(\frac{\xi}{a}\right) - \frac{1}{a}\frac{\xi}{\sqrt{1 - \left(\frac{\xi}{a}\right)^2}} + \frac{\xi}{\sqrt{a^2 - \xi^2}} \\ &= \quad -\left(\frac{\pi}{2} - s\right) + \cos^{-1}\left(\frac{\xi}{a}\right), \end{aligned}$$

and the second derivative is given as

$$G_s''(\xi) = -\frac{1}{\sqrt{a^2 - \xi^2}}. \tag{3.78}$$

Therefore,

$$\xi_0 := a \sin s, \tag{3.79}$$

is a stationary (saddle) point of the phase function G_s:

$$G_s(\xi_0) = a \cos s, \qquad G_s'(\xi_0) = 0, \qquad G_s''(\xi_0) = -\frac{1}{a|\cos s|} = \frac{1}{a \cos s} < 0,$$

since $s \in (\pi/2 + \Delta, \pi - \delta)$.

Note that when $s \to \pi/2$, the stationary point ξ_0 in (3.79) enters the circular domain $\mathcal{B}_1$ around the point $\xi = a$, where the Debye's asymptotic expansion is not valid. Hence in the transitional domain of the boundary of the scatterer, another asymptotic representation of the Hankel function must be used, as we will show later.

We can obtain the asymptotic expansion of the integral (3.74) using the method of steepest descent see e.g. [101]. We deform the contour of integration into a new path of integration

that passes through the stationary (saddle) point ξ_0 of the phase function G. In doing so the integrand can be shown to be exponentially small everywhere on the deformed contour except near ξ_0 where the contour is of the "steepest descent" form.

In detail, we transform the contour of integration to a new contour illustrated in Figure 3.7. Along the deformed path, away from ξ_0, the integrand in (3.74) has the following properties: F and G_s are smooth and the imaginary part of the phase-function G_s is positive. In particular, analogously to Lemma 3.20, we can show that at $\xi = i\beta$, the imaginary part of G_s is positive. Also for $\mathrm{Re}(\xi) \to \infty$ and $\mathrm{Im}(\xi) < 0$, the imaginary part of G_s is positive.

Therefore, along the path $\mathcal{L}'$, away from $\xi = \xi_0$, the function $\exp(ikG_s(\xi))$ is exponentially small, i.e. of order $O(e^{-k\beta\delta}) \to 0$ as $k \to \infty$.

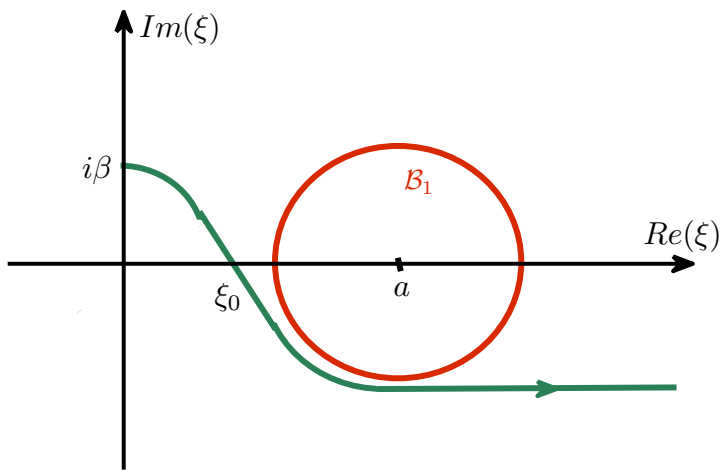


Figure 3.7: *Deformed contour of integration* $\mathcal{L}'$

Thus, the only significant contribution of the integral (3.74) comes from integration over a small sub-contour of the deformed contour containing the stationary point ξ_0, see Figure 3.8.

Using Taylor expansion near $\xi = \xi_0$, we deduce that

$$G_s(\xi) = G_s(\xi_0) + G_s''(\xi_0)\frac{(\xi - \xi_0)^2}{2} + O\left((\xi - \xi_0)^3\right). \tag{3.80}$$

We introduce new variable of integration t,

$$\xi - \xi_0 = e^{i3\pi/4}t. \tag{3.81}$$

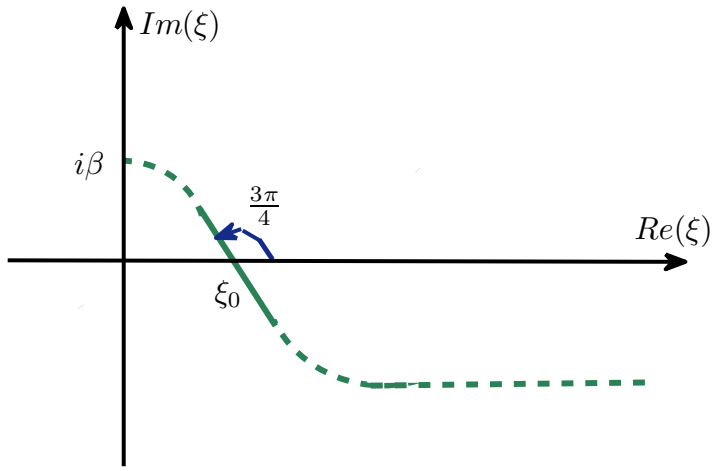


Figure 3.8: *The subcontour of the deformed contour of integration $\mathcal{L}'$ containing the stationary point ξ_0*

Then, we can write

$$\begin{aligned} v(s,k) &= -\frac{2ik}{a\pi}\int_{\mathcal{L}'} F(\xi)\exp(ikG_s(\xi))d\xi + EST \\ &= -\frac{2ik}{a\pi}\int_{\xi_0-e^{i3\pi/4}\epsilon}^{\xi_0+e^{i3\pi/4}\epsilon} F(\xi)\exp(ikG_s(\xi))d\xi + EST, \end{aligned}$$

where $\epsilon \leq ck^{-1/2+\gamma}$, for $0 < \gamma < 1/6$. Then, using (3.80), we deduce

$$\begin{aligned} v(s,k) &= -\frac{2ik}{a\pi}e^{ikG_s(\xi_0)}\int_{\xi_0-e^{i\pi/4}\epsilon}^{\xi_0+e^{i\pi/4}\epsilon} F(\xi)\exp\left(ikG_s''(\xi_0)\frac{(\xi-\xi_0)^2}{2} + O\left(ik(\xi-\xi_0)^3\right)\right)d\xi \\ &+ EST. \end{aligned} \tag{3.82}$$

We introduce the function $\tilde{F}(\xi)$ as follows

$$\tilde{F}(\xi) = F(\xi)\exp\left(ikG_s''(\xi_0)\frac{(\xi-\xi_0)^2}{2} + ikO\left(\xi-\xi_0\right)^3\right)\exp\left(-ikG_s''(\xi_0)\frac{(\xi-\xi_0)^2}{2}\right).$$

Note that the function $\tilde{F}$ is slowly-varying within the integration range in (3.82) since $|\xi-\xi_0| \leq ck^{-1/3}$. Also for notational convenience, we introduce the integral $I(k)$:

$$I(k) := \int_{\xi_0-e^{i\pi/4}\epsilon}^{\xi_0+e^{i\pi/4}\epsilon} \tilde{F}(\xi)\exp\left(ikG_s''(\xi_0)\frac{(\xi-\xi_0)^2}{2}\right)d\xi.$$

Then, from (3.82), the function $v(s,k)$ is equal to

$$v(s,k) = -\frac{2ik}{a\pi} e^{ikG_s(\xi_0)} I(k) + EST. \tag{3.83}$$

Expanding $\tilde{F}$ in Taylor series, we obtain,

$$\begin{aligned}\tilde{F}(e^{-i\pi/4}t+\xi_0) &= \tilde{F}(\xi_0) + \tilde{F}'(\xi_0)e^{-i\pi/4}t + \tilde{F}''(\xi_0)\frac{\left(e^{-i\pi/4}t\right)^2}{2!} + ... + \tilde{F}^{(N)}(\xi_0)\frac{\left(e^{-i\pi/4}t\right)^N}{N!} \\ &+ R_{N+1}(t),\end{aligned}$$

where

$$|R_N(t)| \leq C\frac{t^{N+1}}{(N+1)!}.$$

Then,

$$\begin{aligned} I(k) &= e^{i\pi/4}\int_{-\epsilon}^{+\epsilon} \tilde{F}(e^{-i\pi/4}t+\xi_0)\exp\left(-kG_s''(\xi_0)\frac{t^2}{2}\right)dt \\ &= e^{i\pi/4}\sum_{n=0}^{N}\frac{e^{i\pi n/4}\tilde{F}^{(n)}(\xi_0)}{n!}\int_{-\infty}^{+\infty} t^n \exp\left(-kG_s''(\xi_0)\frac{t^2}{2}\right)dt \\ &\quad + e^{i\pi/4}\int_{-\infty}^{+\infty} R_N(t)\exp\left(-kG_s''(\xi_0)\frac{t^2}{2}\right)dt + EST. \end{aligned} \tag{3.84}$$

To compute the integrals in (3.84), we make further change of variables

$$z = \left(k|G_s''(\xi_0)|\right)^{1/2} t,$$

then

$$\begin{aligned}\int_{-\infty}^{+\infty} t^n \exp\left(-kG_s''(\xi_0)\frac{t^2}{2}\right)dt &= \frac{1}{\left(kG''(\xi_0)\right)^{(n+1)/2}}\int_{-\infty}^{+\infty} z^n \exp\left(\frac{-z^2}{2}\right)dz \\ &= \frac{\sqrt{2\pi}n!}{\left(kG_s''(\xi_0)\right)^{(n+1)/2}}\begin{cases} 0, \text{ if } n \text{ is odd}, \\ 1, \text{ if } n=0, \\ \frac{1}{2^m m!} \text{ if } n=2m. \end{cases}\end{aligned}$$

Then, returning to (3.84), we obtain,

$$\begin{aligned} I(k) &= e^{-i\pi/4}\sum_{n=0}^{N}\frac{e^{i\pi n/4}\tilde{F}^{(n)}(\xi_0)}{n!}\frac{\sqrt{2\pi}n!}{\left(kG_s''(\xi_0)\right)^{(n+1)/2}}\begin{cases} 0, \text{ if } n \text{ is odd}, \\ 1, \text{ if } n=0, \\ \frac{1}{2^m m!} \text{ if } n=2m. \end{cases} \\ &+ \tilde{r}_N(s,k) + EST, \end{aligned}$$

where for all $s \in (\pi/2 + \Delta, \pi - \delta)$,

$$|\tilde{r}_N(s,k)| \leq C \frac{\sqrt{2\pi}}{(kG_s''(\xi_0))^{(N+2)/2}}. \tag{3.85}$$

Furthermore, the estimate (3.85) is also true for all derivatives of $\tilde{r}_N(s,k)$ taken with respect to s.

Now, from (3.83), we obtain,

$$\begin{aligned} v(s,k) &= -\sqrt{\frac{2}{\pi}}\frac{k}{a} e^{ikG_s(\xi_0)} e^{-i\pi/4} \sum_{m=0}^{M} \frac{1}{2^m m!} \frac{i^m \tilde{F}^{(2m)}(\xi_0)}{(kG_s''(\xi_0))^{m+1/2}} \\ &+ r_M(s,k) + EST, \end{aligned} \tag{3.86}$$

where for all $s \in (\pi/2 + \Delta, \pi - \delta)$,

$$|r_M(s,k)| \leq c \frac{1}{k^M}. \tag{3.87}$$

We write (3.86) in the form,

$$v(s,k) = k \exp(ika \cos s) \left\{ \sum_{j=0}^{M} k^{-j} d_j(s) + r_M(s,k) \right\}, \tag{3.88}$$

where $d_j(s)$, $j = 0, 1, 2, \ldots$, are smooth functions, see (3.76), defined as

$$d_j(s) := \pi \frac{(-1)^j}{j!} \tilde{F}^{(2j)}(\xi_0) \frac{(\cos s)^{(j+1)/2}}{2^{j-1}}.$$

The remainder term $r_M(s,k)$ is bounded as follows, see (3.87),

$$|r_M(s,k)| \leq c_0 k^{-M-1},$$

where $c_0 > 0$ is independent of k; and the derivatives are bounded as

$$|D_s^{(n)} r_M(s,k)| \leq c_{M,n} k^{-M-1},$$

where $c_{M,n} > 0$ is independent of k.

From (3.88) we deduce that the asymptotic representation of the function $V(s,k)$ defined in (3.2) is of the form:

$$V(s,k) = \sum_{j=0}^{M} k^{-j} d_j(s) + r_M(s,k). \tag{3.89}$$

The latter expansion is in agreement with (3.3). The leading order term in the expansion

for v is

$$\begin{aligned} v(s,k) &\simeq -\sqrt{\frac{2}{\pi}}\frac{2ik}{a}e^{ikG_s(\xi_0)}e^{-i\pi/4}\frac{\tilde{F}(\xi_0)}{\sqrt{kG_s''(\xi_0)}} \\ &= -\sqrt{\frac{2}{\pi}}\frac{2ik}{a}e^{ikG_s(\xi_0)}e^{-i\pi/4}\frac{e^{i\pi/4}\left(\pi k\sqrt{a^2-\xi_0^2}\right)^{1/2}}{\sqrt{2kG_s''(\xi_0)}} \\ &= -2ike^{ikG(\xi_0)}\sqrt{\frac{a|\cos s|}{G_s''(\xi_0)}} \\ &= 2ike^{ika\cos s}\cos s. \end{aligned}$$

Note $d_0(s)$ is $d_0(s) := 2i\cos s$ is in agreement with $d_0(s) := 2i\boldsymbol{n}(s)\cdot\boldsymbol{a}$, where $\boldsymbol{n}(s)$ is an outer normal to the boundary of the scatterer unit vector and $\boldsymbol{a}$ is a unit vector representing the direction of incidence. □

3.3.3 Asymptotic behaviour of the solution on the transitional parts of the boundary

As we have mentioned earlier, the Debye asymptotic expansion of the Hankel function is not valid near the transition point $\xi = a$. Instead, we use the Cherry's form of the uniform asymptotic expansion for the Hankel function to prove Theorem 3.2.

The Cherry asymptotic representation of Hankel function is given as follows, see (3.40), for complex ξ in $\mathcal{B}_1$, see Figure 3.6,

$$H_{k\xi}^{(1)}(ka) = e^{i\pi/3}(ka)^{1/3}\left(\frac{T(\xi)}{\xi^2-a^2}\right)^{1/4}\alpha(\zeta)\mathrm{A}_+(T(\xi)+\sum_{j=1}^{M}\frac{p_j(\zeta)}{(ka)^{2j-2/3}}+O\left((k)^{-2(M+1)}\right)), \tag{3.90}$$

where

$$\mathrm{A}_+(z) := \mathrm{Ai}\left(e^{\frac{2\pi i}{3}}z\right),$$

is the Airy function. In (3.90), $T(\xi)$ and ζ are defined as follows (see [5, (7.2.4)]),

$$\begin{aligned} T(\xi) := (k\xi)^{2/3}\zeta &= \left[\frac{3}{2}ika\int_1^{\xi/a}\cos^{-1}(x)dx\right]^{2/3} \\ &= 2\left(\frac{ka}{2}\right)^{2/3}\left(\frac{\xi}{a}-1\right)\left[1-\frac{1}{30}\left(\frac{\xi}{a}-1\right)+\ldots\right], \end{aligned} \tag{3.91}$$

and where $p_1(\zeta), p_2(\zeta), \ldots, p_M(\zeta)$ are functions that are analytic at $\zeta = 0$, i.e. at $\xi = a$.

The function $\alpha(\zeta)$, defined as

$$\alpha(\zeta) := \frac{d}{d\zeta}\left(\zeta + \sum_{j=1}^{M} \frac{p_j(\zeta)}{(ak)^{2j}}\right), \tag{3.92}$$

is regular at $\zeta = 0$. Furthermore, by (3.91), the function $\left(\frac{T(\xi)}{\xi^2-a^2}\right)^{1/4}$ is analytic near $\xi = a$. Now we can proceed to the proof of Theorem 3.2 in the case of a circle.

Proof of Theorem 3.2

Proof. From Theorem 3.18, we have the following asymptotically simplified representation of the normal derivative, see (3.57):

$$v(s,k) = \frac{2ik}{\pi a}\int_{i\beta}^{i\beta+\infty} \frac{e^{-ik\xi(s-\pi/2)}}{H^{(1)}_{k\xi}(ka)} d\xi + EST.$$

Our immediate goal is to deform the contour of integration to enable us to retrieve the short-wave asymptotic expansion of the integral. We deform the old contour of integration with the new contour which we denote as $\mathcal{L}_1$ illustrated in Figure 3.9.

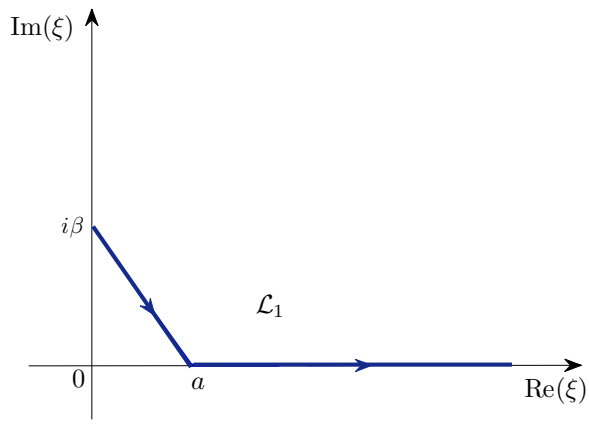


Figure 3.9: *The deformed contour of integration.*

If s is close to $\pi/2$, the biggest contribution of the integral is over the part of the contour that is near $\xi = a$. Indeed, within the accuracy of terms which are exponentially small as $k \to \infty$, we can replace the contour of integration $\mathcal{L}_1$ by $\mathcal{L}_2$ of small enough $O(1)$ size around point $\xi = a$ as illustrated in Figure 3.10. Now we have,

$$v(s,k) = \frac{2ik}{\pi a}\int_{\mathcal{L}_2} \frac{e^{-ik\xi(s-\pi/2)}}{H^{(1)}_{k\xi}(ka)} d\xi + EST.$$

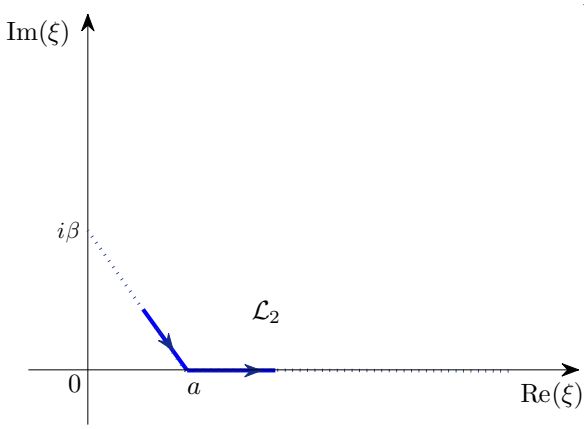


Figure 3.10: *Within an accuracy of terms which are exponentially small as $k \to \infty$, we can replace the contour of integration $\mathcal{L}_1$ with $\mathcal{L}_2$.*

We next replace the Hankel function with its Cherry asymptotic representation (3.90) and obtain

$$v(s,k) = \frac{2ik}{\pi a} \int_{\mathcal{L}_2} F(\xi) e^{-ik\xi(s-\pi/2)} d\xi + EST, \tag{3.93}$$

where the function $F(\xi)$ is defined by

$$F(\xi) = e^{-i\pi/3} \frac{(ka)^{-1/3}}{\alpha(\zeta)} \left(\frac{T(\xi)}{\xi^2 - a^2} \right)^{-1/4} \frac{1}{\mathrm{A}_+ \left(T(\xi) + \sum_{j=1}^{M} \frac{p_j(\zeta)}{(ka)^{2j-2/3}} + O\left(k^{-(2M+4/3)}\right) \right)}, \tag{3.94}$$

The function $\alpha(\zeta)$ defined in (3.92) is a non-zero analytic function at a neighborhood of $\zeta = 0$ (i.e. $\xi = a$) and $p_1(\zeta)$, $p_2(\zeta), \ldots, p_N(\zeta)$, are regular functions at $\zeta = 0$ and are determined from a recurrent relation [29]; T and ζ are defined by (3.91). In order to obtain the asymptotic expansion of $v(s,k)$ from (3.93), in the neighbourhood of $\zeta = 0$ we introduce, for large enough k, the new variable z:

$$z = \mathcal{T}(\xi) = (k\xi)^{2/3}\zeta + \sum_{j=1}^{N} \frac{p_j(\zeta)}{(ak)^{2j-2/3}}. \tag{3.95}$$

Notice that due to (3.91), $\mathcal{T}$ is invertible on $\mathcal{L}_2$ chosen small enough for large enough k. Therefore,

$$\xi = \mathcal{T}^{-1}(z) = a + k^{-2/3} \left(\frac{a}{2}\right)^{1/3} z + r(z,k), \tag{3.96}$$

where $r(z,k) = O\left(k^{-4/3}z^2 + k^{-2}\right)$. See [5, Section 7.2] for the proof of this fact.

In the new variable z, the transformed contour of integration $\mathcal{L}_3$ is illustrated in Figure 3.11.

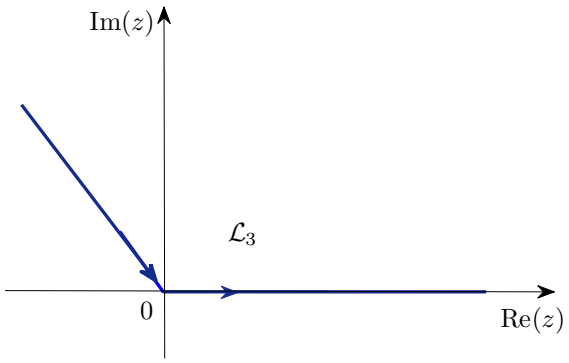


Figure 3.11: *Integration path in the new variable z. Due to the definition of z in (3.95), the contour of integration in z extends to infinity with exponentially small error. The contour $\mathcal{L}_3$ is of the same form as a contour c in Figure 3.1.*

Now, the function $F(\xi)$ takes the following form after the change of variables:

$$F(z) = B_k(z)\frac{1}{A_+\left(z + r_M\right)},$$

where $r_M = O(k^{-(2M+4/3)})$ and $B_k(z)$ is defined as

$$B_k(z) = e^{-i\pi/3}\frac{(ka)^{-1/3}}{\alpha\left(\eta\left(\mathcal{T}^{-1}(z)\right)\right)}\left(\frac{(\mathcal{T}^{-1}(z))^2 - a^2}{T\left(\mathcal{T}^{-1}(z)\right)}\right)^{1/4}\left(\mathcal{T}^{-1}\right)'(z),$$

where the functions T and $\mathcal{T}$ are defined in (3.91) and (3.95) respectively, and the function $\alpha(\zeta)$ is defined in (3.92), and η denotes the function $\zeta = \eta(\xi)$ (that can be explicitly determined from (3.91) and satisfies $\eta(a) = 0$). Note that

$$z = 0 \quad \Longleftrightarrow \quad \mathcal{T}^{-1}(z) = a \quad \Longleftrightarrow \quad \xi = a \quad \Longleftrightarrow \quad \zeta = 0.$$

Further notice that due to (3.91), the function

$$\left(\frac{\left(\mathcal{T}^{-1}(z)\right)^2 - a^2}{T\left(\mathcal{T}^{-1}(z)\right)}\right)^{1/4}$$

is analytic at $z = 0$. Then, due to the uniformity of the expansion (3.94) and the estimate (3.96) and the fact that $\alpha(\zeta)$ is a non-zero, analytic function in the neighborhood of $\zeta = 0$, we can write

$$B_k(z) = \sum_{j=0}^{N}\beta_j(z)k^{-j} + O\left(k^{-N-1}\right), \tag{3.97}$$

where $\beta_j(z)$, $j = 0, 1, 2, \ldots$ are analytic at $z = 0$ (in fact, $\beta_j(z)$ are analytic on $\mathcal{L}_3$ with radius of convergence $O(k^{2/3})$ around $z = 0$).

Therefore, $B_k(z)$ is an analytic function at $z = 0$. Thus, the integral (3.93) can be written as follows:

$$\begin{aligned} v(s,k) &= \frac{ik^{2/3}}{\pi}\left(\frac{2}{a}\right)^{2/3} e^{-ika\left(s-\frac{\pi}{2}\right)}\left(\int_{\mathcal{L}_3} B_k(z)\frac{e^{izk^{1/3}Z(s)-ikr(z,k)\left(s-\frac{\pi}{2}\right)}}{A_+(z)}dz\right)\times \\ &\quad \times\left(1+O\left(k^{-(2M+\frac{4}{3})}\right)\right)+EST, \end{aligned} \tag{3.98}$$

where $r(z,k)$ is defined in (3.96) and Z is defined by

$$Z(s) = 2^{-1/3}a\int_s^{\pi/2}\kappa^{2/3}(s)ds = 2^{-1/3}a^{1/3}\left(\frac{\pi}{2}-s\right), \tag{3.99}$$

where $\kappa(s) = a^{-1}$ is the curvature. Now, let us introduce for convenience the function $\tilde{B}_k$,

$$\tilde{B}_k(z) = B_k(z)e^{-ikr(z,t)\left(s-\frac{\pi}{2}\right)}.$$

Since the function B_k is analytic at $z = 0$, we can expand $\tilde{B}_k$ around 0 using Taylor's expansion:

$$\tilde{B}_k(z) = \sum_{m=0}^{\infty}\tilde{b}_m(s,k)z^m,$$

where $\tilde{b}_m(s,k)$ are smooth in s and, due to (3.97),

$$\tilde{b}_m(s,k) = \sum_{j=0}^{N}k^{-j}\tilde{b}_m(s)+O(k^{-N-1}). \tag{3.100}$$

Let us denote the integral in (3.98) for convenience as follows,

$$I(k) := \int_{\mathcal{L}_3} B_k(z)e^{-ikr(z,k)\left(s-\frac{\pi}{2}\right)}\frac{e^{izk^{1/3}Z(s)}}{A_+(z)}dz.$$

Then, by introducing $\tilde{B}_k(z) := B_k(z)\exp(-ikr(z,k)\left(s-\frac{\pi}{2}\right))$, we can write

$$\begin{aligned} I(k) &= \int_{\mathcal{L}_3}\tilde{B}_k(z)\frac{e^{izk^{1/3}Z(s)}}{A_+(z)}dz = \int_{\mathcal{L}_3}\left(\sum_{m=0}^{\infty}\tilde{b}_m(s,k)z^m\right)\frac{e^{izk^{1/3}Z(s)}}{A_+(z)}dz \\ &= \sum_{m=0}^{\infty}\tilde{b}_m(s,k)e^{ika\left(\frac{\pi}{2}-s\right)^3/6}\int_{\mathcal{L}_3}z^m\frac{e^{izk^{1/3}Z(s)-ikZ^3(s)/3}}{A_+(z)}dz, \end{aligned}$$

where the latter equation is obtained by multiplying and dividing by $\exp(ikZ^3(s)/3)$ and using the fact that $Z^3(s) = (1/2)a(\pi/2-s)^3$ (from (3.99)).

Now, denote

$$\tau = k^{1/3}Z(s). \tag{3.101}$$

Then,

$$\begin{aligned} I(k) &= e^{ika\left(\frac{\pi}{2}-s\right)^3/6} \sum_{m=0}^{\infty} \tilde{b}_m(s,k) e^{-i\tau^3/3} \int_{\mathcal{L}_3} z^m \frac{e^{iz\tau}}{A_+(z)} dz \\ &= e^{ika\left(\frac{\pi}{2}-s\right)^3/6} \sum_{m=0}^{\infty} \tilde{b}_m(s,k) e^{-i\tau^3/3} (-i)^m \frac{d^m}{d\tau^m} \left(\int_{\mathcal{L}_3} \frac{e^{iz\tau}}{A_+(z)} dz \right) \\ &= e^{ika\left(\frac{\pi}{2}-s\right)^3/6} \sum_{m=0}^{\infty} \tilde{\tilde{b}}_m(s,k) \frac{d^m}{d\tau^m} \left(e^{-i\tau^3/3} \int_{\mathcal{L}_3} \frac{e^{iz\tau}}{A_+(z)} dz \right), \end{aligned} \tag{3.102}$$

where $\tilde{\tilde{b}}_m(s,k)$ are smooth. Finally, substituting (3.102) into (3.98), we deduce

$$\begin{aligned} v(s,k) &= \frac{ik^{1/3}}{\pi} \left(\frac{2}{a}\right)^{2/3} e^{-ika\left(s-\frac{\pi}{2}\right)} e^{ika\left(\frac{\pi}{2}-s\right)^3/6} \left(\sum_{m=0}^{\infty} \tilde{\tilde{b}}_m(s,k) \Psi^{(m)}(\tau) \right) \times \\ &\times \left(1 + O\left(k^{-(2M+\frac{4}{3})}\right)\right) + EST, \end{aligned}$$

where the function $\Psi(\tau)$ is defined as

$$\Psi(\tau) = e^{-i\tau^3/3} \int_c \frac{e^{iz\tau}}{A_+(z)} dz,$$

the contour of integration being illustrated in Figure 3.1.

Recall that

$$\cos s = \sin(\pi/2 - s) \simeq \left(\frac{\pi}{2} - s\right) - \frac{1}{6}\left(\frac{\pi}{2} - s\right)^3 + O\left(\left(\frac{\pi}{2} - s\right)^5\right).$$

Thus, we can write the integral for $v(s,k)$ in the form

$$v(s,k) = e^{ika\cos s} \sum_{m=0}^{\infty} k^{-1/3-2m/3} b_m(s,k) \Psi^{(m)}\left(k^{1/3} Z(s)\right) + EST, \tag{3.103}$$

where $b_m(s,k)$, $m = 0, 1, 2, \ldots$ are smooth functions in s and can also be written in the form (3.100). Now, putting

$$b_m(s,k) = \sum_{j=0}^{N} k^{-j} b_{j,m}(s) + R_N(s,k),$$

with

$$|R_N(s,k)| = O\left(k^{-N-1}\right).$$

we write the equation (3.103) as follows

$$\begin{aligned} v(s,k) &= ke^{ika\cos s}\sum_{m=0}^{L}\sum_{j=0}^{N}k^{-1/3-j-2m/3}b_{j,m}(s)\Psi^{(m)}\left(k^{1/3}Z(s)\right) \\ &+ R_{L,N}(s,k)+EST. \end{aligned} \tag{3.104}$$

Equation (3.104) is in agreement with (3.5), where $R_{L,N}(s,k)$ is as in (3.6). This concludes the proof. □

3.4 Model Problem Method

In this section, we will outline the Model problem method [5] which in turn followed [72], that can be used to prove the results (3.3) and (3.5) for general convex obstacles. The main idea of the Model problem method is to construct an asymptotic approximation to the scattering problem (2.1)- (2.3) consistent with (3.3) and (3.5). The resulting approximation solves the BVP (2.1)- (2.3) with small errors on the right hand-side of (2.1). Finally, methods of e.g. [78] yield the error bounds (3.4) and (3.6) respectively.

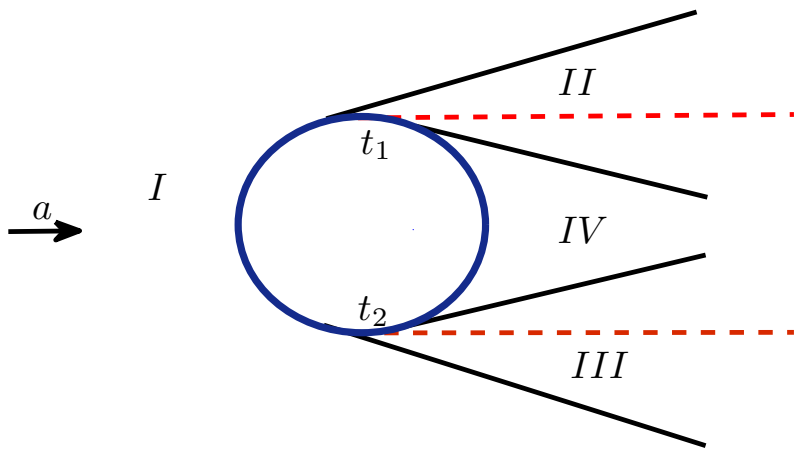


Figure 3.12: *The exterior to the general convex scatterer domain is separated into illuminated (I), transitional (II and III) and shadow (IV) zones with respect to the incident wave.*

In Figure 3.12, the exterior to the scatterer domain is separated into four regions: the illuminated (I), the shadow (IV) and the two transitional subdomains (II and III). We seek the asymptotic representation of the solution u to the exterior scattering problem in each of the four domains.

On the illuminated domain, the solution can be found using the Ray method. We describe the ray method in Section A of the Appendix. Mathematically, ray methods are simply an extension of the WKBJ method: the solution is constructed in the form of a WKBJ-type asymptotic expansion. Moreover, ray methods provide a physical insight into wave propagation phenomena, extending basic principles of the geometrical optics theory.

On the other hand, it is suggested in [5] that we seek the asymptotic expansion of u in the transition zones, in a form similar to that for the circle (3.52) which we present shortly.

The ray approximation in the illuminated zone and the asymptotic approximation in the transition zone have to match with each other in the overlapping regions. Let us denote the resulting approximations by $u_j^a(\boldsymbol{x}, k)$, $\boldsymbol{x} \in \mathbb{R}^2 \setminus \Omega$, with parameter $j = 1, 2, 3, 4$ representing the region in the exterior domain the asymptotic expansion corresponds to.

Then, for $\boldsymbol{x} \in I \cup II$, with $\mathcal{P}$ illustrated in Figure 3.13,

$$u(\boldsymbol{x}) \cong u_I^a(\boldsymbol{x}, k)\chi_1(\boldsymbol{x}) + u_{II}^a(\boldsymbol{x}, k)\chi_2(\boldsymbol{x}), \tag{3.105}$$

where $\chi_j : I \cup II \to [0,1]$, $j = 1,2$ are smooth functions that satisfy the following conditions:

$$\begin{aligned}
\chi_1(\boldsymbol{x}) &= 1, \text{ if } \boldsymbol{x} \in I \setminus \mathcal{P} \quad \text{and} \quad \chi_1(\boldsymbol{x}) = 0, \text{ if } \boldsymbol{x} \in II, \\
\chi_2(\boldsymbol{x}) &= 1, \text{ if } \boldsymbol{x} \in II \setminus \mathcal{P} \quad \text{and} \quad \chi_2(\boldsymbol{x}) = 0, \text{ if } \boldsymbol{x} \in I, \\
&\text{and} \quad \chi_1(\boldsymbol{x}) + \chi_2(\boldsymbol{x}) = 1, \quad \text{if } \boldsymbol{x} \in \mathcal{P}.
\end{aligned}$$

Note that the last equation implies $\nabla\chi_1(\boldsymbol{x}) = -\nabla\chi_2(\boldsymbol{x})$ and $\Delta\chi_1(\boldsymbol{x}) = -\Delta\chi_2(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{P}$. Then, for $\boldsymbol{x} \in \mathcal{P}$,

$$\begin{aligned}
\left(\Delta + k^2\right) u(\boldsymbol{x}) &\cong \left(\Delta + k^2\right)\left(u_1^a(\boldsymbol{x}, k)\chi_1(\boldsymbol{x}) + u_2^a(\boldsymbol{x}, k)\chi_2(\boldsymbol{x})\right) \\
&= \left(\Delta + k^2\right) u_I^a(\boldsymbol{x}, k)\chi_1(\boldsymbol{x}) + \left(\Delta + k^2\right) u_{II}^a(\boldsymbol{x}, k)\chi_2(\boldsymbol{x}) \\
&\quad +2\nabla\left[u_I^a(\boldsymbol{x}, k) - u_{II}^a(\boldsymbol{x}, k)\right] \cdot \nabla\chi_1(\boldsymbol{x}) \qquad (3.106) \\
&\quad + \left[u_I^a(\boldsymbol{x}, k) - u_{II}^a(\boldsymbol{x}, k)\right] \Delta\chi_1(\boldsymbol{x}). \qquad (3.107)
\end{aligned}$$

From (3.106) and (3.107) we deduce that the asymptotic approximations $u_I^a(\boldsymbol{x}, k)$ and

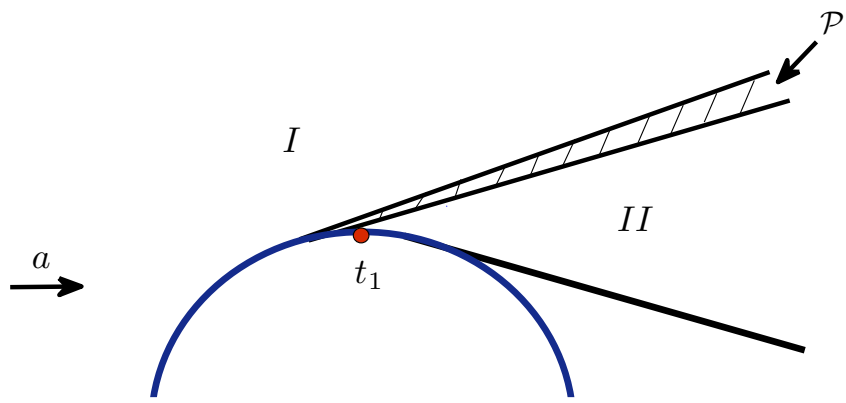


Figure 3.13: *The illuminated (I), transitional (II) zones overlap in the domain $\mathcal{P}$.*

$u_{II}^a(\boldsymbol{x}, k)$ must not only represent the solution u in I and II accurately but also *match* in the overlap region $\mathcal{P}$ illustrated in Figure 3.13, i.e. to ensure the smallness of the term $u_I^a - u_{II}^a$ in (3.106) and (3.107). The same applies to matching II and III with the shadow fields in IV which is set to identical zero. The latter matching will be ensured

automatically by the exponential decay property (3.8) of $\Psi(\tau)$ as $\tau \to -\infty$.

3.4.1 Asymptotic representation of the solution in the transition regions for general convex domains

It is suggested in [5] that we seek the asymptotic expansion of u in the transition zones in the form

$$u(\boldsymbol{x}) = u^I(\boldsymbol{x}) + u^S(\boldsymbol{x}) \simeq k \int_L \left[J(\boldsymbol{x},\gamma) - H(\boldsymbol{x},\gamma) \frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)} \right] e^{ik\xi(\boldsymbol{x},\gamma)} d\gamma. \tag{3.108}$$

Here Ai is the Airy function and $\mathrm{A}_+(z) := \mathrm{Ai}\left(e^{2i/3}z\right)$ and the contour of integration L is a "forked" shaped contour illustrated in Figure 3.14. Integration over the contour L is described in [5, Section 13.1]. The function Ai in (3.108) is replaced by

$$\mathrm{Ai}(t) = \frac{1}{i}\left(\mathrm{Ai}\left(te^{-i\pi/3}\right) + \mathrm{Ai}\left(te^{i\pi/3}\right)\right),$$

and the integrand in (3.108) is split into two terms. The term containing $\mathrm{Ai}(te^{-i\pi/3})$ is integrated along the lower branch of L, i.e. with $\mathrm{Re}(te^{-i\pi/3}) < 0$ and $\mathrm{Im}(te^{-i\pi/3}) < 0$. The term containing $\mathrm{Ai}(te^{-i\pi/3})$ is integrated along the upper branch of L. The function Ai does not need to be split up along the horizontal branch of L.

The point $\boldsymbol{x}$ is the observation point. The functions $\xi(\boldsymbol{x},\gamma)$ and $\mu(\boldsymbol{x},\gamma)$ that appear in equations below, are unknown and $J(\boldsymbol{x},\gamma)$ and $H(\boldsymbol{x},\gamma)$ are given as follows:

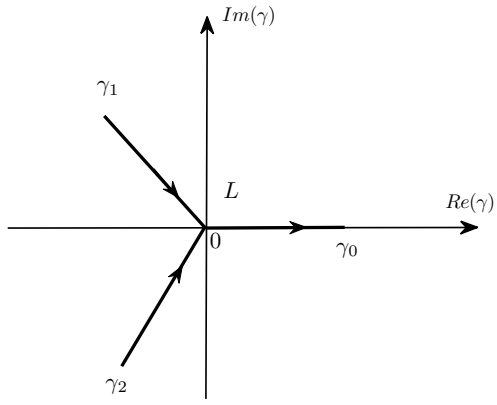


Figure 3.14: *The contour of integration Λ_3 in the complex plane. $|\gamma_1| = |\gamma_2| = \gamma_0$, where γ_0 is a small positive number.*

$$J(\boldsymbol{x},\gamma) = k^{-1/3}\mathrm{Ai}\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right)\mathcal{A}(\boldsymbol{x},\gamma,k) + ik^{-2/3}\mathrm{Ai}'\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right)\mathcal{B}(\boldsymbol{x},\gamma,k), \tag{3.109}$$

$$H(\boldsymbol{x},\gamma) = k^{-1/3}\mathrm{A}_+\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right)\mathcal{A}(\boldsymbol{x},\gamma,k) + ik^{-2/3}\mathrm{A}'_+\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right)\mathcal{B}(\boldsymbol{x},\gamma,k), \tag{3.110}$$

where $\mathcal{A}(\boldsymbol{x},\gamma,k)$ and $\mathcal{B}(\boldsymbol{x},\gamma,k)$ are sought in the form

$$\mathcal{A}(\boldsymbol{x},\gamma,k) = \sum_{j=0}^{N} k^{-j} A_j(\boldsymbol{x},\gamma), \tag{3.111}$$

and

$$\mathcal{B}(\boldsymbol{x},\gamma,k) = \sum_{j=0}^{N} k^{-j} B_j(\boldsymbol{x},\gamma), \tag{3.112}$$

where A_j and B_j are unknown. We in turn seek ξ, μ, A_j and B_j as power series expansions in γ:

$$\xi(\boldsymbol{x},\gamma) = \sum_{j=0}^{\infty} l_j(\boldsymbol{x})\gamma^j \qquad \mu(\boldsymbol{x},\gamma) = \sum_{j=0}^{\infty} m_j(\boldsymbol{x})\gamma^j \tag{3.113}$$

$$A_n(\boldsymbol{x},\gamma) = \sum_{j=0}^{\infty} \mathcal{A}_{nj}(\boldsymbol{x})\gamma^j \qquad B_n(\boldsymbol{x},\gamma) = \sum_{j=0}^{\infty} \mathcal{B}_{nj}(\boldsymbol{x})\gamma^j. \tag{3.114}$$

Substituting the sums into (3.109) and (3.110), we obtain the so-called *approximate caustic sums* (AC sums). The method for constructing the AC sums can be briefly described as follows. The "anzatz" (3.108) - (3.114) is then substituted into the Helmholtz equation (2.1) and the boundary conditions (2.2). We require that the coefficients of successive powers of k^{-1} on both sides of these equations are identical. This will lead to a recurrence relation between A_{nj} and B_{nj} and l_j and m_j.

In more detail, the integrand takes the form of the following asymptotic expansion:

$$\begin{aligned}
&\left[J(x,\gamma) - H(\boldsymbol{x},\gamma)\frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)}\right]e^{ik\xi(\boldsymbol{x},\gamma)} = \\
&\quad k^{-1/3}\left\{\mathrm{Ai}\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right) - \frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)}\mathrm{A}_+\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right)\right\}\mathcal{A}(\boldsymbol{x},\gamma,k)e^{ik\xi(\boldsymbol{x},\gamma)} \\
&\quad + ik^{-2/3}\left\{\mathrm{Ai}'\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right) - \frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)}\mathrm{A}'_+\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right)\right\}\mathcal{B}(\boldsymbol{x},\gamma,k)e^{ik\xi(\boldsymbol{x},\gamma)}
\end{aligned}$$

with the unknown coefficients $A_j(\boldsymbol{x},\gamma)$ and $B_j(\boldsymbol{x},\gamma)$, as well as $\mu(\boldsymbol{x},\gamma)$ and $\xi(\boldsymbol{x},\gamma)$.

The boundary condition (2.2) for $u(\boldsymbol{x})$ translate into following conditions for the unknowns: for $\boldsymbol{x} \in \Gamma$,

$$\mu(\boldsymbol{x},\gamma) = -\gamma, \tag{3.115}$$

$$B_j(\boldsymbol{x},\gamma) \equiv 0. \tag{3.116}$$

For notational convenience, we introduce the function D as follows,

$$D(\boldsymbol{x},\gamma,k) = W\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right)\mathcal{A}(\boldsymbol{x},\gamma,k) + ik^{-1/3}W'\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right)\mathcal{B}(\boldsymbol{x},\gamma,k),$$

where $W(z)$ is either the Airy function $Ai(z)$ or $A_+(z)$. Note that both functions $Ai(z)$ and $A_+(z)$ solve the Airy differential equation (3.23). Then, in order to determine the unknown functions, we solve asymptotically the following equation, to ensure an approximation to (2.1),

$$(\Delta + k^2)\left(D(\boldsymbol{x},\gamma,k)e^{ik\xi(\boldsymbol{x},\gamma)}\right) = 0. \tag{3.117}$$

Then by equating the terms of order k^2 in (3.117) the functions $\xi(\boldsymbol{x},\gamma)$ and $\mu(\boldsymbol{x},\gamma)$ can be found.

Proposition 3.22. *If the function $u(\boldsymbol{x})$ in (3.108) satisfies*

$$(\Delta + k^2)u(\boldsymbol{x}) = 0,$$

then the functions $\xi(\boldsymbol{x},\gamma)$ and $\mu(\boldsymbol{x},\gamma)$ satisfy

$$\begin{cases} |\nabla\xi|^2 + \mu|\nabla\mu|^2 &= 1, \\ \nabla\mu\cdot\nabla\xi &= 0. \end{cases} \tag{3.118}$$

Proof. See Section G of Appendix. □

3.4.2 Determining the eikonal analogue

The system of equations (3.118) is seen to be equivalent to the system of two eikonal equations which can be written as follows:

$$\left|\nabla\left(\xi \pm \frac{2}{3}\mu^{3/2}\right)\right|^2 = 1. \tag{3.119}$$

To solve this system, we seek further expansions in the powers of γ,

$$\xi(\boldsymbol{x},\gamma) = \sum_{j=0}^{\infty}\xi_j(\boldsymbol{x})\gamma^j, \quad \text{and} \quad \mu(\boldsymbol{x},\gamma) = \sum_{j=0}^{\infty}\mu_j(\boldsymbol{x})\gamma^j. \tag{3.120}$$

Substituting the expansions (3.120) for ξ and μ into equation (3.118), and equating coefficients of γ^j, we obtain, for $j = 0$,

$$\begin{cases} |\nabla\xi_0|^2 + \mu_0|\nabla\mu_0|^2 = 1, \\ \nabla\xi_0\cdot\nabla\mu_0 = 0. \end{cases} \tag{3.121}$$

Regarding the functions $\mu_0(\boldsymbol{x})$ and $\xi_0(\boldsymbol{x})$, we make the following observations:

(i)

When $\boldsymbol{x} \in \Gamma$, since $\mu_0(\boldsymbol{x}) = 0$ by (3.115), vector $\nabla\mu_0$ must be parallel to $\boldsymbol{n}(\boldsymbol{x})$, the normal to the boundary at $\boldsymbol{x}$. But from the second equation in (3.121), we deduce that $\nabla\mu_0$ is perpendicular to $\nabla\xi_0$. Thus $\nabla\xi_0 = \boldsymbol{t}(\boldsymbol{x})$, where $\boldsymbol{t}(\boldsymbol{x})$ is the unit tangent vector. On Γ, we have

$$\xi_0(s) = s + c_0.$$

Here c_0 is a constant of integration that has to be taken to equal to zero on Γ for matching with the geometrical optics asymptotics in the domain I.

(ii)

Set

$$\tau_0^{\pm}(\boldsymbol{x}) = \xi_0(\boldsymbol{x}) \pm \frac{2}{3}\mu_0^{3/2}(\boldsymbol{x}), \tag{3.122}$$

then, cf (3.119), $\tau_0^{\pm}$ must satisfy

$$\begin{cases} \left|\nabla\tau_0^{\pm}(\boldsymbol{x})\right| = 1, \\ \tau_0^{\pm}(\boldsymbol{x}_0) = s \quad \text{for} \quad \boldsymbol{x}_0 \in \Gamma \end{cases} \tag{3.123}$$

With reference to Figure 3.15, we deduce in "normal" coordinates (s, n), from (3.123),

$$\tau_{\pm}(s, n) = s_0 \pm t(s, n). \tag{3.124}$$

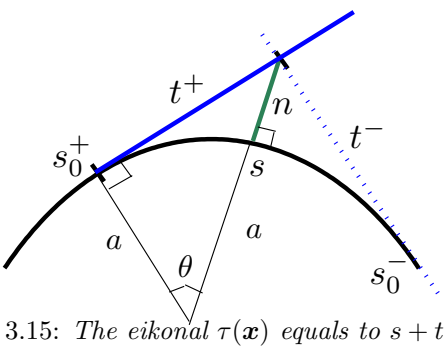


Figure 3.15: *The eikonal $\tau(\boldsymbol{x})$ equals to $s + t$.*

Lemma 3.23. *The function $\tau_{\pm}$ in (3.124) satisfies,*

$$\tau_{\pm}(s, n) = s \pm \frac{2}{3}\sqrt{2\kappa(s)}n^{3/2} + O\left(n^2\right), \qquad \text{as } n \to 0,$$

where $\kappa(s)$ denotes the curvature of the boundary at s, see (3.3.19) in [5].

Proof for the case of a cirle. From Figure 3.15 we deduce,

$$\begin{aligned}\tau_{\pm} &= s_0 \pm t \\ &= s - (s - s_0) \pm t.\end{aligned}$$

Note that

$$s - s_0 = \pm a\theta, \tag{3.125}$$

and

$$\begin{aligned}t &= (a+n)\sin\theta = (a+n)\left(\theta - \frac{\theta^3}{6} + O(\theta^5)\right) \\ &= a\theta + n\theta - \frac{1}{6}a\theta^3 + O\left(n^2\right). \end{aligned}\tag{3.126}$$

Now, from (3.125) and (3.126), we deduce,

$$\tau_{\pm} = s \pm [t - a\theta].$$

$$t - a\theta = n\theta - \frac{1}{6}a\theta^3 + O(n^2).$$

Since $\theta = \sqrt(2/a)n^{1/2}$ and $a = \kappa(s)^{-1}$, the result follows. □

Thus we can find the Eikonal, τ, in the transition zones and write it in the form

$$\tau(s,m) = s + \frac{2}{3}\sqrt{2\kappa(s)}m^3 + O(m^4), \tag{3.127}$$

where s is the parameter of the arc-length parametrisation of the boundary and $m^2 = n$ is the distance from the point on the diffracted ray to the boundary, along the normal to the boundary, see Figure 3.15. Define

$$\tau_e(s,m) = \frac{\tau(s,m) + \tau(s,-m)}{2}, \quad \text{and} \quad \tau_o(s,m) = \frac{\tau(s,m) - \tau(s,-m)}{2}.$$

Then,

$$\tau_e(s,m) = \xi(s,m) = s + O(m^4), \quad \text{and} \quad \tau_o(s,m) = \pm\frac{2}{3}\mu_o^{3/2}(s,m),$$

where

$$\mu_o(s,m) = (2\kappa(s))^{2/3}\, m^2 + O(m^4). \tag{3.128}$$

Returning to the main system of equations (3.118) with substituted expansions (3.120), we obtain by equating coefficients of γ^m, $m = 1$, the following system for ξ_1 and μ_1:

$$\begin{cases} \nabla\xi_0 \cdot \nabla\xi_1 + \mu_1|\nabla\mu_0|^2 + 2\mu_0\nabla\mu_0 \cdot \nabla\mu_1 = 0, \\ \nabla\xi_0 \cdot \nabla\mu_1 + \nabla\xi_1 \cdot \nabla\mu_0 = 0. \end{cases} \tag{3.129}$$

From (3.128), we deduce that when $\boldsymbol{x} \in \Gamma$

$$|\nabla \mu_0(\boldsymbol{x})|^2 = (2\kappa(\boldsymbol{x}))^{2/3}. \tag{3.130}$$

From (3.115), for $\boldsymbol{x} \in \Gamma$, $\mu_1(\boldsymbol{x}) = -1$. Therefore, $\nabla\mu_1$ is perpendicular to the tangent vector to Γ. This implies that $\nabla\xi_0(\boldsymbol{x})\cdot\nabla\mu_1(\boldsymbol{x}) = 0$ on Γ. Thus, from (3.129), $\nabla\xi_1\cdot\nabla\mu_0 = 0$, and hence

$$\xi_1(\boldsymbol{x}) = 2^{-1/3}\int_{\boldsymbol{x}_0}^{\boldsymbol{x}} \kappa(s)^{2/3} ds. \tag{3.131}$$

In a similar way, unknowns μ_1, ξ_2, μ_2,... can be obtained by solving the relevant recurrence relations and determining the constants of integration so that the ray asymptotic expansion is matched, see [5, Sections 13.2, 13.4].

We proceed to estimating the normal derivative $v(s,k)$.

3.4.3 The normal derivative of the solution.

We start by finding the normal derivative of the following term in the integrand in (3.108):

$$\begin{aligned}
&\nabla\left[\left(J(\boldsymbol{x},\gamma) - H(\boldsymbol{x},\gamma)\frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)}\right)e^{ik\xi(\boldsymbol{x},\gamma)}\right]\cdot\boldsymbol{n} = \\
&k^{-1/3}\left(\mathrm{Ai}(-k^{2/3}\mu) - \frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)}\mathrm{A}_+(-k^{2/3}\mu)\right)\nabla\mathcal{A}\cdot\boldsymbol{n}\; e^{ik\xi} \\
+\;& ik^{-2/3}\left(\mathrm{Ai}'(-k^{2/3}\mu) - \frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)}\mathrm{A}_+^{'}(-k^{2/3}\mu)\right)\nabla\mathcal{B}\cdot\boldsymbol{n}\; e^{ik\xi} \\
+\;& k^{-1}\left(\mathrm{Ai}'(-k^{2/3}\mu) - \frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)}\mathrm{A}_+^{'}(-k^{2/3}\mu)\right)(\boldsymbol{n}\cdot\nabla\mu)\mathcal{A}\; e^{ik\xi} \\
+\;& ik^{-4/3}\left(\mathrm{Ai}''(-k^{2/3}\mu) - \frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)}\mathrm{A}_+^{''}(-k^{2/3}\mu)\right)(\boldsymbol{n}\cdot\nabla\mu)\mathcal{B}\; e^{ik\xi} \\
+\;& \Bigg[k^{-1/3}\left(\mathrm{Ai}(-k^{2/3}\mu) - \frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)}\mathrm{A}_+(-k^{2/3}\mu)\right)\mathcal{A} \\
&\qquad + ik^{-2/3}\left(\mathrm{Ai}'(-k^{2/3}\mu) - \frac{\mathrm{Ai}(k^{2/3}\gamma)}{\mathrm{A}_+(k^{2/3}\gamma)}\mathrm{A}_+^{'}(-k^{2/3}\mu)\right)\mathcal{B}\Bigg]\, ik\,(\nabla\xi\cdot\boldsymbol{n})\, e^{ik\xi}
\end{aligned}$$

where $\mathcal{A}(\boldsymbol{x},\gamma,k)$ and $\mathcal{B}(\boldsymbol{x},\gamma,k)$ are defined in (3.111) and (3.112) respectively. Note that the first and the second to last terms vanish due to (3.115) and (3.116). Furthermore, the last term also vanishes when $\boldsymbol{x} \in \Gamma$ due to (3.115) and (3.116). The second and third terms simplify significantly for $\boldsymbol{x} \in \Gamma$ since for all t, the following identity holds due to Proposition 3.10,

$$\mathrm{Ai}'(t)\mathrm{A}_+(t) - \mathrm{Ai}(t)\mathrm{A}_+^{'}(t) = i.$$

This leads to the following representation of the normal derivative: for $\boldsymbol{x} \in \Gamma$,

$$\nabla u \cdot \boldsymbol{n}(\boldsymbol{x}) = k^{2/3} \int_L \frac{\tilde{B}(\boldsymbol{x}, k, \gamma)}{A_+(k^{2/3}\gamma)} e^{ik\xi(\boldsymbol{x},\gamma)} d\gamma, \tag{3.132}$$

where

$$\begin{aligned} \tilde{B}(\boldsymbol{x}, k, \gamma) &= \sum_{j=0}^{N} k^{-j} \left(iA_j(\boldsymbol{x}, \gamma)(\boldsymbol{n} \cdot \nabla \mu(\boldsymbol{x}, \gamma)) - k^{-1/3} \nabla B_j(\boldsymbol{x}, \gamma) \cdot \boldsymbol{n} \right) + R_N(\boldsymbol{x}, k) \\ &= \sum_{j=0}^{N} k^{-j} \tilde{B}_j(\boldsymbol{x}, \gamma) + R_N(\boldsymbol{x}, k, \gamma), \end{aligned}$$

where

$$\left| D_{\boldsymbol{x}}^{(n)} R_N(\boldsymbol{x}, k, \gamma) \right| \leq C_n \frac{1}{k^{N+1}}.$$

Although the functions $\tilde{B}_j(\boldsymbol{x}, \gamma)$ depend on k, we do not reflect this in the notation since $\tilde{B}_j(\boldsymbol{x}, \gamma) = O(1)$ as $k \to \infty$. Now, from (3.132), we deduce,

$$\nabla u \cdot \boldsymbol{n}(\boldsymbol{x}) = k^{-1/3} \sum_{j=0}^{N} k^{-j} \int_L \frac{\tilde{B}_j(\boldsymbol{x}, \gamma)}{\mathrm{A}_+(k^{2/3}\gamma)} e^{ik\xi(\boldsymbol{x},\gamma)} d\gamma + \tilde{R}_N(\boldsymbol{x}, k), \tag{3.133}$$

where

$$\left| D_{\boldsymbol{x}}^{(n)} \tilde{R}_N(\boldsymbol{x}, k) \right| \leq \tilde{C}_n \frac{1}{k^{N+1}}.$$

Now put

$$\tilde{B}_j(\boldsymbol{x}, \gamma) = \sum_{m=0}^{L} B_{j,m}(\boldsymbol{x}) \gamma^m + r_L(\boldsymbol{x}, \gamma),$$

where

$$r_L(\boldsymbol{x}, \gamma) = O\left(\frac{1}{\gamma^{L+1}} \right).$$

Thus, we deduce that

$$\begin{aligned} \nabla u \cdot \boldsymbol{n}(\boldsymbol{x}) &= \sum_{j=0}^{N} k^{-1/3-j} \left(\sum_{m=0}^{L} B_{j,m}(\boldsymbol{x}) \int_L \frac{\gamma^m e^{ik\xi(\boldsymbol{x},\gamma)}}{\mathrm{A}_+(k^{2/3}\gamma)} d\gamma + \int_L \frac{r_L(\boldsymbol{x}, \gamma) e^{ik\xi(\boldsymbol{x},\gamma)}}{\mathrm{A}_+(k^{2/3}\gamma)} d\gamma \right) \\ &+ R_N(\boldsymbol{x}, k). \end{aligned}$$

We introduce a change of variables: $z = k^{2/3}\gamma$, then

$$\begin{aligned} \nabla u \cdot \boldsymbol{n}(\boldsymbol{x}) &= \sum_{j=0}^{N} \sum_{m=0}^{L} k^{-1/3-2m/3-j} B_{j,m}(\boldsymbol{x}) \int_L \frac{r_L(\boldsymbol{x}, \gamma) e^{ik\xi(\boldsymbol{x}, k^{-2/3}z)}}{\mathrm{A}_+(z)} dz \\ &+ N \int_L \frac{r_L(\boldsymbol{x}, zk^{-2/3}) e^{ik\xi(\boldsymbol{x}, k^{-2/3}z)}}{\mathrm{A}_+(z)} dz + R_N(\boldsymbol{x}, k). \end{aligned} \tag{3.134}$$

We take $\xi(\boldsymbol{x},\gamma) = \xi_0(\boldsymbol{x}) + \xi_1(\boldsymbol{x})\gamma + O(\gamma^2)$ and then, using (3.131),

$$\begin{aligned}\int_L \frac{z^m e^{ik\xi(\boldsymbol{x},k^{-2/3}z)}}{\mathrm{A}_+(z)}dz &= \int_L \frac{z^m e^{ik\xi_0(\boldsymbol{x})}e^{ik\xi_1(\boldsymbol{x})k^{-2/3}z)}}{\mathrm{A}_+(z)}dz \\ &= e^{ik\xi_0(\boldsymbol{x})}\int_L \frac{z^m e^{ik^{1/3}\xi_1(\boldsymbol{x})z}}{\mathrm{A}_+(z)}dz = e^{ik\xi_0(\boldsymbol{x})}\int_L \frac{z^m e^{ik^{1/3}Z(\boldsymbol{x})z}}{\mathrm{A}_+(z)}dz,\end{aligned}$$

where

$$Z(s) = 2^{-1/3}\int_{\boldsymbol{x}_0}^{\boldsymbol{x}} \kappa(s)^{2/3}ds. \tag{3.135}$$

We can now rewrite the equation (3.134) as follows, for $\boldsymbol{x} \in \Gamma$ as $\boldsymbol{x} = \boldsymbol{x}(s)$:

$$v(s,k) = ke^{ik\boldsymbol{x}\cdot\boldsymbol{a}}\sum_{j=0}^{N}\sum_{m=0}^{L} k^{-1/3-2m/3-j}b_{j,m}(\boldsymbol{x})\Psi^{(m)}\left(k^{1/3}Z(s)\right) + R_{N,L}(\boldsymbol{x},k), \tag{3.136}$$

with

$$|D_s^n R_{L,N}(s,k)| \leq C_{L,N,n}(1+k)^{\mu+n/3},$$

and where $\mu := -\min\{2(L+1)/3, (N+1)\}$ and $C_{L,N,n}$ are independent of k and the Fock's integral $\Psi(\tau)$ is defined as,

$$\Psi(\tau) := \exp\left(-i\tau^3/3\right)\int_c \frac{e^{-iz\tau}}{\mathrm{Ai}(e^{2\pi i/3}z)}dz,$$

where the contour c is illistrated in Figure 3.1. The function $\Psi^{(l)}$ denotes the l-th derivative of Ψ and this integral converges exponentially due to asymptotic properties of Airy function Ai defined earlier in (3.31).

Chapter 4

Computation of highly-oscillatory double integrals

4.1 Introduction

In this chapter, we describe a numerical method for the efficient computation of the highly-oscillatory double integrals that arise from the Galerkin discretisation of the boundary integral equations for solving scattering problems which was described in Chapter 2.

We will often use terms *highly-oscillatory* and *slowly-varying* to describe functions or integrals. Thus, before we proceed to the main content of the chapter, we consider an example of slowly-varying and highly-oscillatory functions and integrals.

Let $I = [a, b]$ be a bounded interval. In the simplest case, a family of smooth functions $f_k : I \to \mathbb{R}$ depending on a parameter $k \in [1, \infty)$ is said to be *slowly-varying* if for any $j = 0, 1, 2, \ldots$

$$\left| f_k^{(j)}(x) \right| \leq C_j, \quad x \in I, \tag{4.1}$$

where C_j is a constant independent of k. On the other hand, a family of functions $F_k : I \to \mathbb{R}$ of the form

$$F_k(x) := f_k(x) \exp(ikx), \tag{4.2}$$

is said to be *highly-oscillatory*. Note that F_k satisfy

$$\left| F_k^{(j)}(x) \right| \geq C_j k^j, \quad x \in I, \tag{4.3}$$

where C_j is a constant independent of k.

A slowly-varying function can be well approximated by a polynomial interpolation at a suitable set of points independent of k with error remaining bounded as $k \to \infty$. However, this property does not hold for highly-oscillatory functions.

Similarly, *integrals* of slowly-varying functions can be accurately approximated using classical quadratures that are based on the polynomial approximation of the integrand. In this case, the number of quadrature points would be independent of k. However, the number of quadrature points required for the computation of highly-oscillatory integrals via classical

quadrature typically grows with k as $k \to \infty$ to achieve the same level of accuracy.

To illustrate and emphasize the difficulty of approximating highly-oscillatory integrals using classical quadratures consider the following example of error bounds for the cases when the same quadrature rule is applied to an integral with a slowly-varying and a highly-oscillatory integrand.

We consider the classical Clenshaw-Curtis quadrature described in Chapter 5 based on polynomial interpolation of the integrand at Clenshaw-Curtis points. Error estimates follow from Theorem 5.2 in Chapter 5.

$N+1$**-point classical Clenshaw-Curtis applied to a slowly-varying integrand**, i.e. to f_k satisfying (4.1),

$$I^{[a,b]}[f_k] := \int_a^b f_k(x)dx \simeq \int_a^b [\mathcal{P}_N f_k](x)dx =: Q_{1,N}^{classical,[a,b]}[f_k],$$

where $\mathcal{P}_N f_k$ is a polynomial of degree N satisfying $\mathcal{P}_N f_k(x_j) = f_k(x_j)$, where x_j, $j = 0, ..., N$ are Chebyshev points mapped to the interval $[a, b]$.

$$\left| I^{[a,b]}[f_k] - Q_{1,N}^{classical,[a,b]}[f_k] \right| \leq A_m \frac{(b-a)^{m+1}}{N^m} \sup_k \left\| f_k^{(m)} \right\|_{\infty,[a,b]}, \qquad m \geq 1, \tag{4.4}$$

where $A_m > 0$ are constants independent of k and N.

$N+1$**-point classical Clenshaw-Curtis applied to a highly-oscillatory integrand**, i.e. F_k satisfying (4.2),

$$I^{[a,b]}[F_k] := \int_a^b F_k(x)\exp(ikx)dx \simeq \int_a^b [\mathcal{P}_N F_k](x)dx := Q_{2,N}^{classical,[a,b]}[F_k],$$

$$\begin{aligned} \left| I^{[a,b]}[F_k] - Q_{2,N}^{classical,[a,b]}[F_k] \right| &\leq A_m \frac{(b-a)^{m+1}}{N^m} \left\| F_k^{(m)} \right\|_{\infty,[a,b]} \\ &\leq B_m \frac{(b-a)^{m+1} k^m}{N^m}, \end{aligned} \tag{4.5}$$

where $A_m > 0$ and $B_m > 0$ are constants independent of k and N. The error bound (4.4) shows that the classical quadrature applied to the slowly-varying integrand converges superalgebraically with N independently of k. On the other hand, the error bound (4.5) only shows convergence when the number of quadrature points N grows linearly with k as $k \to \infty$ or the interval (a, b) shrinks with $O(k^{-1})$.

Note that estimates (4.1) and (4.2) do not allow functions with singularities. However the notion of slowly- and highly-oscillatory can easily be extended to this case. We will consider the extension to singular functions as well as two dimensional functions with singularities when we encounter them later in this chapter.

4.1.1 Motivation of the chapter

In Chapter 2, a Galerkin boundary integral method for solving high-frequency acoustic scattering problems was described. This method requires the assembly and solution of a system of linear equations (2.39). The coefficients of this system of linear equations are double integrals. Assuming that the integrals can be computed exactly, the number of degrees of freedom of the Galerkin method only needs to grow at a rate slightly higher than $k^{1/9}$ in order to maintain the accuracy as $k \to \infty$. This follows from Theorem 6.5 in Chapter 2.

In practice, however, the double integrals can not be computed exactly and numerical methods are required for their approximation. The difficulty with the approximation of these integrals is that they are often highly-oscillatory. As described above, this means that the number of integration points required for their computation grows with k in practice at least quadratically if integration methods based on polynomial interpolation of the integrand are applied. The assembly of the system of linear equations (2.39) then becomes computationally unfeasible for large k. Moreover, the integrands of the double integrals often suffer from algebraic singularities. This fact must also be carefully considered for an accurate approximation of the integrals.

In this chapter, we present a numerical technique for the computation of the double integrals that arise from the Galerkin boundary integral method described in Chapter 2. The numerical method has the following properties:

- we can efficiently obtain an accuracy for the approximation of the double integrals so that the overall convergence rate of the theoretical Galerkin boundary integral method is not degraded by the application of the quadrature to capture the integrals,
- the number of quadrature points required to maintain this accuracy does not depend on the wavenumber k, and therefore the integration method is efficient for all wavenumbers k.

Our quadrature methods are devised by adapting Filon-type ideas, which are discussed in detail in Chapter 5. There Filon-type quadratures are introduced for integrals of the form:

$$I_k^{[a,b]}[f] = \int_a^b f(x) \exp(ikx) dx, \tag{4.6}$$

by replacing the function f by a suitable interpolating polynomial $P_N f$ of degree N and integrating exactly. One feature of these methods is that for a fixed number of quadrature points, Filon-type quadratures become more accurate as k increases provided the function f is sufficiently smooth. Moreover, and very important for our purposes, for fixed k the method converges superalgebraically with respect to N. In this sense good accuracy is obtained for both small and large k.

The oscillatory term in (4.6), $\exp(ikx)$, is called a *canonical oscillator*. On the other hand, if the frequency of the oscillator changes non-linearly with x, e.g. $\exp(ikg(x))$, we call such function a *non-canonical oscillator*.

Definition 4.1. *We denote highly-oscillatory integrals with non-canonical oscillator by* $I_k^{g,[a,b]}[f]$, *i.e.*

$$I_k^{g,[a,b]}[f] := \int_a^b f(x)\exp(ikg(x))dx. \tag{4.7}$$

Note

$$I_k^{[a,b]}[f] = I_k^{g,[a,b]}[f], \quad \text{when} \quad g(x) = x.$$

Before we proceed to describing the structure of the double integrals, we introduce the following notation that we use later for the derivatives of functions two variables.

Notation 4.2. *For* $g : [a,b] \times [c,d] \to \mathbb{R}$, *denote* $D^{\boldsymbol{p}}g(s,t)$, *with* $\boldsymbol{p} = (p_1, p_2)^T \in \mathbb{N}^2$, *as*

$$D^{\boldsymbol{p}}g(s,t) := \frac{\partial^{|\boldsymbol{p}|}}{\partial s^{p_1}\partial t^{p_2}} \ (g(s,t)) \tag{4.8}$$

where $|\boldsymbol{p}| = p_1 + p_2$.

Furthermore, we say g *is smooth, i.e.* $g \in C^\infty([a,b] \times [c,d])$, *if for any* $\boldsymbol{p} \in \mathbb{N}^2$, $D^{\boldsymbol{p}}g(s,t)$ *exist and is continuous.*

4.1.2 The structure of the double integrals

In this section, we discuss the structure of the double integrals arising from the Galerkin discretisation (2.39). Recall that the Galerkin method of Chapter 2 leads to integrals of the following form, (see equation (6.26b)),

$$J_k := \int_\Lambda \int_{\Lambda'} g(s,t) \left[\partial_{\boldsymbol{n}(s)} \Phi_k(\boldsymbol{\gamma}(s), \boldsymbol{\gamma}(t)) - ik\Phi_k(\boldsymbol{\gamma}(s), \boldsymbol{\gamma}(t))\right] \exp\left(ik\boldsymbol{a} \cdot [\boldsymbol{\gamma}(t) - \boldsymbol{\gamma}(s)]\right) dtds, \tag{4.9}$$

where Λ and Λ' are two subintervals of $[0, 2\pi]$ and $g \in C^\infty([0, 2\pi] \times [0, 2\pi])$ is independent of k and

$$\Phi_k(\boldsymbol{x}, \boldsymbol{y}) = \frac{ik}{4} H_0^{(1)}\left(k\left|\boldsymbol{x} - \boldsymbol{y}\right|\right), \tag{4.10}$$

is the fundamental solution of the Helmholtz equation in 2D.

The first step in computing J_k is to separate the highly-oscillatory part of the integrand from the slowly-varying part. We do this in Theorem 4.3.

Theorem 4.3. *The double integral, J_k, defined in (4.9) can be written in the form:*

$$J_k = \int_\Lambda \int_{\Lambda'} M(s,t) \exp\left(ik\Psi(s,t)\right) dtds, \tag{4.11}$$

where the phase function $\Psi : \Lambda \times \Lambda' \to \mathbb{R}$ is given by

$$\Psi(s,t) = \left|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)\right| - \boldsymbol{a} \cdot \left(\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)\right), \tag{4.12}$$

and the function $M : \Lambda \times \Lambda' \to \mathbb{R}$, is defined as

$$M(s,t) := g(s,t) \left[\partial_{\boldsymbol{n}(s)} \Phi_k(\boldsymbol{\gamma}(s), \boldsymbol{\gamma}(t)) - ik\Phi_k(\boldsymbol{\gamma}(s), \boldsymbol{\gamma}(t))\right] \exp\left(-ik\left|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)\right|\right), \tag{4.13}$$

satisfies for $s \neq t$

$$|M(s,t)| \leq C_1 k \max\left\{1, \log\left(\frac{1}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|}\right)\right\}, \tag{4.14}$$

where $C_1 > 0$ is independent of k. Furthermore, for all $\boldsymbol{p} \in \mathbb{N}^2$,

$$|D^{\boldsymbol{p}} M(s,t)| \leq C_2 k \frac{1}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|^{|\boldsymbol{p}|}}, \tag{4.15}$$

where $C_2 > 0$ is independent of k.

Note that the bounds on the function M given in (4.14) grow linearly with k. This is due to the factor of k that appears in the function Φ_k defined in (4.10). However, the derivatives of M do not produce any additional powers of k.

The proof of Theorem 4.3 requires several intermediate results that we present in Lemma 4.5, Lemma 4.6 and Lemma 4.7. For convenience, we introduce the following notation,

Notation 4.4. *For $m \geq 0$, we define*

$$h_m(z) = \exp(-iz)H_m^{(1)}(z), \qquad z > 0. \tag{4.16}$$

The function $h_m(z)$ is singular at $z = 0$ due to the singularity in the Hankel function $H_m^{(1)}(z)$.

In Lemma 4.5, we bound $h_0(z)$, $z > 0$ by another function that has an explicit singularity for $z > 0$. Furthermore, we derive estimates for the derivatives of $h_0(z)$ for $z > 0$.

Lemma 4.5. *There exists $C > 0$ such that*

$$|h_0(z)| \leq C \begin{cases} 1 - \log z, & z \in (0, 1], \\ z^{-1/2}, & z \in [1, \infty). \end{cases} \tag{4.17}$$

Moreover, for each $n \geq 1$, there exist $C_n > 0$ such that

$$\left| \left(\frac{d}{dz} \right)^n h_0(z) \right| \leq C_n \begin{cases} z^{-n}, & z \in (0, 1], \\ z^{-(n+1/2)}, & z \in [1, \infty). \end{cases} \tag{4.18}$$

Proof. We use the integral representation of Hankel function [81, Section 13.3]:

$$H_0^{(1)}(z) = -\frac{2i}{\pi} \exp(iz) \int_0^\infty \frac{\exp(-zt)}{t^{1/2}(t - 2i)^{1/2}} dt, \quad z > 0, \tag{4.19}$$

where $(t - 2i)^{1/2}$ is chosen with positive real part for $t > 0$. The latter implies

$$|t - 2i|^{1/2} \geq \max\{t^{1/2}, 2^{1/2}\}. \tag{4.20}$$

The representation (4.19) yields

$$h_0(z) = -\frac{2i}{\pi} \int_0^\infty \frac{\exp(-zt)}{t^{1/2}(t - 2i)^{1/2}} dt.$$

Therefore, for $n \geq 0$,

$$\left(\frac{d}{dz} \right)^n h_0(z) = (-1)^{n+1} \frac{2i}{\pi} \int_0^\infty \frac{\exp(-zt) t^{n-1/2}}{(t - 2i)^{1/2}} dt. \tag{4.21}$$

We now write the integral (4.21) as a sum of two integrals:

$$\begin{aligned}\left(\frac{d}{dz}\right)^n h_0(z) &= (-1)^{n+1}\frac{2i}{\pi}\left(\int_0^1 \frac{\exp(-zt)t^{n-1/2}}{(t-2i)^{1/2}}dt + \int_1^\infty \frac{\exp(-zt)t^{n-1/2}}{(t-2i)^{1/2}}dt\right)\\ &=: (-1)^{n+1}\frac{2i}{\pi}\left(I_1(z)+I_2(z)\right).\end{aligned}$$

Using (4.20) and then changing variables $y = zt$ we deduce,

$$|I_1(z)| \leq \frac{1}{\sqrt{2}}\int_0^1 \exp(-zt)t^{n-1/2}dt = \frac{1}{\sqrt{2}z^{n+1/2}}\int_0^z \exp(-y)y^{n-1/2}dy, \tag{4.22}$$

$$|I_2(z)| \leq \int_1^\infty \exp(-zt)t^{n-1}dt = \frac{1}{z^n}\int_z^\infty \exp(-y)y^{n-1}dy. \tag{4.23}$$

For the case when $n = 0$, the integral $I_1(z)$ is bounded as follows,

$$|I_1(z)| \leq C\begin{cases}1, & z \in (0,1],\\ z^{-1/2}, & z \in [1,\infty),\end{cases} \tag{4.24}$$

where $C > 0$. Similarly $I_2(z)$ for $z \leq 1$ is estimated by

$$\begin{aligned}|I_2(z)| &\leq \frac{1}{\sqrt{2}}\int_z^1 \frac{\exp(-y)}{y}dy + \frac{1}{\sqrt{2}}\int_1^\infty \frac{\exp(-y)}{y}dy \leq \int_z^1 \frac{1}{y}dy + \int_1^\infty \exp(-y)dy\\ &\leq C_1\log\left(\frac{1}{z}\right) + C_2,\end{aligned} \tag{4.25}$$

where $C_1 > 0$ and $C_2 > 0$. On the other hand, for $z \geq 1$,

$$|I_2(z)| \leq \int_1^\infty \frac{\exp(-y)}{y}dy \leq C, \tag{4.26}$$

where $C > 0$. Hence the result (4.17) follows from estimates (4.24) - (4.26).

The case when $n \geq 1$, i.e. the result (4.18), follows from estimates (4.22) and (4.23) by considering cases when $z \leq 1$ and $z > 1$ separately. □

In Lemma 4.6, we bound $zh_1(z)$, $z > 0$ and its derivatives using Lemma 4.5.

Lemma 4.6. *For the function $h_1(z)$ defined as in (4.16), the following estimate holds:*

$$|zh_1(z)| \leq C\begin{cases}1, & z \in (0,1]\\ z^{1/2}, & z \in [1,\infty)\end{cases}. \tag{4.27}$$

Moreover, the derivatives are bounded by,

$$\left|\left(\frac{d}{dz}\right)^n (zh_1(z))\right| \leq C \begin{cases} z^{-n}, & z \in (0,1] \\ z^{-(n-1/2)}, & z \in [1,\infty) \end{cases}, \quad n \geq 1. \tag{4.28}$$

Proof. Since

$$\frac{d}{dz}H_0^{(1)}(z) = -H_1^{(1)}(z),$$

we find the function $h_1(z)$ is related to $h_0(z)$ by

$$h_1(z) = -ih_0(z) - \frac{d}{dz}h_0(z).$$

Therefore, $zh_1(z) = -izh_0(z) - z((d/dz)h_0(z))$. Then to see (4.27), for $z \in (0,1]$, we use Lemma 4.5 and the fact that $\lim_{z\to 0} z\log(z) = 0$.

On the other hand, to show (4.28), we use the identity,

$$\left(\frac{d}{dz}\right)^n (zf(z)) = n\left(\frac{d}{dz}\right)^{n-1} f(z) + z\left(\frac{d}{dz}\right)^n f(z),$$

to deduce,

$$\begin{aligned} \left|\left(\frac{d}{dz}\right)^n (zh_1(z))\right| &\leq \left|n\left(\frac{d}{dz}\right)^{n-1} h_0(z)\right| + \left|z\left(\frac{d}{dz}\right)^n h_0(z)\right| \\ &+ \left|n\left(\frac{d}{dz}\right)^n h_0(z)\right| + \left|z\left(\frac{d}{dz}\right)^{n+1} h_0(z)\right|. \end{aligned}$$

Finally, using Lemma 4.5, we obtain (4.28). □

In the following lemma, we prove an intermediate result required for the proof of Theorem 4.3.

Lemma 4.7. *Define the function* $L : [0,2\pi] \times [0,2\pi] \to \mathbb{R}$ *as,*

$$L(s,t) := \frac{(\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)) \cdot \boldsymbol{n}(s)}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|^2}, \tag{4.29}$$

where $\boldsymbol{\gamma}$ *is a smooth parametrisation of a 2D contour* Γ*, i.e.* $\boldsymbol{\gamma} \in C^\infty([0,2\pi]) \times C^\infty([0,2\pi])$*. Then* $L \in C^\infty([0,2\pi] \times [0,2\pi])$*.*

Proof. Clearly L is a C^∞ function for any t outside a neighborhood of s. To investigate what happens when t near s, note that for $i = 1,2$, we may write,

$$\boldsymbol{\gamma}_i(s) - \boldsymbol{\gamma}_i(t) = \int_0^1 \frac{d}{d\lambda}\left\{\boldsymbol{\gamma}_i(t + \lambda(s-t))\right\} d\lambda.$$

Then, integrating by parts, we have,

$$\begin{aligned}\gamma_i(s)-\gamma_i(t) &= \left[\lambda\frac{d}{d\lambda}\left\{\gamma_i(t+\lambda(s-t))\right\}\right]_{\lambda=0}^{\lambda=1}-\int_0^1\lambda\frac{d^2}{d\lambda^2}\left\{\gamma_i(t+\lambda(s-t))\right\}d\lambda\\ &= \gamma_i'(s)(s-t)-(s-t)^2\int_0^1\lambda\gamma_i''(t+\lambda(s-t))d\lambda\\ &= \gamma_i'(s)(s-t)-G(s,t)(s-t)^2,\end{aligned}$$

where $G\in C^\infty([0,2\pi]\times[0,2\pi])$. Therefore,

$$\boldsymbol{\gamma}(s)-\boldsymbol{\gamma}(t)=\boldsymbol{\gamma}'(s)(s-t)+G(s,t)(s-t)^2, \tag{4.30}$$

with $G\in C^\infty([0,2\pi]\times[0,2\pi])$. Hence, since $\boldsymbol{\gamma}'(s)\cdot\boldsymbol{n}(s)=0$, we have,

$$(\boldsymbol{\gamma}(s)-\boldsymbol{\gamma}(t))\cdot\boldsymbol{n}(s)=\left[G(s,t)\cdot\boldsymbol{n}(s)\right](s-t)^2,$$

and

$$|\boldsymbol{\gamma}(s)-\boldsymbol{\gamma}(t)|^2=|\boldsymbol{\gamma}'(s)|^2(s-t)^2+2\left[\boldsymbol{\gamma}'(s)\cdot G(s,t)\right](s-t)^3+|G(s,t)|^2(s-t)^4. \tag{4.31}$$

Thus,

$$L(s,t)=\frac{G(s,t)\cdot\boldsymbol{n}(s)}{|\boldsymbol{\gamma}'(s)|^2+2\left[\boldsymbol{\gamma}'(s)\cdot G(s,t)\right](s-t)+|G(s,t)|^2(s-t)^2}.$$

Since $|\boldsymbol{\gamma}'(s)|>0$, it follows that L is also C^∞ in a sufficiently small neighborhood of $t=s$. □

Lemma 4.8. *If $\boldsymbol{\gamma}$ is a smooth parammetrisation of a 2D contour Γ and a function r is defined as*

$$r(s,t):=|\boldsymbol{\gamma}(s)-\boldsymbol{\gamma}(t)|, \tag{4.32}$$

then for each $\boldsymbol{p}\in\mathbb{N}^2$ there exists $C_{\boldsymbol{p}}>0$ such that

$$\left|D^{\boldsymbol{p}}_{(s,t)}r(s,t)\right|\le C_{\boldsymbol{p}}r(s,t)^{1-|\boldsymbol{p}|}, \tag{4.33}$$

where $D^{\boldsymbol{p}}_{(s,t)}$ is defined as in (4.8),

$$D^{\boldsymbol{p}}_{(s,t)}r(s,t):=\frac{\partial^{|\boldsymbol{p}|}}{\partial s^{p_1}\partial t^{p_2}}\ (r(s,t)).$$

Proof. See Section F of Appendix. □

Remark 4.9. *Although, the first derivatives of the function r are bounded, they are not continuous at $s=t$. For example let us consider the partial derivative of r with respect to*

the variable s,

$$\frac{\partial}{\partial s} r(s,t) = \frac{(\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)) \cdot \boldsymbol{\gamma}'(s)}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|}.$$

By (4.30) and (4.31), we have

$$(\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)) \cdot \boldsymbol{\gamma}'(s) = -(s-t)\left(|\boldsymbol{\gamma}'(s)|^2 + (s-t)G(s,t) \cdot \boldsymbol{\gamma}'(s)\right),$$

and

$$|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)| = |s-t| \left\{|\boldsymbol{\gamma}'(s)|^2 + 2\left[\boldsymbol{\gamma}'(s) \cdot G(s,t)\right](s-t) + |G(s,t)|^2 (s-t)^2\right\}^{1/2},$$

so

$$\lim_{t \to s+} \left\{ \frac{(\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)) \cdot \boldsymbol{\gamma}'(s)}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|} \right\} = -1 = -\lim_{t \to s-} \left\{ \frac{(\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)) \cdot \boldsymbol{\gamma}'(s)}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|} \right\}.$$

Let us now return to the proof of Theorem 4.3.

Proof of Theorem 4.3

Proof. For convenience assume $k \in [1, \infty)$. The formulae (4.11) -(4.12) are obtained by straightforward algebra. It remains to show that function M satisfies (4.14) and (4.15). Recall the definition of the function M in (4.11),

$$M(s,t) = g(s,t)\mathcal{K}(s,t)\exp(-ikr(s,t)), \tag{4.34}$$

where $g \in C^\infty([0,2\pi] \times [0,2\pi])$ and

$$\begin{aligned}\mathcal{K}(s,t) &:= \left[\partial_{\boldsymbol{n}(s)} \Phi_k(\boldsymbol{\gamma}(s), \boldsymbol{\gamma}(t)) - ik\Phi_k(\boldsymbol{\gamma}(s), \boldsymbol{\gamma}(t))\right] \\ &= -\frac{ik}{4} H_1^{(1)}\left(kr(s,t)\right) \frac{(\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)) \cdot \boldsymbol{n}(s)}{r(s,t)} + \frac{k}{4} H_0^{(1)}\left(kr(s,t)\right).\end{aligned}$$

Since g in (4.34) is smooth, we require only bounds on the derivatives of

$$\begin{aligned}\mathcal{K}(s,t)\exp(-ikr(s,t)) &:= -\frac{ik}{4} h_1(kr(s,t)) \frac{(\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)) \cdot \boldsymbol{n}(s)}{r(s,t)} + \frac{k}{4} h_0(kr(s,t)) \\ &= -\frac{ik}{4}\left[h_1(kr(s,t))r(s,t)\right] L(s,t) + \frac{k}{4} h_0(kr(s,t)),\end{aligned} \tag{4.35}$$

where functions h_0 and h_1 are defined in (4.16) and L is defined in (4.29) and proved to be smooth in Lemma 4.7.

From Lemma 4.5, we deduce, for $0 < kr(s,t) \leq 1$,

$$|h_0\left(kr(s,t)\right)| \lesssim \log\left(\frac{1}{kr(s,t)}\right) + 1 \lesssim \log\left(\frac{1}{r(s,t)}\right) + 1, \tag{4.36}$$

since $\log(1/x)$ is a decreasing function and $kr(s,t) \geq r(s,t)$, where the notation $A \lesssim B$ in (4.36) means $A \leq cB$, where c is independent of k.

On the other hand, when $kr(s,t) > 1$,

$$|h_0\left(kr(s,t)\right)| \lesssim 1, \tag{4.37}$$

Combining equations (4.36) and (4.37), we obtain

$$|h_0\left(kr(s,t)\right)| \lesssim \max\left\{1, \log\left(\frac{1}{r(s,t)}\right)\right\}, \qquad s,t \in [0,2\pi]. \tag{4.38}$$

Let us now investigate the bounds on derivatives of $h_0\left(kr(s,t)\right)$. For this, we require two-dimensional Faa di Bruno's formula in it's combinatorical form [1].

For integers $p_1, p_2 \geq 0$, we introduce the set

$$S := \{1, \ldots, p_1, p_1+1, \ldots, p_1+p_2\}.$$

We denote $\boldsymbol{p} = (p_1, p_2)^T$. A *partition* of S is a set (here denoted by σ) of non-empty, non-overlapping subsets of S whose union is all of S. The number of sets in σ is denoted $|\sigma|$. For each $\beta \in \sigma$ we let β_1 denote the number of elements of β which are $\leq p_1$ and let β_2 denote the number of elements of β which are $> p_1$ and $\leq p_1 + p_2$. Denote the number of elements in σ by $|\sigma|$ and denote $|\beta|$ as $\beta_1 + \beta_2$.

Then, using Faa di Bruno's formula [55, 35], we deduce,

$$\left(\frac{\partial}{\partial s}\right)^{p_1}\left(\frac{\partial}{\partial t}\right)^{p_2}\{h_0(kr(s,t))\} = \sum_{\sigma} h_0^{|\sigma|}(kr(s,t)) \prod_{\beta\in\sigma}\left(\frac{\partial}{\partial s}\right)^{\beta_1}\left(\frac{\partial}{\partial t}\right)^{\beta_2}\{kr(s,t)\}, \tag{4.39}$$

where the sum is over all possible partitions σ of S.

For example, if $p_1 = 1 = p_2$, then $S = \{1,2\}$ and the possible partitions of S are:

$$\sigma = \{\{1\},\{2\}\} \qquad \text{and} \qquad \sigma = \{\{1,2\}\},$$

[1] **Faa di Bruno's formula in 2D** [55, 35]. For a composite function $F \circ G$, regardless of whether $x_1, \ldots, x_p$ are all distinct or identical, the following holds,

$$\frac{\partial^p}{\partial x_1 \ldots \partial x_p} F(G) = \sum_{\sigma\in S} F^{(|\sigma|)}(G) \prod_{\beta\in\sigma} \frac{\partial^{|\beta|} G}{\prod_{j\in\beta} \partial x_j},$$

where σ denotes a set of all possible partitions of the set $S = \{1,2,\ldots,p\}$. Here σ is a set of non-empty, non-overlapping subsets of S whose union is all of S. We denote the number of elements in σ by $|\sigma|$. For $\beta \in \sigma$, we denote $|\beta|$ as the number of elements in β.

with $|\sigma| = 2$ and $|\sigma| = 1$ respectively. The right hand side of (4.39) is

$$h_0^{(2)}\left(kr(s,t)\right)\frac{\partial}{\partial s}\left\{kr(s,t)\right\} + h_0^{(1)}\left(kr(s,t)\right)\frac{\partial}{\partial s}\frac{\partial}{\partial t}\left\{kr(s,t)\right\},$$

which is easily seen to equal the left hand side.

Now, consider a typical term in the sum (4.39),

$$\begin{aligned}\left|\prod_{\beta\in\sigma}\left(\frac{\partial}{\partial s}\right)^{\beta_1}\left(\frac{\partial}{\partial t}\right)^{\beta_2}\left\{kr(s,t)\right\}\right| &= k^{|\sigma|}\left|\prod_{\beta\in\sigma}\left(\frac{\partial}{\partial s}\right)^{\beta_1}\left(\frac{\partial}{\partial t}\right)^{\beta_2}\left\{r(s,t)\right\}\right| \quad \text{by (4.33)}\\ &\leq k^{|\sigma|}\prod_{\beta\in\sigma} r(s,t)^{1-|\beta|}\\ &= Ck^{|\sigma|}r(s,t)^{|\sigma|}r(s,t)^{-|\boldsymbol{p}|}, \end{aligned} \tag{4.40}$$

with C independent of k, s and t. Also, by Lemma 4.5,

$$\left|h^{(|\sigma|)}(kr(s,t))\right| \leq C(kr(s,t))^{-|\sigma|}. \tag{4.41}$$

Therefore, a typical term in (4.39) is bounded by

$$Ck^{-|\sigma|}(r(s,t))^{-|\sigma|}k^{|\sigma|}r(s,t)^{|\sigma|}r(s,t)^{-|\boldsymbol{p}|} = Cr(s,t)^{-|\boldsymbol{p}|},$$

which proves

$$D^{\boldsymbol{p}}\left\{h_0(kr(s,t))\right\} \lesssim r(s,t)^{-|\boldsymbol{p}|}. \tag{4.42}$$

On the other hand, from Lemma 4.6, we obtain, for $0 < kr(s,t) \leq 1$,

$$\left|h_1\left(kr(s,t)\right)r(s,t)\right| \lesssim \frac{1}{k} \lesssim 1, \tag{4.43}$$

while for $kr(s,t) > 1$,

$$\begin{aligned}\left|h_1\left(kr(s,t)\right)r(s,t)\right| &= \frac{1}{k}\left[h_1\left(kr(s,t)\right)kr(s,t)\right]\\ &\lesssim \frac{1}{k}\left(kr(s,t)\right)^{1/2} \lesssim 1. \end{aligned} \tag{4.44}$$

Combining (4.43) and (4.44), we deduce

$$\left|h_1\left(kr(s,t)\right)r(s,t)\right| \lesssim 1, \qquad s,t \in [0,2\pi]. \tag{4.45}$$

To estimate the derivatives, first note

$$D^{\boldsymbol{p}}\left\{h_1(kr(s,t))r(s,t)\right\} = \frac{1}{k}D^{\boldsymbol{p}}_{(s,t)}\left\{l(kr(s,t))\right\},$$

where $l(z) := zh_1(z)$.

Applying formula (4.39) with h_0 replaced by l and applying analogous estimates using Lemma 4.6, we obtain a sum of terms each of which can be estimated by,

$$\frac{1}{k}(kr(s,t))^{-|\sigma|+\frac{1}{2}}k^{|\sigma|}r(s,t)^{|\sigma|}r(s,t)^{-|\boldsymbol{p}|} = \frac{1}{k}r(s,t)^{-|\boldsymbol{p}|} \lesssim r(s,t)^{-|\boldsymbol{p}|},$$

which proves

$$|D^{\boldsymbol{p}}\left[h_1\left(kr(s,t)\right)r(s,t)\right]| \lesssim (r(s,t))^{-|\boldsymbol{p}|}, \qquad s,t \in [0,2\pi]. \tag{4.46}$$

Using (4.38) and (4.45) in (4.35) and then substituting the estimates in (4.34), we obtain the bound on $|M(s,t)|$, $s,t \in [0,2\pi]$, hence proving (4.14).

Finally, using (4.42) and (4.46) in (4.35) and then in (4.34), we obtain the bounds on the derivatives of M, therefore proving (4.15). □

Theorem 4.3 implies that that the double integral J_k is highly-oscillatory with the integrand function M that has a logarithmic singularity. Moreover, J_k has a non-canonical oscillator. All these factors make it difficult to approximate J_k accurately and efficiently.

4.1.3 Outline of the remainder of the chapter

The plan for this chapter is as follows. In Section 4.2 we discuss an abstract methodology for the computation of single highly-oscillatory integrals with non-canonical oscillators using Filon-type quadratures. Also in this section, we demonstrate how this idea can be extended to double integrals with non-canonical oscillators. In Section 4.3, we apply the idea to the integrals of the form (4.11) arising in scattering problems.

We demonstrate how the highly-oscillatory integral J_k defined in (4.11) can be written as a repeated integral where the outer integral is highly-oscillatory with a canonical oscillator, i.e. can be written in the form (4.6), and the inner integral is slowly-varying. Adaptation of this idea to double integrals such as (4.11) is a novel result of this chapter. In the next chapter, we obtain explicit error bounds for the method.

For the case when the phase-function Ψ in (4.11) has stationary points in the domain of integration, the integration method described in Section 4.3 is not applicable. In Section 4.4, we describe the stationary points of the phase-function Ψ and investigate the behaviour of the integrand in (4.11) around these points.

Finally, in Section 4.4.1, we describe the methodology for the computation of integrals (4.11) in domains containing stationary points.

4.2 Methodology for the computation of highly-oscillatory integrals

In this section, we develop a methodology for the computation of highly-oscillatory integrals of the form:

$$I_k^{\psi,[a,b]}[f] := \int_a^b f(s)\exp(ik\psi(s))ds, \tag{4.47}$$

using classical Filon-type quadratures, where the function f is a slowly-varying function and the phase-function ψ is smooth. We also demonstrate how to compute double integrals of the form,

$$J_k^{\Psi,D}[M] := \int\int_D M(s,t)\exp(ik\Psi(s,t))dtds, \tag{4.48}$$

where the function M is slowly-varying and the phase-function Ψ is a smooth function satisfying certain properties, and where D is ether a rectangular or a triangular domain.

4.2.1 Abstract methodology for single integrals

One class of methods specifically designed for the computation of highly-oscillatory integrals is Filon-type methods. Classical Filon-type quadratures are designed to compute integrals of the form:

$$I_k^{[a,b]}[f] := \int_a^b f(s)\exp(iks)ds, \tag{4.49}$$

by replacing the slowly-varying function f by a suitable interpolating polynomial $P_N f$ of degree N and integrating exactly. Filon-type quadratures require computation of the so-called *moments* $\mu_n(k)$ defined as $\mu_n(k) := I_k^{[a,b]}[p_n]$ where p_n is a suitable polynomial basis. Clearly, the moments can be computed exactly for all n and k in the case of the canonical oscillator.

However, when the oscillating term in the integrand is non-canonical, i.e. $\exp(ik\psi(s))$:

$$I_k^{\psi,[a,b]}[f] := \int_a^b f(s)\exp(ik\psi(s))ds, \tag{4.50}$$

it is generally not possible to compute the moments $\mu_n^{\psi}(k) := I_k^{\psi,[a,b]}[p_n]$ exactly or express them in terms of special functions.

Let us first consider computing (4.50) when the phase-function ψ does not have any stationary points in $[a,b]$. Later we consider the case when stationary points are present in $[a,b]$.

The case when the phase function does not have stationary points

In order to avoid computing moments the following technique can be applied: the integral with non-canonical oscillator (4.50) is transformed into an integral with the canonical oscillator (4.49) with the change of variable,

$$\tau = \psi(s). \tag{4.51}$$

The change of variable is valid since there are no stationary points in $[a, b]$, i.e.

$$\psi'(s) \neq 0 \quad \text{for all} \quad s \in [a, b]. \tag{4.52}$$

The change of variable yields the Jacobian in the integrand:

$$ds = \frac{1}{|\psi'(\psi^{-1}(\tau))|} d\tau, \qquad \text{with} \quad s = \psi^{-1}(\tau).$$

Then we can write,

$$\begin{aligned} I_k^{\psi,[a,b]}[f] : &= \int_a^b f(s) \exp(ik\psi(s)) ds \\ &= \int_{\psi(a)}^{\psi(b)} \frac{f(\psi^{-1}(\tau))}{|\psi'(\psi^{-1}(\tau))|} \exp(ik\tau) d\tau \\ &= \int_{\psi(a)}^{\psi(b)} F(\tau) \exp(ik\tau) d\tau, \end{aligned} \tag{4.53}$$

where

$$F(\tau) := \frac{f\left(\psi^{-1}(\tau)\right)}{|\psi'\left(\psi^{-1}(\tau)\right)|}.$$

The resulting integral in (4.53) can be computed using classical Filon-type quadrature. The additional cost is the evaluation of the inverse function ψ^{-1}.

Some observations about the regularity of F can be made if certain properties of f and ψ are known. For example, if both functions f and ψ are smooth and (4.52) holds then the function F is smooth in $[\psi(a), \psi(b)]$.

On the other hand, if the condition (4.52) is not satisfied, i.e. the phase-function has stationary points, then the change of variables (4.51) can not be applied on the whole interval $[a, b]$.

The case when the phase function has stationary points

Let us consider in more detail the case when the phase-function has a stationary point $\xi \in (a,b)$ of order n:

$$\psi'(\xi) = \psi''(\xi) = \ldots = \psi^{(n)}(\xi) = 0, \tag{4.54}$$

$$\psi^{(n+1)}(\xi) \neq 0, \tag{4.55}$$

and $\psi'(s) \neq 0$, $s \in [a,b] \setminus \{\xi\}$.

Then, the change of variables (4.51) can be applied on two intervals $[a,\xi)$ and $(\xi,b]$ separately. The function ψ is strictly monotone in $[a,\xi]$ and in $[\xi,b]$. Without loss of generality, assume ξ is a minimum. Then, we write

$$\begin{aligned} I_k^{\psi,[a,b]}[f] &:= \int_a^{\xi} f(s)\exp(ik\psi(s))ds + \int_{\xi}^{b} f(s)\exp(ik\psi(s))ds \\ &= \int_{\psi(\xi)}^{\psi(a)} F_1(\tau)\exp(ik\tau)d\tau + \int_{\psi(\xi)}^{\psi(b)} F_2(\tau)\exp(ik\tau)d\tau, \end{aligned} \tag{4.56}$$

where $F_1 : [\psi(\xi),\psi(a)] \to \mathbb{R}$ and $F_2 : [\psi(\xi),\psi(b)] \to \mathbb{R}$ are defined by

$$F_1(\tau) := -\, f\left(\psi^{-1}(\tau)\right)\left(\psi^{-1}\right)'(\tau), \qquad \text{and} \qquad F_2(\tau) := f\left(\psi^{-1}(\tau)\right)\left(\psi^{-1}\right)'(\tau). \tag{4.57}$$

In the following theorem, we provide estimates for the derivatives of $F_1(\tau)$ and $F_2(\tau)$.

Theorem 4.10. *Assuming the functions ψ and f in (4.57) are smooth, the derivatives of F_1 and F_2 are bounded as follows:*

$$\left|F_1^{(p)}(\tau)\right| \leq C_{1,p}\,|\tau - \psi(\xi)|^{\frac{1}{n+1}-p-1}, \qquad \tau \in [\psi(\xi),\psi(a)] \tag{4.58}$$

and

$$\left|F_2^{(p)}(\tau)\right| \leq C_{2,p}\,|\tau - \psi(\xi)|^{\frac{1}{n+1}-p-1}, \qquad \tau \in [\psi(\xi),\psi(b)], \tag{4.59}$$

where $C_{1,p}$ and $C_{2,p}$ are positive constants independent of n.

The proof of Theorem 4.10 requires Lemma 4.11 presented below.

Lemma 4.11. *The derivatives of ψ^{-1} are bounded as follows,*

$$\left|\left(\psi^{-1}\right)^{(j)}(\tau)\right| \leq C_j\,|\tau - \psi(\xi)|^{\frac{1}{n+1}-j}, \tag{4.60}$$

for all $j = 1,2,\ldots$, where C_j are constants independent of τ and n.

We prove lemma for the case $j = 1$. Following the proof, we give an example where we consider a particular function ψ and verify Lemma 4.11 for all j using the chain rule.

Proof. Choose $\delta > 0$ such that $I_\delta := (\xi, \xi + \delta) \subset [a, b]$. We use Taylor's expansion of $\psi(s)$ around ξ,

$$\psi(s) = \psi(\xi) + (s - \xi)\psi'(\xi) + \ldots + \frac{(s-\xi)^n}{n!}\psi^{(n)}(\xi) + R_\xi(s), \tag{4.61}$$

for $s \in I_\delta$, where the remainder term is given in the integral form, [97, Theorem 4, Chapter 19],

$$R_\xi(s) = \frac{1}{n!}\int_\xi^s (s-t)^n \psi^{(n+1)}(t)dt. \tag{4.62}$$

By making the change of variables $y = (t - \xi)/(s - \xi) \in [0, 1]$ for $t \in [\xi, s]$ in (4.62), we obtain

$$R_\xi(s) = \frac{1}{n!}(s-\xi)^{n+1}\int_0^1 (1-y)^n \psi^{(n+1)}(\xi + y(s-\xi))dy. \tag{4.63}$$

Therefore $R_\xi \in C^\infty(I_\delta)$ since ψ is smooth on $[a, b]$.

Now, substituting (4.54) into (4.61) we obtain, for $s \in I_\delta$, $\psi(s) = \psi(\xi) + R_\xi(s)$. Hence, using (4.63), we deduce,

$$|\psi(s) - \psi(\xi)| \le |s - \xi|^{n+1}|G(s)|, \tag{4.64}$$

where for $s \in I_\delta$,

$$\begin{aligned} |G(s)| &\le \frac{1}{n!}\int_0^1 (1-y)^n dy \max_{t\in[s,\xi]}\left|\psi^{(n+1)}(t)\right| \\ &\le \frac{1}{(n+1)!}\max_{t\in[s,\xi]}\left|\psi^{(n+1)}(t)\right|. \end{aligned}$$

Next, similarly to (4.61), we expand the derivative of ψ around ξ,

$$\psi'(s) = \psi'(\xi) + (s-\xi)\psi''(\xi) + \ldots + \frac{(s-\xi)^{(n-1)}}{(n-1)!}\psi^{(n)}(\xi) + \tilde{R}_\xi(s), \tag{4.65}$$

where

$$\tilde{R}_\xi(s) = \frac{1}{(n-1)!}\int_\xi^s (s-t)^{n-1}\psi^{(n+1)}(t)dt.$$

Hence substituting (4.54) into (4.65), we obtain $\psi'(s) := \tilde{R}(\xi)$. Changing the variable from t to y as before, we obtain,

$$\psi'(s) = \frac{1}{(n-1)!}(s-\xi)^n\int_0^1 (1-y)^{n-1}\psi^{(n+1)}(\xi + y(s-\xi))dy.$$

Thus, we find,

$$|\psi'(s)| \ge |s - \xi|^n|\tilde{G}(s)|, \tag{4.66}$$

where $\tilde{G} \in C^{\infty}(I_\delta)$ satisfies for $s \in I_\sigma$ and σ sufficiently small,

$$\begin{aligned} |\tilde{G}(s)| &\geq \frac{1}{(n-1)!}\int_0^1 (1-y)^{n-1} dy \min_{t\in[s,\xi]} \left|\psi^{(n+1)}(t)\right| \\ &\geq \frac{1}{n!}\min_{t\in[s,\xi]} \left|\psi^{(n+1)}(t)\right| \geq C > 0, \end{aligned}$$

for some constant C, due to (4.55). Using (4.64) and (4.66), we obtain,

$$\begin{aligned} \frac{|\psi(s)-\psi(\xi)|}{|\psi'(s)|^{\frac{n+1}{n}}} &\leq \frac{|s-\xi|^{n+1}\,|G(s)|}{|s-\xi|^{n+1}\left|\tilde{G}(s)\right|^{\frac{n+1}{n}}} \\ &\leq C. \end{aligned} \tag{4.67}$$

Now, since

$$\left(\psi^{-1}\right)'(\tau) = \frac{1}{\psi'(\psi^{-1}(\tau))},$$

we can write (4.67) with $s = \psi^{-1}(\tau)$, as follows,

$$|\tau - \psi(\xi)| \left|\left(\psi^{-1}\right)'(\tau)\right|^{\frac{n+1}{n}} \leq C.$$

Finally, rearranging the latter equation, we obtain,

$$\left|\left(\psi^{-1}\right)'(\tau)\right| \leq C\,|\tau-\psi(\xi)|^{-\frac{n}{n+1}}, \tag{4.68}$$

for $\tau = \psi(s)$, $s \in I_\delta$. □

Example

Consider a function

$$\psi(s) = s^2, \quad \text{near } s = 0, \ s \geq 0.$$

The function $\psi(s)$ has a stationary pointof order $n = 1$ at $s = 0$. We also have,

$$\psi^{-1}(\tau) = \tau^{-1/2},$$

and $\psi\left(\psi^{-1}(\tau)\right) = \tau$. The first and second derivatives of ψ are

$$\begin{aligned} \psi'(s) &= 2s, \\ \psi''(s) &= 2. \end{aligned}$$

Then using the chain rule, we deduce,

$$\begin{aligned} \psi\left(\psi^{-1}(\tau)\right) &= \tau, \\ \Rightarrow \qquad \psi'\left(\psi^{-1}(\tau)\right)\ \left(\psi^{-1}\right)'(\tau) &= 1. \end{aligned}$$

Therefore,

$$\begin{aligned} \left(\psi^{-1}\right)'(\tau) &= \frac{1}{\psi'\left(\psi^{-1}(\tau)\right)} \\ &= \frac{1}{2\psi^{-1}(\tau)} \\ &= \frac{1}{2}\tau^{-1/2}. \end{aligned} \tag{4.69}$$

Equation (4.69) is consistent with (4.60) with $j = 1$. Furthermore, by the second application of the chain rule,

$$\psi''\left(\psi^{-1}(\tau)\right)\ \left[\left(\psi^{-1}\right)'(\tau)\right]^2 + \psi'\left(\psi^{-1}(\tau)\right)\ \left(\psi^{-1}\right)''(\tau) = 0.$$

Then,

$$\begin{aligned} \left(\psi^{-1}\right)''(\tau) &= -\frac{2\left[\left(\psi^{-1}\right)'(\tau)\right]^2}{\psi'\left(\psi^{-1}(\tau)\right)} \\ &= -\frac{1}{2\tau}\frac{1}{2\tau^{1/2}} \\ &= -\frac{1}{4}\tau^{-3/2}. \end{aligned} \tag{4.70}$$

Equation (4.69) is consistent with (4.60) with $j = 2$. Applying the chain rule repeatedly and using the known derivatives of ψ, Lemma 4.11 can be verified for all j.

Proof of Theorem 4.10

Proof. We first observe

$$F_1(\tau) := -f(\psi^{-1}(\tau))\left(\psi^{-1}\right)'(\tau).$$

Faa di Bruno's formula [37] tells us that

$$\frac{\partial^p}{\partial\tau^p}\left\{f\left(\psi^{-1}(\tau)\right)\right\}, \qquad \text{for} \quad p \geq 1,$$

is a linear combination of terms of the form

$$\left[f^{(m_1+\ldots+m_p)}\left(\psi^{-1}(\tau)\right)\right]\left[\prod_{j=1}^{p}\left(\left(\psi^{-1}\right)^{(j)}(\tau)\right)^{m_j}\right], \tag{4.71}$$

where

$$m_1 + 2m_2 + \ldots + pm_p = p. \tag{4.72}$$

Now, the first term in (4.71) is bounded, while in view of estimates (4.60), the second term in (4.71) can be estimated by

$$C_p|\tau - \psi(\xi)|^{(\alpha-1)m_1+(\alpha-2)m_2+\ldots+(\alpha-p)m_p}, \tag{4.73}$$

where C_p depends on p and $\alpha = 1/(n+1)$. Then using (4.72), the index in (4.73) satisfies

$$\alpha \sum_{j=1}^{p} m_j - \sum_{j=1}^{p} jm_j = \alpha \sum_{j=1}^{p} m_j - p \geq \alpha - p.$$

So each term of the form (4.71) may be bounded by $C_p|\tau - \psi(\xi)|^{\alpha-p}$. Now let us consider

$$\frac{\partial^p}{\partial \tau^p} \left\{ f\left(\psi^{-1}(\tau)\right) \psi^{-1}(\tau) \right\}, \qquad \text{for} \quad p \geq 1.$$

By Leibnitz rule this is a linear combination of

$$\frac{\partial^l}{\partial \tau^l} \left\{ f\left(\psi^{-1}(\tau)\right) \right\} \frac{\partial^{p-l}}{\partial \tau^{p-l}} \left\{ \psi^{-1}(\tau) \right\}, \quad l = 0, \ldots, p. \tag{4.74}$$

From the above discussion and (4.60) again, for $l \neq 0$ and $l \neq p$, each term of the form (4.74) can be estimated by

$$C_p|\tau - \psi(\xi)|^{\alpha-l}|\tau - \psi(\xi)|^{\alpha-p+l} = C_p|\tau - \psi(\xi)|^{2\alpha-p}.$$

However, when $l = 0$ or p the bound is

$$C_p|\tau - \psi(\xi)|^{\alpha-p}.$$

Hence the result follows. The proof of estimates (4.59) for F_2 follows analogously. □

Theorem 4.10 implies $F_1(\tau)$ and $F_2(\tau)$ have an algebraic singularity at $\tau = \psi(\xi)$. Filon-type quadratures applied on a mesh graded towards the singularity can be used to compute integrals (4.56) with F_1 and F_2 satisfying (4.58) and (4.59) accurately as we discuss in detail in Chapter 5.

4.2.2 The idea for the computation of the double integrals

Let us return to the double integral introduced in Theorem 4.3:

$$J_k := \int_a^b \int_{c(s)}^{d(s)} M(s,t)\exp(ik\Psi(s,t))dtds, \tag{4.75}$$

where M is slowly-varying, i.e. all derivatives bounded independently of k and the phase-function Ψ is smooth. We imagine that M may depend on k, although we do not reflect this in the notation.

Definition 4.12. *We define the function $\psi_{[s]}(t)$ as*

$$\psi_{[s]}(t) := \Psi(s,t).$$

Here the notation indicates that we are thinking of $\psi_{[s]}$ as a family of functions of t parameterised by s. Similarly, we define

$$\psi^{[t]}(s) := \Psi(s,t),$$

as a family of functions of s parameterised by t.

Clearly,

$$\psi_{[s]}(t) = \Psi(s,t) = \psi^{[t]}(s).$$

As in the one dimensional case, explained above, our immediate aim is to rewrite (4.75) so that the oscillator is canonical. We do this in Lemma 4.13 under conditions defined in the following hypothesis.

Hypothesis A. *For all $s \in [a,b]$,*

$$\psi'_{[s]}(t) \neq 0, \quad \textit{for all} \quad t \in [c(s), d(s)]. \tag{4.76}$$

We will see later in this thesis that this hypothesis is often (but not always) satisfied in the integrals with arise from the Galerkin discretisation.

Lemma 4.13. *Under Hypothesis A,*

$$J_k = \int_a^b \int_{\psi_{[s]}(c(s))}^{\psi_{[s]}(d(s))} H(s,t)\exp(ik\tau)d\tau ds, \tag{4.77}$$

where

$$H(s,\tau) := \frac{M(s,\psi_{[s]}^{-1}(\tau))}{\left|\psi'_{[s]}\left(\psi_{[s]}^{-1}(\tau)\right)\right|}. \tag{4.78}$$

Proof. For fixed s we make the change of variable:

$$\tau := \psi_{[s]}(t). \tag{4.79}$$

This yields,

$$dt = \frac{1}{|\psi'_{[s]}(\psi^{-1}_{[s]}(\tau))|} d\tau, \qquad \text{with} \qquad t = \psi^{-1}_{[s]}(\tau),$$

and the result follows. □

Depending on the shape of the functions $\psi_{[s]}(c(s))$ and $\psi_{[s]}(d(s))$, the integral J_k in (4.77) may be rewritten by *changing the order of integration* in the form,

$$\begin{aligned} J_k &:= \int_{\tau_0}^{\tau_1} F_1(\tau)\exp(ik\tau)d\tau + \ldots + \int_{\tau_{N-1}}^{\tau_N} F_N(\tau)\exp(ik\tau)d\tau \\ &= \sum_{j=0}^{N-1} I_k^{[\tau_j,\tau_{j+1}]}[F_{j+1}], \end{aligned} \tag{4.80}$$

where F_j, $j = 1, \ldots, N$, are themselves defined as integrals. We now give two examples to illustrate this process.

Example 1. Let us consider a double integral (4.75),

$$J_k := \int_a^b \int_c^d M(s,t)\exp(ik\Psi(s,t))dtds,$$

where the domain of integration D is a rectangular domain

$$D := \{(s,t), \quad s \in [a,b], \quad t \in [c,d], \quad [a,b] \cap [c,d] = \varnothing\}.$$

Under Hypothesis A, we have,

$$\begin{aligned} J_k &:= \int_a^b \int_c^d M(s,t)\exp(ik\Psi(s,t))dtds \\ &= \int_a^b \int_{\psi_{[s]}(c)}^{\psi_{[s]}(d)} H(s,\tau)\exp(ik\tau)d\tau ds, \end{aligned} \tag{4.81}$$

where H is defined in (4.78).

In order to write J_k in the form (4.80) we need to change the order of integration. Consider the case when $\psi^{[c]}$ and $\psi^{[d]}$ are monotone functions on $[a,b]$.

Remark 4.14. *The latter condition is often satisfied by the integrals (4.11) arising from the scattering problems as we will show in Theorem 4.17 later.*

The monotonicity condition can be written in the following form, for all $s \in [a, b]$,

$$\left(\psi^{[c]}\right)'(s) \neq 0, \qquad \text{and} \qquad \left(\psi^{[d]}\right)'(s) \neq 0. \tag{4.82}$$

Then, the rectangular domain of integration in $s - t$ space in (4.75) transforms into a curvilinear rectangular domain in $s - \tau$ space with straight edges and curved edges as shown in Figure 4.1. The transformation is governed by (4.79).

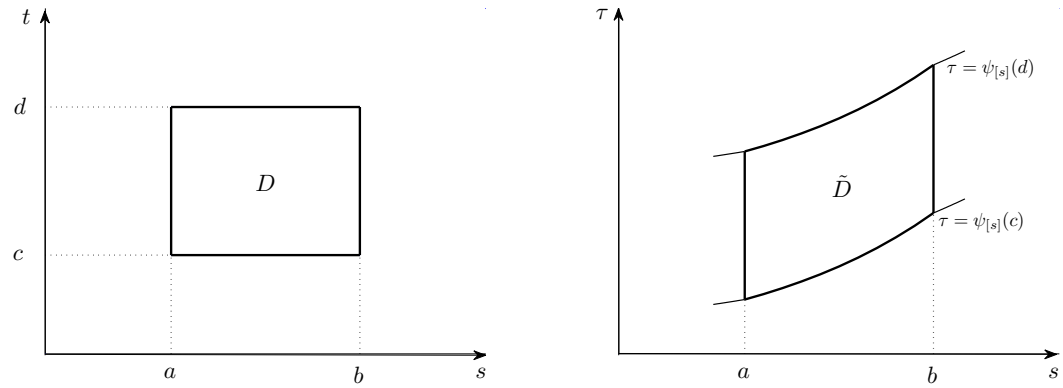


Figure 4.1: *Original (left) and transformed (right) domains of integration under transformation (4.79). Note the upper and lower curves bounding $\tilde{D}$ are monotone due to assumption (4.82).*

To obtain the form (4.80), we need to subdivide $\tilde{D}$ into subdomains so that the integral over each subdomain can be written as a repeated integral. In this case, we subdivide $\tilde{D}$ into three subdomains as shown in Figure 4.2.

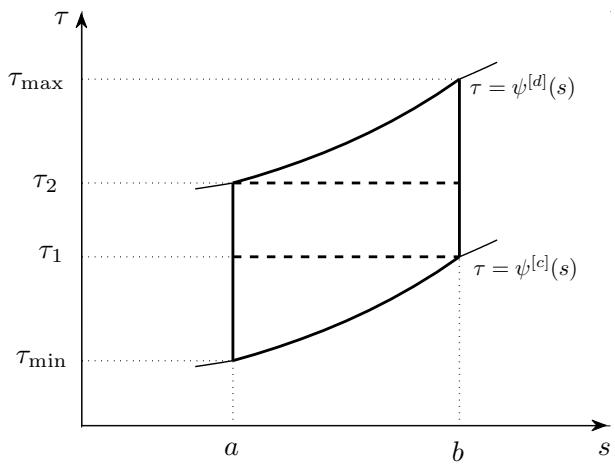


Figure 4.2: *Transformed domain $\tilde{D}$ is subdivided into three subdomains.*

Then, (4.75) can be written in the form (4.80),

$$\begin{aligned} J_k &= \int_{\tau_{\min}}^{\tau_1} F_1(\tau)\exp(ik\tau)d\tau + \int_{\tau_1}^{\tau_2} F_2(\tau)\exp(ik\tau)d\tau \\ &+ \int_{\tau_2}^{\tau_{\max}} F_3(\tau)\exp(ik\tau)d\tau, \end{aligned} \tag{4.83}$$

where $\tau_{\min} = \psi^{[c]}(a)$, $\tau_1 = \psi^{[c]}(b)$, $\tau_2 = \psi^{[d]}(a)$, $\tau_{\max} = \psi^{[d]}(b)$. The functions F_1, F_2 and F_3 are defined as

$$F_1(\tau) := \int_a^{\left(\psi^{[c]}\right)^{-1}(\tau)} H(s,\tau)ds, \tag{A1}$$

$$F_2(\tau) := \int_a^b H(s,\tau)ds, \tag{A2}$$

$$F_3(\tau) := \int_{\left(\psi^{[d]}\right)^{-1}(\tau)}^b H(s,\tau)ds. \tag{A3}$$

The integrals in (4.83) are in the form amenable to classical Filon-type quadratures and the integrals (A1) - (A3) have slowly-varying integrands.

Similarly to the one-dimensional case, we can make some observations about the behaviour of F_1, F_2 and F_3, provided particular properties of M and Ψ are known.

Since M is slowly-varying and $\psi^{[c]}$ and $\psi^{[d]}$ are assumed monotone, we can show by differentiating (A1), (A2), and (A3) that F_j, $j = 1, \ldots, 3$, have all derivatives bounded independently of k. In other words, each F_j, $j = 1, 2, 3$, is slowly-varying in $[\tau_{\min}, \tau_1]$, $[\tau_1, \tau_2]$ and $[\tau_2, \tau_{\max}]$ respectively.

In cases when either the upper boundary $\tau = \psi^{[d]}(s)$ or the lower boundary $\tau = \psi^{[c]}(s)$ are not monotone, (i.e. (4.82) does not hold), the functions F_j, $j = 1, 2, 3$, will have singularities at these points.

In the next example, we consider the case when the original domain is triangular and the condition (4.82) is violated.

Example 2. Consider an integral,

$$J_k := \int_b^c \int_s^c M(s,t)\exp(ik\Psi(s,t))dtds, \tag{4.85}$$

where the domain of integration T is an upper triangular domain:

$$T := \{(s,t)), \quad s \in [b,c], \quad t \in [s,c]\},$$

and the function M is slowly-varying and the phase-function Ψ is smooth and satisfies the

following conditions,

$$\psi_{[s]}(s) = \Psi(s,s) = \psi^{[s]}(s) = 0, \tag{4.86}$$

also there exist a unique point $\xi \in [b,c]$ such that

$$\left(\psi^{[c]}\right)'(\xi) = 0, \qquad \text{and} \qquad \left(\psi^{[c]}\right)''(\xi) \neq 0. \tag{4.87}$$

The condition (4.87) is often satisfied by the integrals (4.11) arising from the scattering problems as we will see in the next section.

The change of variable (4.79) and assumption (4.86) yields

$$J_k = \int_b^c \int_0^{\psi_{[s]}(c)} H(s,\tau)\exp(ik\tau)d\tau dx. \tag{4.88}$$

The transformed domain $\tilde{T}$ is illustrated in Figure 4.3.

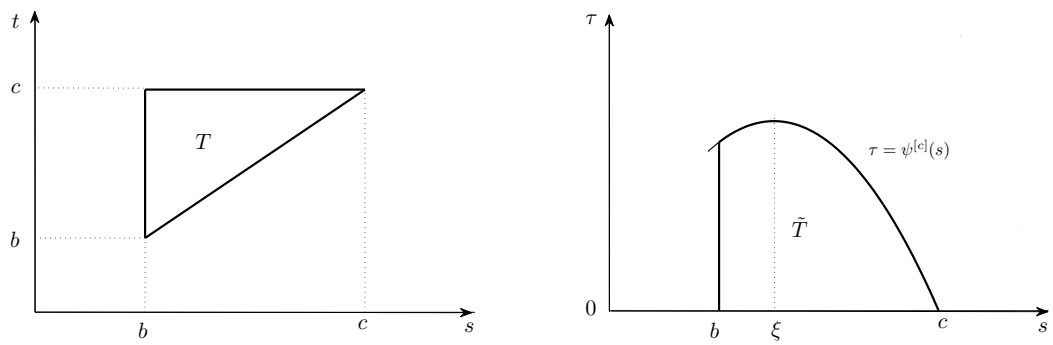


Figure 4.3: *Original and transformed domains of integration. Note the upper curve bounding $\tilde{D}$ contains a turning point and the lower boundary is $\tau = 0$.*

We want to change the order of integration in (4.88) to rewrite the integral in analogous way to (4.83). In order to do this, we need to subdivide $\tilde{T}$ into two subdomains as shown in Figure 4.4. Then, we write J_k in the form (4.80),

$$J_k = \int_{\tau_{\min}}^{\tau_1} F_1(\tau)\exp(ik\tau)d\tau + \int_{\tau_1}^{\tau_{\max}} F_2(\tau)\exp(ik\tau)d\tau, \tag{4.89}$$

where $\tau_{\min} = 0$, $\tau_1 = \psi^{[c]}(b)$ and $\tau_{\max} = \psi^{[c]}(\xi)$, and the functions F_1 and F_2 are defined as

$$F_1(\tau) := \int_b^{\left(\psi^{[c]}\right)^{-1}(\tau)} H(s,\tau)\,ds, \qquad \text{and} \qquad F_2(\tau) := \int_{r_1(\tau)}^{r_2(\tau)} H(s,\tau)\,ds, \tag{4.90}$$

where

$$r_1 : [\tau_1, \tau_{\max}] \to [b,\xi], \qquad r_1(\tau) := \left(\psi^{[c]}\right)^{-1}(\tau),$$

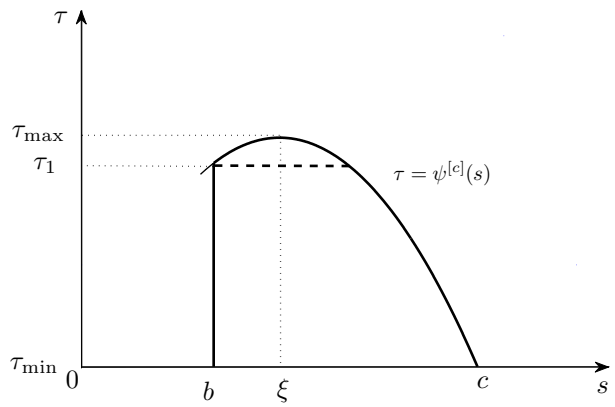


Figure 4.4: *Transformed domain $\tilde{T}$ is subdivided into two subdomains.*

is monotone increasing and

$$r_2 : [\tau_1, \tau_{\max}] \to [\xi, c], \qquad r_2(\tau) := \left(\psi^{[c]}\right)^{-1}(\tau),$$

is monotone decreasing, and $H(s, \tau)$ is defined in (4.78). After the change of variable, the integrals in (4.89) are in the form amenable to classical Filon-type quadratures.

Due to (4.87), functions r_1 and r_2 have a square-root singularity. This can be shown similarly to the proof of Theorem 4.10. Although we have assumed that M is smooth, similar arguments can be applied if M has a weak singularity. In the stiffness matrix $\boldsymbol{R}$ in (2.43), the entries are of the form (4.11) with $M(s, t)$ that has a log-singularity at $t = s$. This induces a singularity in F_1. We prove that F_1 has a log-singularity at $\tau = 0$ in Section E for a particular case. We also show in Chapter 5 that Filon quadrature applied on graded meshes can be used for approximating (4.89) when F_1 and F_2 have algebraic singularities.

We will describe typical geometries of transformed domains arising in scattering problems in Section 4.3. As we will see, in all our applications, there is never more than one turning point on the upper and lower boundaries of the transformed domains. This follows from the assumption that Γ, the boundary of the scatterer, is convex.

4.3 Application to the integrals arising from scattering problem

The double integrals (4.11) arising from the Galerkin discretisation are of the form

$$J_k := \int_{\Lambda_l} \int_{\Lambda_j} M(s,t) \exp\left(ik\Psi(s,t)\right) dtds, \tag{4.91}$$

with Ψ given in (4.10) and $M(s,t)$ is typically (weakly) singular at $s = t$ and Λ_l and Λ_j are subintervals of $[0, 2\pi]$, see below. By Theorem 4.3, M satisfies the estimates (4.15). The phase-function Ψ given in (4.12) satisfies (4.86). Recall the intervals Λ_1, Λ_2 and Λ_3 in $[0, 2\pi]$ are defined in (2.40) as follows,

$$\Lambda_1 = [a, b], \quad \Lambda_2 = [b, c], \quad \Lambda_3 = [c, d], \tag{4.92}$$

with

$$a = t_1 - \delta, \quad b = t_1 + \varepsilon, \quad c = t_2 + \varepsilon, \quad d = t_2 + \delta,$$

with $\varepsilon > \delta$, where t_1 and t_2 are transition points on the boundary (also called shadow boundaries, where $\boldsymbol{n}(t_j) \cdot \boldsymbol{a} = 0$, $j = 1, 2$).

Remark 4.15. *The condition $\varepsilon > \delta$ ensures that transition parts of the boundary extend further into illuminated part than the shadow. This ensures that Hypothesis A is satisfied for all $(s,t) \in ([a,d] \times [a,d])$ apart from only two domains, see Theorem 4.16. We discuss the computation of integrals over the two domains where Hypothesis A is not satisfied, later in Section 4.4.1.*

In Figure 4.5 we illustrate all the possible rectangular domains of integration which arise in the computation of (4.11). In domains which include the diagonal $s = t$ (at which M is singular), we decompose the integral (4.91) as a sum of two integrals: one over the upper triangular domain and one over the lower triangular domain.

In this section, we describe the geometries of the corresponding transformed domains governed by the transformation (4.79). The following two theorems examine the validity of Hypothesis A as well as conditions (4.82) and (4.87).

Theorem 4.16. *For $s, t \in ([a,d] \times [a,d])$, excluding $(s,t) \in \Lambda_1 \times \Lambda_1$ and $(s,t) \in \Lambda_3 \times \Lambda_3$, the following holds*

$$\left(\psi_{[s]}\right)'(t) \neq 0.$$

The proof is derived in Lemmas C.2 and C.5 of Section C of the Appendix. Theorem 4.16 implies that in all domains in Figure 4.5 apart from two rectangular domains, the conditions of Hypothesis A are satisfied.

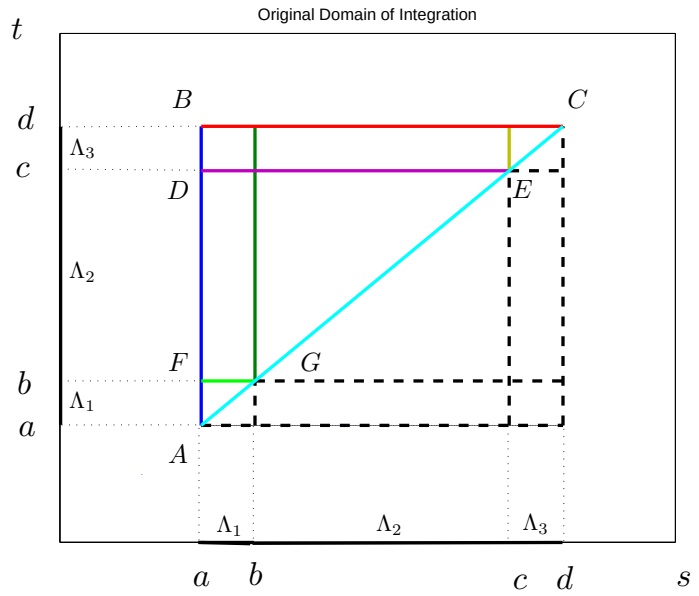


Figure 4.5: *The original rectangular domains of integration which include the diagonal $s = t$ (where M is singular). The rectangular domains containing the diagonal are separated into the upper triangular and lower triangular subdomains.*

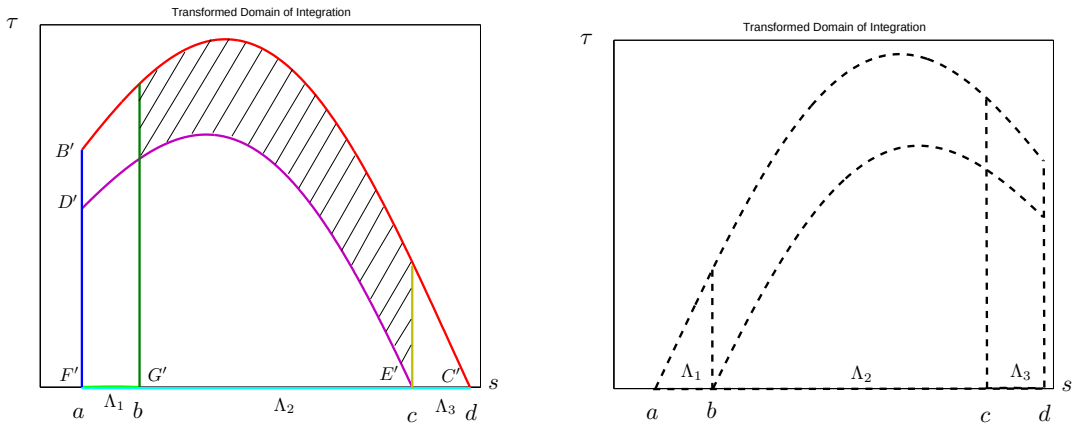


Figure 4.6: *On the left, the transformed domains of integration, corresponding to the above-the-diagonal part of the original domains are illustrated; on the right, the transformed domains corresponding to the below-the-diagonal part of the original domains are illustrated. The transformation is governed by (4.79). We consider the integration over the triangular domain AFG in more detail later, see Figure 4.16.*

Theorem 4.17. *For each $t \in [a, d]$, there exists a unique point $\xi \in [a, t)$ such that*

$$\left(\psi^{[t]}\right)'(\xi) = 0, \qquad \textit{and} \qquad \left(\psi^{[t]}\right)''(\xi) \neq 0.$$

The proof of this theorem is derived similarly to the proof of Theorem 4.16 in the Appendix. Theorem 4.17 implies that in some transformed under (4.79) domains, the upper, the lower

or both curvilinear boundaries would contain a turning point.

In Figure 4.6, the geometries of the domains of Figure 4.5 transformed under (4.79) are depicted. In Figure 4.6, the plot on the left corresponds to the above-the-diagonal domains of the original domain in Figure 4.5. The plots are shown for the case of an elliptic scatterer although the geometries of the transformed domains are illustrative of the corresponding domains for more general convex obstacles.

A number of transformed domains have boundaries that are curves with turning points. As we have seen from Section 4.2.1, the resulting transformed integrals over these domains will contain singularities.

In Table 4.1, we present the geometries of the individual domains of Figure 4.5 transformed under (4.79).

The double integrals can be written as a sum of integrals (4.80), with singularities confined to the end points,

$$J_k := \sum_{j=0}^{J-1} I_k^{[\tau_j,\tau_{j+1}]}[F_{j+1}], \tag{4.93}$$

with $\tau_0 = \tau_{\min}$ and $\tau_N = \tau_{\max}$. See for example (4.83). We introduce the following notation for the integrand function F_j, $j = 1, \ldots, J$

$$F_j(\tau) = \int_{lower\ boundary}^{upper\ boundary} H\left(s, \tau\right) ds, \tag{4.94}$$

where the *upper* and the *lower boundaries* are presented in Table 4.2 for the transformed domains corresponding to the above-the-diagonal regions of the original domains. See for example F_1, F_2, F_3 defined in (A1), (A2) and (A3) following (4.83).

For the below-the-diagonal subdomains, the table can be constructed in a similar way.

In the Table 4.2, the notation $(\Lambda_2 \times \Lambda_2)^+$ is used to denote the domain $\{(s,t) : s,t \in [b,c], t > s\}$ while $(\Lambda_3 \times \Lambda_3)^+$ is used to denote $\{(s,t) : s,t \in [c,d], t > s\}$. In the second column of Table 4.2, we display the singularities in the interval $[\tau_{\min}, \tau_{\max}]$. We emphasize what type of singularity F_j has at τ by adding **(L)** or **(S)** next to τ, where **(L)** represents a logarithmic singularity and **(S)** represents a square-root singularity.

In Section E of the Appendix, we prove that

- if the original domain of integration contains the diagonal, i.e. points $s = t$, then in the transformed domain the function F_1 has a log-singularity at $\tau = 0$
- if upper or lower boundaries of the transformed domain contain turning points, then at least one of F_j, $j = 2, \ldots, J$, has a square-root singularity at $\tau = \psi(\xi)$, where ξ is a turning point.

- away from these singularities, the functions F_j, $j = 1, \ldots, J$, are smooth.

Under the change of variables (4.79), the horizontal lines $t = b$, $t = c$ and $t = d$ become curves $\tau = \psi^{[b]}(s)$, $\tau = \psi^{[c]}(s)$ and $\tau = \psi^{[d]}(s)$ respectively. These curves represent the upper and lower boundaries of the transformed domains, see e.g. Figure 4.6 and Table 4.1. In Table 4.2, we denote ξ_b,ξ_c and ξ_d as the turning points of the curves $\tau = \psi^{[b]}(s)$, $\tau = \psi^{[c]}(s)$ and $\tau = \psi^{[d]}(s)$ respectively:

$$\left(\psi^{[b]}\right)'(\xi_b) = 0, \qquad \left(\psi^{[c]}\right)'(\xi_c) = 0, \qquad \left(\psi^{[d]}\right)'(\xi_d) = 0. \tag{4.95}$$

Let us now consider in more detail two examples of integration over the triangular domain $(\Lambda_2 \times \Lambda_2)^+$ and the rectangular domain $(\Lambda_2 \times \Lambda_3)$.

Example of computing the double integral over the domain $(\Lambda_2 \times \Lambda_2)^+$

As we have shown in (4.89) the integral J_k over the triangular domain $(\Lambda_2 \times \Lambda_2)^+$ can be written as

$$\begin{aligned} J_k &:= \int_b^c \int_s^c M(s,t) \exp\left(ik\Psi(s,t)\right) dtds \\ &= \int_0^{\tau_1} F_1(\tau) \exp(ik\tau) d\tau + \int_{\tau_1}^{\tau_{\max}} F_2(\tau) \exp(ik\tau) d\tau, \end{aligned} \tag{4.96}$$

where $F_1 : [0, \tau_1] \to \mathbb{R}$ and $F_2 : [\tau_1, \tau_{\max}] \to \mathbb{R}$ are defined as

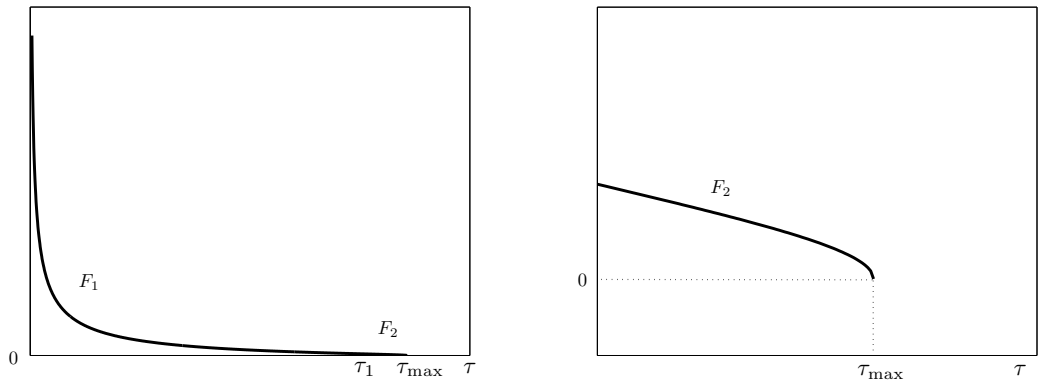


Figure 4.7: *Plot of function $F_1(\tau)$ and $F_2(\tau)$. The function $F_1(\tau)$ has a logarithmic singularity at $\tau = 0$ and the function $F_2(\tau)$ has the square-root singularity at $\tau = \tau_{\max}$.*

$$F_1(\tau) = \int_b^{\left(\psi^{[c]}\right)^{-1}(\tau)} H(s,\tau)\, ds, \qquad F_2(\tau) = \int_{r_1(\tau)}^{r_2(\tau)} H(s,\tau)\, ds,$$

where,

$$r_1(\tau) = \left(\psi^{[c]}\right)^{-1}(\tau) \in [b, \xi_c], \quad \text{and} \quad r_2(\tau) = \left(\psi^{[c]}\right)^{-1}(\tau) \in [\xi_c, c],$$

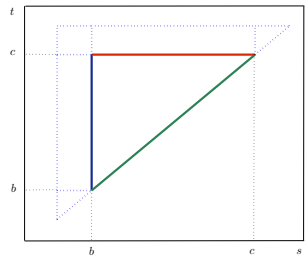
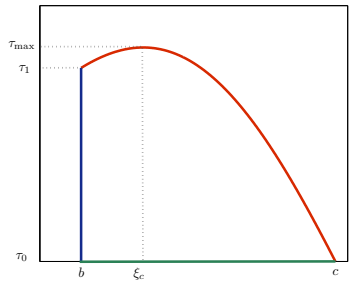
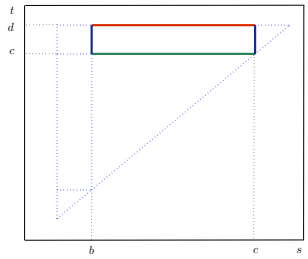
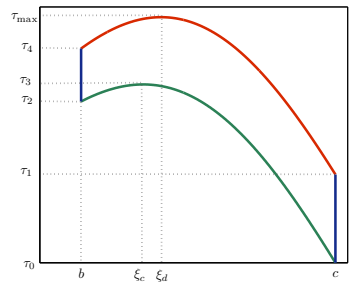
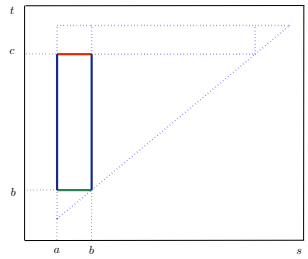
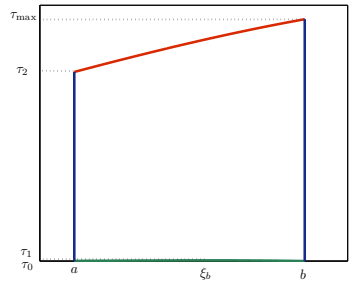
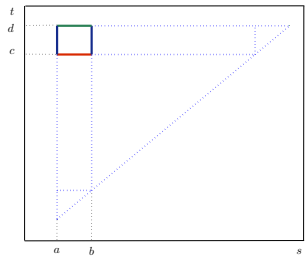
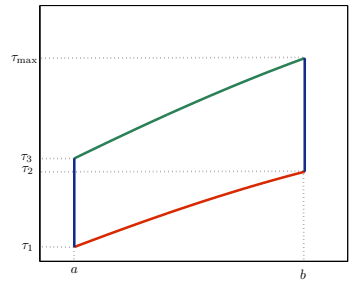
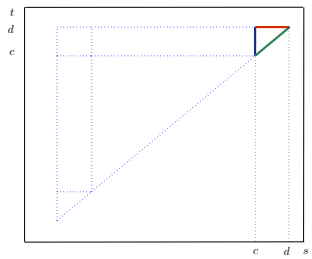
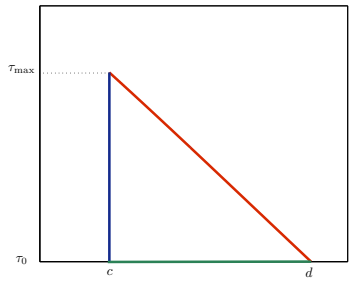
	original domain	transformed domain
$(\Lambda_2 \times \Lambda_2)^+$		
$(\Lambda_2 \times \Lambda_3)$		
$(\Lambda_1 \times \Lambda_2)$		
$(\Lambda_1 \times \Lambda_3)$		
$(\Lambda_3 \times \Lambda_3)^+$		

Table 4.1: *Geometries of the original and transformed under (4.79) domains respectively.*

original domain	singularities	τ-interval	lower boundary	upper boundary
$(\Lambda_2 \times \Lambda_2)^+$	(L) $\tau_0 = 0$, $\tau_1 = \Psi(b,c)$	$[\tau_0, \tau_1]$	b	$\left(\psi^{[c]}\right)^{-1}(\tau)$
	(S) $\tau_{\max} = \Psi(\xi_c, c)$	$[\tau_1, \tau_{\max}]$	$\left(\psi^{[c]}\right)^{-1}(\tau)$	$\left(\psi^{[c]}\right)^{-1}(\tau)$
$(\Lambda_2 \times \Lambda_3)$	(L) $\tau_0 = 0$, $\tau_1 = \Psi(c,d)$	$[\tau_0, \tau_1]$	$\left(\psi^{[c]}\right)^{-1}(\tau)$	c
	$\tau_2 = \Psi(b,c)$	$[\tau_1, \tau_2]$	$\left(\psi^{[c]}\right)^{-1}(\tau)$	$\left(\psi^{[d]}\right)^{-1}(\tau)$
	(S) $\tau_3 = \Psi(\xi_c, c)$	$[\tau_2, \tau_3]$	b	$\left(\psi^{[c]}\right)^{-1}(\tau)$
			$\left(\psi^{[c]}\right)^{-1}(\tau)$	$\left(\psi^{[d]}\right)^{-1}(\tau)$
	$\tau_4 = \Psi(b,d)$	$[\tau_3, \tau_4]$	b	$\left(\psi^{[d]}\right)^{-1}(\tau)$
	(S) $\tau_{\max} = \Psi(\xi_d, d)$	$[\tau_4, \tau_{\max}]$	$\left(\psi^{[d]}\right)^{-1}(\tau)$	$\left(\psi^{[d]}\right)^{-1}(\tau)$
$(\Lambda_1 \times \Lambda_2)$	(L) $\tau_0 = 0$, $\tau_1 = \Psi(a,b)$	$[\tau_0, \tau_1]$	$\left(\psi^{[b]}\right)^{-1}(\tau)$	b
	(S) $\tau_2 = \Psi(\xi_b, b)$	$[\tau_1, \tau_2]$	a	$\left(\psi^{[b]}\right)^{-1}(\tau)$
			$\left(\psi^{[b]}\right)^{-1}(\tau)$	b
	$\tau_3 = \Psi(a,c)$	$[\tau_2, \tau_3]$	a	b
	$\tau_{\max} = \Psi(b,c)$	$[\tau_3, \tau_{\max}]$	$\left(\psi^{[c]}\right)^{-1}(\tau)$	b
$(\Lambda_1 \times \Lambda_3)$	$\tau_1 = \Psi(a,c)$, $\tau_2 = \Psi(b,c)$	$[\tau_1, \tau_2]$	a	$\left(\psi^{[c]}\right)^{-1}(\tau)$
	$\tau_3 = \Psi(a,d)$	$[\tau_2, \tau_3]$	a	b
	$\tau_{\max} = \Psi(b,d)$	$[\tau_3, \tau_{\max}]$	$\left(\psi^{[d]}\right)^{-1}(\tau)$	b
$(\Lambda_3 \times \Lambda_3)^+$	(L) $\tau_0 = 0$, $\tau_{\max} = \Psi(c,d)$	$[\tau_0, \tau_{\max}]$	c	$\left(\psi^{[d]}\right)^{-1}(\tau)$

Table 4.2: *Description of the transformed domains. In the first column we display the original domain of integration in J_k that is either a rectangular or a triangular domain. See Table 4.1 for plots of these domains. The remaining four columns of the table describe the corresponding transformed domain and contain a sufficient information to enable us to write the double integral J_k in the form (4.93). The intervals $[\tau_j, \tau_{j+1}]$, $j = 0, \ldots, J-1$, in (4.93) are displayed in the third column of the table with τ_j defined in the second column. The turning points ξ_b, ξ_c and ξ_d are defined in (4.95). The lower and the upper boundaries of functions F_j defined in (4.94) are displayed in the fourth and fifth columns respectively. For each $j = 1, \ldots, J$, we emphasize what type of singularity F_j has at τ by adding **(L)** or **(S)** next to τ in the second column, where **(L)** represents a logarithmic singularity and **(S)** represents a square-root singularity. The function F_j is smooth away from these singularities. The entries of this table were obtained by investigating the geometries of the transformed domains and can be easily verified by looking at the geometry of the transformed domains in Table 4.1.*

with ξ_c defined in (4.95).

The geometry of the transformed domain of integration can be observed in Figure 4.3 (see also Table 4.1).

In Figure 4.7 the plots of functions F_1 and F_2 are illustrated. The function F_1 has a logarithmic singularity at $\tau = 0$ and the function F_2 has a square-root singularity at $\tau = \tau_{\max}$. Away from these points, the funtions F_1 and F_2 are smooth. The proofs of these statements can be found in Lemma E.1, Lemma E.3 and Lemma E.5 in Section E of the Appendix.

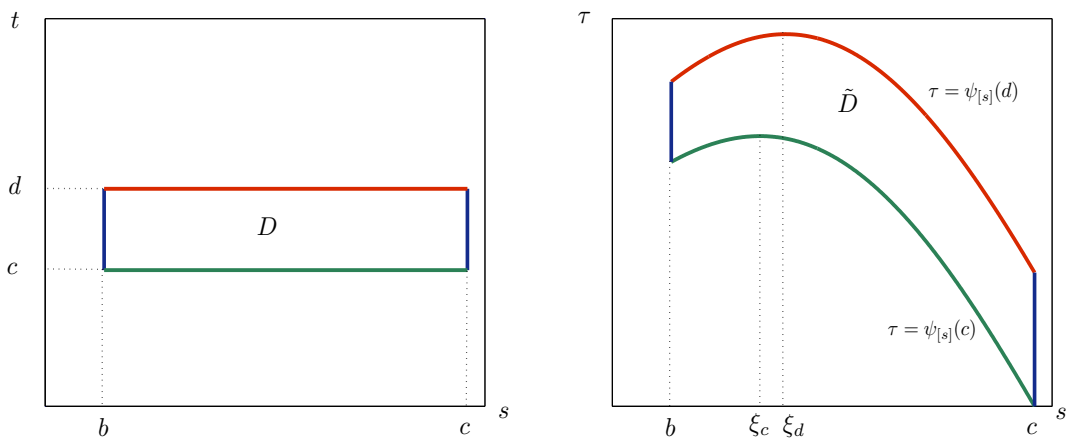


Figure 4.8: *The original domain of integration,* $(\Lambda_2 \times \Lambda_3)$ *and the transformed domain of integration.*

Example of computing the double integral over the domain $(\Lambda_2 \times \Lambda_3)$

Let us now consider another example, where the domain of integration is the rectangle $\Lambda_2 \times \Lambda_3$,

$$J_k = \int_{\Lambda_2} \int_{\Lambda_3} M(s,t) \exp\left(ik\Psi(s,t)\right) dtds,$$

The original and transformed domains of integration are illustrated in Figure 4.8 (see also Table 4.1).

In Figure 4.9, the transformed domain is divided into subdomains over which the integral J_k can be written in the form (4.80). Note that in the interval $\tau \in [\tau_2, \tau_3]$, the domain is further divided into two subdomains: one on the left of $s = \xi_c$ and another on the right. We then write the double integral J_k as a sum of five highly-oscillatory one-dimensional integrals,

$$J_k := \sum_{j=0}^{5} \int_{\tau_j}^{\tau_{j+1}} F_{j+1}(\tau) \exp(ik\tau) d\tau,$$

with $\tau_0 = \tau_{\min} = 0$, $\tau_1 = \Psi(c,d)$, $\tau_2 = \Psi(b,c)$, $\tau_3 = \Psi(\xi_c, c)$, $\tau_4 = \Psi(b,d)$ and $\tau_{\max} =$

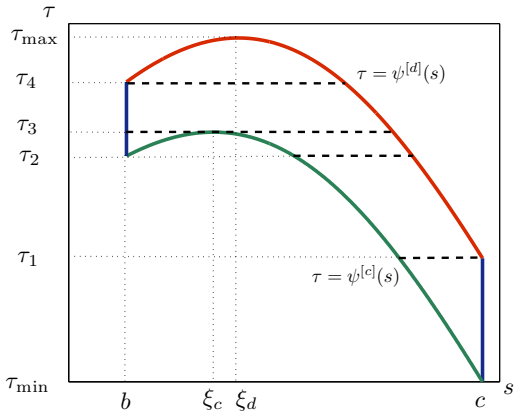


Figure 4.9: *The transformed domain is subdivided by the dashed lines into five subdomains.*

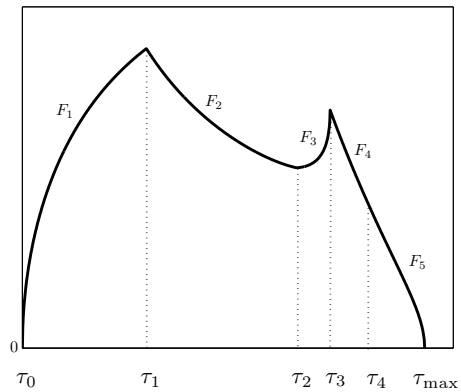


Figure 4.10: *Plot of functions F_j, $j = 1, \ldots, 5$. The function has the square-root singularity at $\tau = \tau_3$ and $\tau = \tau_{\max}$, and a* log*-singularity at $\tau = \tau_{\min} = 0$.*

$\Psi(\xi_d, d)$. The functions F_j are defined as follows,

$$F_j(\tau) = \int_{r_{1,j}(\tau)}^{r_{2,j}(\tau)} H\left(s, \psi_{[s]}^{-1}(\tau)\right) ds, \tag{4.97}$$

where $F_3(\tau)$ is a sum of two integrals:

$$F_3(\tau) = \int_{r_{1,3}^{I}}^{r_{2,3}^{I}} H\left(s, \psi_{[s]}^{-1}(\tau)\right) ds + \int_{r_{1,3}^{II}}^{r_{2,3}^{II}} H\left(s, \psi_{[s]}^{-1}(\tau)\right) ds,$$

where the functions $r_{1,j}$ and $r_{2,j}$ are defined, for subintervals of τ, as follows:

$$\begin{aligned}
&\tau = [0, \tau_1], && r_{1,1}(\tau) = \left(\psi^{[c]}\right)^{-1}(\tau), && r_{2,1}(\tau) = c,\\
&\tau = [\tau_1, \tau_2], && r_{1,2}(\tau) = \left(\psi^{[c]}\right)^{-1}(\tau), && r_{2,2}(\tau) = \left(\psi^{[d]}\right)^{-1}(\tau),\\
&\tau = [\tau_2, \tau_3], && r_{1,3}^{I}(\tau) = b, && r_{2,3}^{I}(\tau) = \left(\psi^{[c]}\right)^{-1}(\tau) \in [b, \xi_c],\\
& && r_{1,3}^{II}(\tau) = \left(\psi^{[c]}\right)^{-1}(\tau) \in [\xi_c, c], && r_{2,3}^{II}(\tau) = \left(\psi^{[d]}\right)^{-1}(\tau),\\
&\tau = [\tau_3, \tau_4], && r_{1,4}(\tau) = b, && r_{2,4}(\tau) = \left(\psi^{[d]}\right)^{-1}(\tau),\\
&\tau = [\tau_4, \tau_{\max}], && r_{1,5}(\tau) = \left(\psi^{[d]}\right)^{-1}(\tau) \in [b, \xi_4], && r_{2,5}(\tau) = \left(\psi^{[d]}\right)^{-1}(\tau) \in [\xi_4, c].
\end{aligned}$$

In Figure 4.10 we present the plots of F_j, $j = 1, \ldots, 5$. The functions F_1, F_3 and F_5 have singularities at $\tau_0 = \tau_{\min} = 0$, τ_3 and $\tau_{\max}$ respectively. At the point $\tau_0 = \tau_{\min} = 0$, the function $F_1(\tau)$ has a log-singularity inherited from the singularity in M at $s = t$. At the points τ_3 and τ_5, functions F_3 and F_5 respectively have a square-root singularity.

4.4 Stationary points for the phase function

In this section, we investigate the typical locations of points $(s,t) \in [0,2\pi] \times [0,2\pi]$ where

$$\psi'_{[s]}(t) = 0, \tag{4.98}$$

i.e. where the condition of Hypothesis A is not satisfied.

In Figure 4.11, the points t where Hypothesis A is not satisfied are plotted against s for the case of an elliptical scatterer in blue. The diagonal is also plotted in Figure 4.11 in red.

In Figure 4.12, the rectangles plotted in black represent the domains of integration $(\Lambda_l \times \Lambda_j)$, with $l, j = 1, 2, 3$ in the double integrals (4.11). Along with the black rectangles, the diagonal $s = t$ is also plotted in red. Also in Figure 4.12, the points where Hypothesis A is not satisfied in the domains contained within the black rectangles, are plotted in blue.

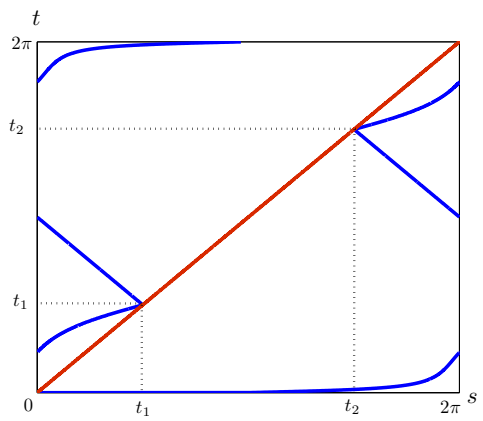


Figure 4.11: *The curves plotted in blue represent points $(s,t) \in ([0,2\pi] \times [0,2\pi])$ where Hypothesis A is not satisfied for the case of an elliptical scatterer $(a = 3, b = 1)$. The points where the integrand is singular are located on the diagonal $s = t$ which is plotted in red.*

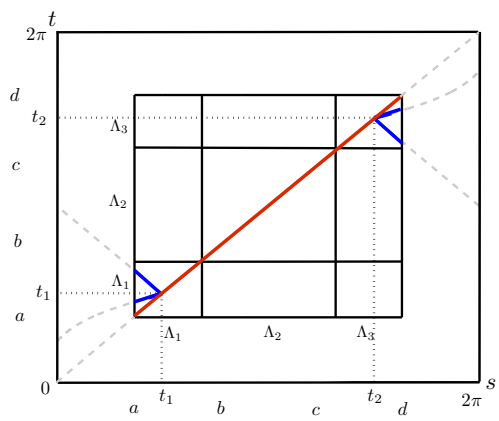


Figure 4.12: *The boundaries of the illuminated $\Lambda_2 := [b,c]$ and transition domains $\Lambda_1 := [a,b]$ and $\Lambda_3 := [c,d]$ are plotted along with the diagonal $s = t$ in red and blue curves representing points where Hypothesis A is not satisfied within the domain contained in the rectangles.*

From Figure 4.12 we observe that (4.98) holds only in two domains,

$$(\Lambda_1 \times \Lambda_1) \quad \text{and} \quad (\Lambda_3 \times \Lambda_3)\,. \tag{4.99}$$

Computing the double integrals (4.11) over the domains in (4.99) using integration method described in the previous section is not possible due to Hypothesis A not being satisfied. However the integrals can still be computed efficiently. The first result towards showing this is given in Theorem 4.18 below.

Theorem 4.18. *We define domains A_1 and A_2 as follows*

$$A_1 := \{(s,t) : s,t \in [t_1-\delta, t_1+\delta], t > s\}, \quad \text{and} \quad A_2 := \{(s,t) : s,t \in [t_2-\delta, t_2+\delta], t < s\}, \tag{4.100}$$

where δ is of order $O(k^{-1/3})$. The integrands of the double integrals

$$\int\int_{A_j} M(s,t)\exp(ik\Psi(s,t))dsdt, \qquad \text{for } \ j = 1,2, \tag{4.101}$$

where M and Ψ are defined in (4.13) and (4.12) respectively, satisfy in the domains A_1 and A_2,

$$k^{-\frac{1}{3}|\boldsymbol{p}|}\left|D^{\boldsymbol{p}}\left(M(s,t)\exp(ik\Psi(s,t))\right)\right| \leq C_{\boldsymbol{p}}\left|D^{\boldsymbol{p}}M(s,t)\right|, \tag{4.102}$$

where $C_{\boldsymbol{p}} > 0$ is a constant independent of k, with the estimates of the derivatives of M obtained in (4.15). Therefore, the integrals (4.101) can be approximated accurately and efficiently using classical quadratures.

Remark 4.19. *Theorem 4.18 essentially states that the domain A_1 (that contains stationary point) is small enough for the integrand in (4.101) to be slowly-varying, i.e. the derivatives of the integrand do not produce any additional powers of k (as stated in (4.102)). However, the integrand in (4.101) is singular as known from (4.15). In practice, classical quadrature rules can be applied on a graded mesh to capture the singularity accurately. In detail, the triangular domain A_1 with the singularities of M on the diagonal can be transformed into a rectangular domain with the singularities of M on one side of the rectangle. Then, the new integral is of the form,*

$$\int\int_R M(\tilde{s},\tilde{t})\exp(ik\Psi(\tilde{s},\tilde{t}))d\tilde{s}d\tilde{t},$$

where R is the rectangular domain. In one of the variables $\tilde{s}$ or $\tilde{t}$, the integrand is smooth and integration can be carried out with classical quadratures. In the other variable, the integrand is singular and classical quadratures can be applied on a graded mesh to compute the integral accurately. The application of graded meshes to singular integrals will be discussed in Chapter 5 in more detail later. The integration over the domain A_2 can be performed identically to the technique described above. In the numerical examples presented in Chapter 6, classical Clenshaw-Curtis rule (in its standard form and over a graded mesh) is used in these domains.

The domains A_1 and A_2 are plotted in Figure 4.13. The proof of Theorem 4.18 requires a few intermediate results. We first show in Lemma 4.20 that at points (t_1,t_1) and (t_2,t_2), the phase function $\Psi(s,t)$ defined in (4.12) and its first and second derivatives, vanish. Then in Proposition 4.21 we find k-dependent estimates on higher derivatives of Ψ. Finally, using Faa di Bruno's formula in 2D (that we have used before in the proof of Theorem

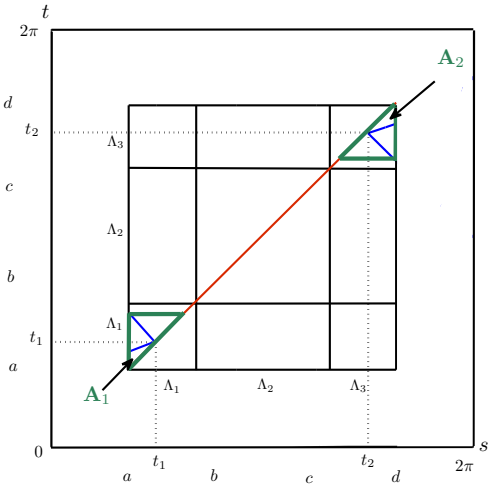


Figure 4.13: *The boundaries of the triangular domains A_1 and A_2 are plotted in green.*

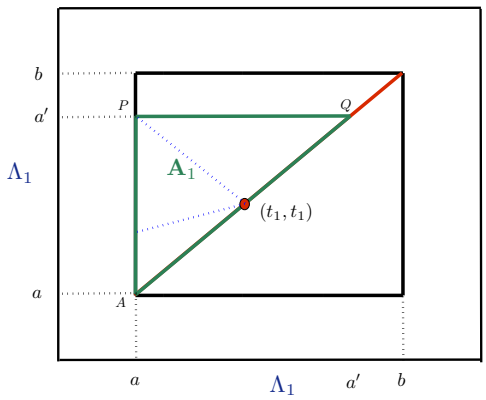


Figure 4.14: *The subdomain A_1 represents the domain where the integrand is slowly-oscillatory and is defined in (4.100).*

4.3, see (4.39)) we prove Theorem 4.18.

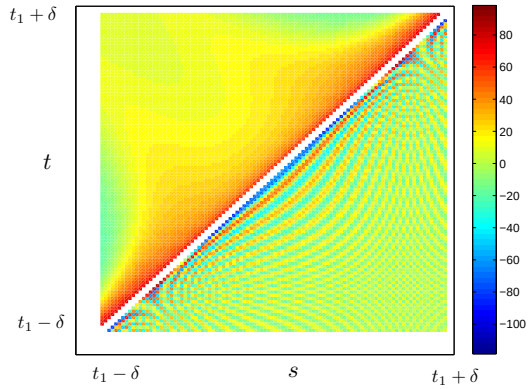


Figure 4.15: *The behaviour of the integrand in the domain $[t_1-\delta, t_1+\delta]\times[t_1-\delta, t_1+\delta]$. Above the diagonal, the integrand of (4.11) is slowly-varying, while below the diagonal it is highly-oscillatory.*

Denote

$$F(x^-, y) := \lim_{\varepsilon\to 0} F(x-\varepsilon, y) \quad \text{and} \quad F(x^+, y) := \lim_{\varepsilon\to 0} F(x+\varepsilon, y).$$

Lemma 4.20. *The phase-function Ψ defined in (4.12) has the following properties:*

$$\begin{aligned} &\Psi(t_1^-, t_1) = 0, \quad \Psi_s(t_1^-, t_1) = 0, \quad \Psi_t(t_1^-, t_1) = 0, \\ &\Psi_{ss}(t_1^-, t_1) = 0, \quad \Psi_{st}(t_1^-, t_1) = 0, \quad \Psi_{tt}(t_1^-, t_1) = 0, \end{aligned} \tag{4.103}$$

and

$$\Psi(t_2^+, t_2) = 0, \quad \Psi_s(t_2^+, t_2) = 0, \quad \Psi_t(t_2^+, t_2) = 0,$$
$$\Psi_{ss}(t_2^+, t_2) = 0, \quad \Psi_{st}(t_2^+, t_2) = 0, \quad \Psi_{tt}(t_2^+, t_2) = 0. \tag{4.104}$$

Proof. Denote the function $\boldsymbol{\rho}(s,t)$ as follows,

$$\boldsymbol{\rho}(s,t) := \frac{\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|}.$$

Differentiating $\boldsymbol{\rho}(s,t)$ we obtain,

$$\begin{aligned}
\frac{d}{ds}\boldsymbol{\rho}(s,t) &= \frac{1}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|}\left(\boldsymbol{\gamma}'(s) - \left(\boldsymbol{\rho}(s,t)\cdot\boldsymbol{\gamma}'(s)\right)\boldsymbol{\rho}(s,t)\right), \\
\frac{d}{dt}\boldsymbol{\rho}(s,t) &= \frac{1}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|}\left(-\boldsymbol{\gamma}'(t) + \left(\boldsymbol{\rho}(s,t)\cdot\boldsymbol{\gamma}'(t)\right)\boldsymbol{\rho}(s,t)\right)
\end{aligned}$$

Also, the function $\boldsymbol{\rho}(s,t)$ has the following properties:

$$\boldsymbol{\rho}(t_1^-, t_1) = \lim_{\varepsilon\to 0}\frac{\boldsymbol{\gamma}(t_1-\varepsilon) - \boldsymbol{\gamma}(t_1)}{|\boldsymbol{\gamma}(t_1-\varepsilon) - \boldsymbol{\gamma}(t_1)|} = -\frac{\boldsymbol{\gamma}'(t_1)}{|\boldsymbol{\gamma}'(t_1)|} = \boldsymbol{a}, \tag{4.105}$$

and

$$\boldsymbol{\rho}(t_2^+, t_2) = \lim_{\varepsilon\to 0}\frac{\boldsymbol{\gamma}(t_2+\varepsilon) - \boldsymbol{\gamma}(t_2)}{|\boldsymbol{\gamma}(t_2+\varepsilon) - \boldsymbol{\gamma}(t_2)|} = \frac{\boldsymbol{\gamma}'(t_2)}{|\boldsymbol{\gamma}'(t_2)|} = \boldsymbol{a}. \tag{4.106}$$

The equalities in (4.105) and (4.105) follow because the incident plane wave is tangential to the boundary Γ at $\boldsymbol{\gamma}(t_1)$ and $\boldsymbol{\gamma}(t_2)$.

Now, differentiating Ψ defined in (4.12), we obtain

$$\begin{aligned}
\Psi_t(s,t) &= \boldsymbol{\gamma}'(t)\cdot(\boldsymbol{a} - \boldsymbol{\rho}(s,t)), \quad \Psi_s(s,t) = \boldsymbol{\gamma}'(s)\cdot(\boldsymbol{\rho}(s,t) - \boldsymbol{a}), \\
\Psi_{tt}(s,t) &= \boldsymbol{\gamma}''(t)\cdot(\boldsymbol{a} - \boldsymbol{\rho}(s,t)) + \frac{1}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|}\left(|\boldsymbol{\gamma}'(t)|^2 - (\boldsymbol{\gamma}'(t)\cdot\boldsymbol{\rho}(s,t))^2\right), \\
\Psi_{ss}(s,t) &= \boldsymbol{\gamma}''(s)\cdot(\boldsymbol{\rho}(s,t) - \boldsymbol{a}) + \frac{1}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|}\left(|\boldsymbol{\gamma}'(s)|^2 - (\boldsymbol{\gamma}'(s)\cdot\boldsymbol{\rho}(s,t))^2\right), \\
\Psi_{st}(s,t) &= \frac{1}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|}\left(-\boldsymbol{\gamma}'(s)\cdot\boldsymbol{\gamma}'(t) + (\boldsymbol{\gamma}'(s)\cdot\boldsymbol{\rho}(s,t))(\boldsymbol{\gamma}'(t)\cdot\boldsymbol{\rho}(s,t))\right),
\end{aligned}$$

Hence using (4.105) and (4.106), and the fact that $\gamma'(t_i)$, is in the direction $\boldsymbol{a}$ for $i = 1, 2$, the result follows. □

Proposition 4.21. *The derivatives of the phase-function Ψ can be bounded as follows:*

$$\left|\left(\frac{\partial}{\partial s}\right)^n\left(\frac{\partial}{\partial t}\right)^m \Psi(s,t)\right| \le C_{n,m}k^{-1+\frac{1}{3}(n+m)}, \qquad \textit{for} \;\; (s,t)\in A_1. \tag{4.107}$$

Proof. Expanding the function $\Psi(s,t)$ in a Taylor series around the point (t_1^-, t_1) and using Lemma 4.20, we obtain

$$\begin{aligned}\Psi(s,t) &= \Psi(t_1^-,t_1) + (s-t_1)\Psi_s(t_1^-,t_1) + (t-t_1)\Psi_t(t_1^-,t_1) + \frac{(s-t_1)^2}{2!}\Psi_{ss}(t_1^-,t_1)\\ &+ \frac{(t-t_1)^2}{2!}\Psi_{tt}(t_1^-,t_1) + 2\frac{(t-t_1)(s-t_1)}{2!}\Psi_{st}(t_1^-,t_1) + \frac{(s-t_1)^3}{3!}\Psi_{sss}(t_1^-,t_1)\\ &+ 3\frac{(s-t_1)^2(t-t_1)}{3!}\Psi_{sst}(t_1^-,t_1) + 3\frac{(s-t_1)(t-t_1)^2}{3!}\Psi_{stt}(t_1^-,t_1)\\ &+ \frac{(t-t_1)^3}{3!}\Psi_{ttt}(t_1^-,t_1) + O(|s-t_1|^4) + O(|t-t_1|^4).\end{aligned}$$

Since the functions $\Psi_{sss}(s,t)$, $\Psi_{sst}(s,t)$, $\Psi_{stt}(s,t)$, $\Psi_{ttt}(s,t)$ are bounded independently of k, and since the diameter of A_1 is $O(k^{-1/3})$, we deduce that $|\Psi(s,t)| \le c_1 k^{-1}$. Similarly,

$$|\Psi_s(s,t)| \le c_2 k^{-2/3}, \quad |\Psi_t(s,t)| \le c_3 k^{-2/3},$$

$$|\Psi_{tt}(s,t)| \le c_4 k^{-1/3}, \quad |\Psi_{st}(s,t)| \le c_5 k^{-1/3}, \quad |\Psi_{ss}(s,t)| \le c_6 k^{-1/3},$$

where the constants c_j, $j = 1,\dots,6$ are independent of k. All the remaining derivatives of Ψ are bounded independently of k. Therefore, for $(s,t) \in A_1$,

$$\left|\left(\frac{\partial}{\partial s}\right)^n \left(\frac{\partial}{\partial t}\right)^m \Psi(s,t)\right| \le C_{n,m} \min\{k^{-1+\frac{1}{3}(n+m)}, 1\}.$$

Hence the result follows. □

proof of Theorem 4.18.

Proof. Applying product rule twice, we obtain

$$D^{\boldsymbol{p}}\left[M(s,t)\exp(ik\Psi(s,t))\right] = \sum_{n_1=0}^{p_1}\sum_{n_2=0}^{p_2} C_{\boldsymbol{n},\boldsymbol{p}}\left(D^{\boldsymbol{p}-\boldsymbol{n}}M(s,t)\right)\left(D^{\boldsymbol{n}}\left[\exp(ik\Psi(s,t))\right]\right), \tag{4.108}$$

where $\boldsymbol{n} = (n_1,n_2)^T$. In order to obtain estimates on $D^{\boldsymbol{p}}\left[M(s,t)\exp(ik\Psi(s,t))\right]$ we require estimates on the derivatives of $\exp(ik\Psi(s,t))$. For this, we require Faa di Bruno's formula.

For integers $p_1, p_2 \ge 0$, we introduce the set

$$S := \{1,\dots,p_1,p_1+1,\dots,p_1+p_2\}.$$

We denote $\boldsymbol{p} = (p_1,p_2)^T$. A partition of S is a set (that we denote by σ) of non-empty, non-overlapping subsets of S whose union is all of S. The number of sets in σ is denoted $|\sigma|$. For each $\beta \in \sigma$ we let β_1 denote the number of elements of β which are $\le p_1$ and let

β_2 denote the number of elements of β which are $> p_1$ and $\leq p_1 + p_2$. We denote $|\beta|$ as $\beta_1 + \beta_2$.

Then, using Faa di Bruno's formula, we deduce,

$$\left(\frac{\partial}{\partial s}\right)^{p_1}\left(\frac{\partial}{\partial t}\right)^{p_2}\{\exp(ik\Psi(s,t))\} = \sum_{\sigma}\exp(ik\Psi(s,t))k^{|\sigma|}\prod_{\beta\in\sigma}\left(\frac{\partial}{\partial s}\right)^{\beta_1}\left(\frac{\partial}{\partial t}\right)^{\beta_2}\{\Psi(s,t)\}, \tag{4.109}$$

where the sum is over all possible partition of σ.

Now consider a typical term in the sum (4.109),

$$\begin{aligned} k^{|\sigma|}\left|\prod_{\beta\in\sigma}\left(\frac{\partial}{\partial s}\right)^{\beta_1}\left(\frac{\partial}{\partial t}\right)^{\beta_2}\{\Psi(s,t)\}\right| &\leq k^{|\sigma|}\prod_{\beta\in\sigma}k^{-1+\frac{1}{3}|\beta|} \\ &= Ck^{|\sigma|}k^{-|\sigma|}k^{\frac{1}{3}|\boldsymbol{p}|}, \end{aligned} \tag{4.110}$$

with C independent of k, s and t. Therefore, a typical term in (4.109) is bounded by $Ck^{\frac{1}{3}|\boldsymbol{n}|}$, which proves

$$D^{\boldsymbol{n}}_{(s,t)}\{\exp(ik\Psi(s,t))\} \leq C_{\boldsymbol{n}}k^{\frac{1}{3}|\boldsymbol{n}|}. \tag{4.111}$$

Returning to equation (4.108), we deduce that $|D^{\boldsymbol{p}}[M(s,t)\exp(ik\Psi(s,t))]|$ is bounded by a linear combination of

$$C_{\boldsymbol{p},\boldsymbol{n}}k^{\frac{1}{3}|\boldsymbol{n}|}\left|D^{\boldsymbol{p}-\boldsymbol{n}}M(s,t)\right|,$$

for $n_1 = 0, ..., p_1$ and $n_2 = 0, ..., p_2$, with $\boldsymbol{n} = (n_1, n_2)$. Multiplying each term by $k^{-\frac{1}{3}|\boldsymbol{p}|}$, we deduce that $k^{-\frac{1}{3}|\boldsymbol{p}|}|D^{\boldsymbol{p}}[M(s,t)\exp(ik\Psi(s,t))]|$ is bounded by a linear combination of

$$C_{\boldsymbol{p},\boldsymbol{n}}k^{\frac{1}{3}(|\boldsymbol{n}|-|\boldsymbol{p}|)}\left|D^{\boldsymbol{p}-\boldsymbol{n}}M(s,t)\right| \leq \tilde{C}_{\boldsymbol{p},\boldsymbol{n}}\left|D^{\boldsymbol{p}-\boldsymbol{n}}M(s,t)\right|,$$

since $k^{\frac{1}{3}(|\boldsymbol{n}|-|\boldsymbol{p}|)} \leq 1$. Hence (4.102) follows. □

In Figure 4.15, the integrand of (4.11) is plotted on the domain $[t_1 - \delta, t_1 + \delta] \times [t_1 - \delta, t_1 + \delta]$ for the case of an elliptic scatterer. On the diagonal, the integrand has a log-singularity. Above the diagonal, the integrand is slowly-oscillatory and below the diagonal the integrand is highly-oscillatory.

The analogous result can be proved in a similar way for the domain A_2. Therefore, the double integrals over A_1 and A_2 can be computed using conventional quadratures.

By separating the domain $\Lambda_1 \times \Lambda_1$ into above-the-diagonal and below-the-diagonal sub-domains, we have shown that

- above the diagonal, in the domain A_1 defined in (4.100), the integral (4.11) is slowly-varying- see Figure 4.14 and Figure 4.15;

- below the diagonal, the integral (4.11) is highly-oscillatory and Hypothesis A is satisfied, see Remark C.8 in Section C of Appendix.

In the next example, we consider the integration over the above-the-diagonal domain of $(\Lambda_1 \times \Lambda_1)$ containing A_1.

4.4.1 Integration over the transition domains

Let us consider computing the integral

$$J_k := \int\int_{(\Lambda_1 \times \Lambda_1)^+} M(s,t) \exp\left(ik\Psi(s,t)\right) dtds,$$

where the domain of integration is defined as

$$(\Lambda_1 \times \Lambda_1)^+ = \{(s,t) : s,t \in [a,b], t > s\}. \tag{4.112}$$

The domain $(\Lambda_1 \times \Lambda_1)^+$ is illustrated in Figure 4.16.

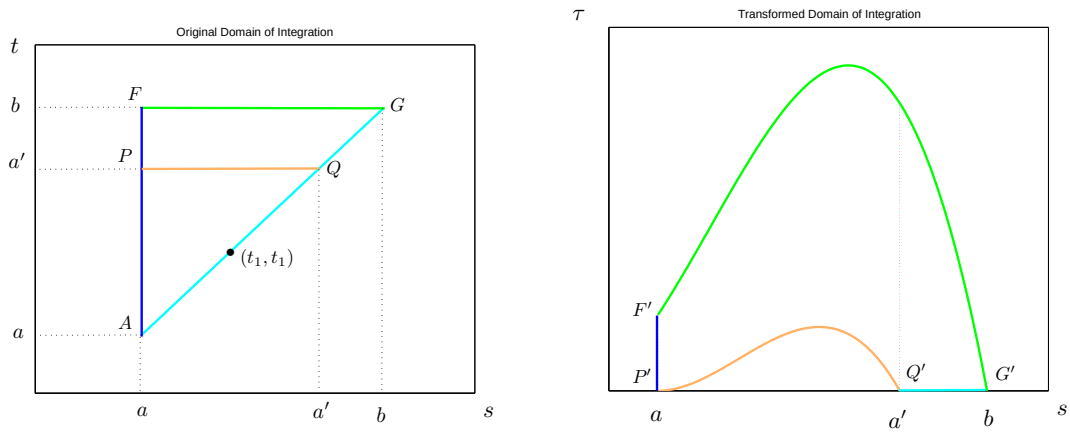


Figure 4.16: *On the left, is the plot of the original domain $(\Lambda_1 \times \Lambda_1)^+$ separated into two subdomains by a line PQ: the triangle APQ and the remaining trapezium FGPQ; the right plot corresponds to the transformed domain of the subdomain FGPQ.*

We separate the domain $(\Lambda_1 \times \Lambda_1)^+$ into two:

- the triangular domain APQ plotted in Figure 4.16 that represents $O(k^{-1/3})$ region around the point (t_1^-, t_1), this is the same as A_1 in (4.100) and
- the trapezium FGPQ plotted in Figure 4.16, where Hypothesis A is satisfied.

We write the integral J_k as a sum of two integrals,

$$J_k := \int\int_{APQ} M(s,t)\exp\left(ik\Psi(s,t)\right)dtds + \int\int_{FGPQ} M(s,t)\exp\left(ik\Psi(s,t)\right)dtds.$$

The integral over the domain APQ can be efficiently computed using classical quadrature rules since the integral is slowly-varying.

Over the trapezium FGPQ (that consists of a rectangle and a triangle), the integral in J_k is highly-oscillatory. The transformed domain of integration for the the trapezium FGPQ is illustrated in Figure 4.16. Then

$$\int\int_{FGPQ} M(s,t)\exp\left(ik\Psi(s,t)\right)dtds = \sum_{j=0}^{2} I_k^{[\tau_j,\tau_{j+1}]}[F_{j+1}],$$

where, with $a' = t_1 + \delta$,

$$\begin{aligned}
F_1(\tau) &= \int_a^{\left(\psi^{[a']}\right)^{-1}(\tau)} H(s,\tau)\,ds + \int_{\left(\psi^{[a']}\right)^{-1}(\tau)}^{\left(\psi^{[b]}\right)^{-1}(\tau)} H(s,\tau)\,ds, \qquad \tau \in [\tau_0,\tau_1],\\
F_2(\tau) &= \int_a^{\left(\psi^{[b]}\right)^{-1}(\tau)} H(s,\tau)\,ds, \qquad \tau \in [\tau_1,\tau_2],\\
F_3(\tau) &= \int_{r_1(\tau)}^{r_2(\tau)} H(s,\tau)\,ds, \qquad \tau \in [\tau_2,\tau_3],
\end{aligned}$$

where

$$\begin{aligned}
r_1(\tau) &= \left(\psi^{[b]}\right)^{-1}(\tau) \quad \in [a,\xi_2],\\
r_2(\tau) &= \left(\psi^{[b]}\right)^{-1}(\tau) \quad \in [\xi_2,b].
\end{aligned}$$

In the Table 4.3, we tabulate the boundaries of the inner integral to complete the Table 4.2. The domain in the table is defined as follows:

$$(\Lambda_1 \times \Lambda_1)^* = \left\{(s,t) : s \in [a,t], t \in [a',b]\right\}, \qquad a := t_1 - \delta, \quad a' := t_1 + \delta.$$

Also ξ_a and ξ_b are defined as in (4.95):

$$\left(\psi^{[a]}\right)'(\xi_a) = 0, \qquad \text{and} \qquad \left(\psi^{[b]}\right)'(\xi_b) = 0.$$

original domain	singularities	τ-interval	lower boundary	upper boundary
$(\Lambda_1 \times \Lambda_1)^*$	(L) $\tau_0 = 0$,	$[\tau_0, \tau_1]$	a	$\left(\psi^{[a']}\right)^{-1}(\tau)$
	(S) $\tau_1 = \Psi(\xi_a, a)$		$\left(\psi^{[a']}\right)^{-1}(\tau)$	$\left(\psi^{[b]}\right)^{-1}(\tau)$
	$\tau_2 = \Psi(a, b)$	$[\tau_1, \tau_2]$	a	$\left(\psi^{[b]}\right)^{-1}(\tau)$
	(S) $\tau_3 = \tau_{\max} = \Psi(\xi_b, b)$	$[\tau_2, \tau_{\max}]$	$\left(\psi^{[b]}\right)^{-1}(\tau)$	$\left(\psi^{[b]}\right)^{-1}(\tau)$

Table 4.3: *Table of transformed domains*

Chapter 5

Filon-Clenshaw-Curtis quadrature

5.1 Introduction

In Chapter 4, we described the double integrals that arise from the Galerkin discretisation of the boundary integral equation (2.13). The double integrals can be written as a sum of repeated integrals with the outer integral of the form,

$$I_k^{[a,b]}[f] = \int_a^b f(x)\exp(ikx)dx, \tag{5.1}$$

where f may have algebraic singularities at the end points and the wavenumber k can be large. In this chapter we describe methods which compute the integrals of the form (5.1) so that they:

- efficiently compute the integral with high accuracy for both high and low wavenumbers k;
- allow algebraic singularities in the integrand so that the rate of convergence of the method is not affected by the singularity;
- permit error analysis resulting in error bounds that are explicit in the wavenumber k, the number of quadrature points N, the regularity of f and other parameters that account for the singularities in the integrand.

Many classical integration rules are based on the polynomial interpolation of the integrand. When applied to (5.1), classical rules typically require a fixed number of quadrature points N per wavelength. The wavelength is inversely proportional to the wavenumber k. Hence, in order to maintain the accuracy, these rules require N to grow linearly with k as $k \to \infty$. Therefore, the classical quadrature rules fail to be efficient when approximating $I_k^{[a,b]}[f]$ for large wavenumbers k.

There are at least three classes of methods designed for computing (5.1) accurately as $k \to \infty$. These are asymptotic methods, Filon-type methods (Filon 1928) and Levin-type methods (Levin 1982). The main property of all these methods is that the error bounds decrease with inverse powers of k as $k \to \infty$. Recently, the computation of the highly-oscillatory integrals has been a subject of substantial renewed research and the three

classes of methods have been revisited, analysed and enhanced. Indeed, new numerical methods have been developed such as numerical steepest descent method [60]. We refer the reader for a general overview of these methods to [59].

5.1.1 Outline of the chapter

The plan for this chapter is as follows. In the remainder of this section, we will discuss the development of Filon-type methods and their properties and other novel numerical methods for computing (5.1). We will motivate the choice of Filon-Clenshaw-Curtis (FCC) quadrature for our target application, namely the integrals arising in Chapter 4. In Section 5.2, we will discuss the accuracy of Chebyshev interpolating polynomials and derive a non-composite version of FCC quadrature and discuss an efficient method for the computation of its weights. In Section 5.3, we will derive two error bounds for the FCC quadrature that are explicit in k, N and regularity of f: the first in terms of the Sobolev norm of the cosine transform of f and the second in terms of Chebyshev norm of the derivatives of f. The second error bound is one of the novel results of this chapter. The reason for the development of error bounds in terms of the regularity of the function f rather that its cosine mapping will become apparent in Section 5.4. In the section, we consider the case when the function f has algebraic singularities and we extend the error analysis to the case when FCC quadrature is applied on graded meshes in order to resolve the singularity in f. This is the second novel result of this chapter. We conclude the chapter with the Section 5.5 where we carry out numerical experiments.

5.1.2 Survey of existing methods

Filon-type methods

The idea behind Filon-type quadratures is simple, the function $f(x)$ in (5.1) is replaced by a suitable interpolating function so that the integral of $\exp(ikx)$ against this interpolant is easily computed. In our case, the function f is replaced by an algebraic polynomial $(P_N f)$ that interpolates the function f at $x_0,\ldots,x_N$:

$$(P_N f)(x_j) = f(x_j), \quad j = 0, \ldots, N. \tag{5.2}$$

The resulting integral can then be integrated exactly to obtain an approximation to $I_k^{[a,b]}[f]$:

$$I_k^{[a,b]}[f] := \int_a^b f(x)\exp(ikx)dx \simeq \int_a^b (P_N f)(x)\exp(ikx)dx =: Q_{k,N}^{[a,b]}[f]. \tag{5.3}$$

Note that writing

$$(P_N f)(x) = \sum_{n=0}^{N} \alpha_n p_n(x),$$

where $\{p_0(x), \ldots, p_1(x)\}$ is a suitable polynomial basis and α_n, $n = 0, \ldots, N$ are constants that can be determined from (5.2), the quadrature can be written as follows

$$Q_{k,N}^{[a,b]}[f] := I_k^{[a,b]}(P_N f) = \sum_{n=0}^{N} \alpha_n \mu_n(k) \tag{5.4}$$

where

$$\mu_n(k) := I_k^{[a,b]} p_n = \int_a^b p_n(x)\exp(ikx)dx,$$

are called *moments*.

Filon [47] was the first to suggest in 1928 computing integrals of the form (5.1) by replacing the function f with a quadratic polynomial which takes the same values as f at the end points and the midpoint of the interval of integration. This rule can be understood as a modified Simpson's rule or Filon-Simpson quadrature. In the limit as $k \to 0$, the weights of the Filon-Simpson quadrature reduce to the weights of the classical Simpson rule. The idea has been subsequently enhanced by Luke 1954 [73] and Flinn 1960 [48] where higher order polynomial approximations of f (with degree ≤ 10) were considered. However the convergence rates have not been discussed by neither Filon, Luke nor Flinn.

In Iserles et al 2003 [62] the Filon-type methods have been generalized: the function f in (5.1) is replaced by a suitable approximating polynomial and the analysis of the *asymptotic order of convergence* of such Filon-type methods is derived in [62] for $k \to \infty$.

The asymptotic order of convergence of a method represents the rate at which the error of the method decreases with respect to inverse powers of k as $k \to \infty$. In [62] the error of Filon-type methods applied to (5.3) with $[a, b] = [-1, 1]$ is proved to be of asymptotic order $O(k^{-2})$, provided the quadrature points include the end points of the interval of integration. This has been further generalized in Iserles et al 2005 [63] to the case when the oscillating term in the integrand in (5.1) is replaced with the *non-canonical* exponential term $\exp(ikg(x))$ where the function g may have a finite number of stationary points of order 1 in $[-1, 1]$:

$$I_k^{g,[-1,1]}[f] := \int_{-1}^{1} f(x) \exp(ikg(x))dx. \tag{5.5}$$

The corresponding moments $\mu_n(k) = I_k^{g,[-1,1]}[p_n]$ are assumed to be known. The error of the Filon-type quadrature is shown in this case to be of asymptotic order $O(k^{-3/2})$, provided the quadrature points include the end points and all stationary points.

In Iserles and Nørset 2005 [64] a *generalized Filon method* for integrals with non-canonical oscillating term $\exp(ikg(x))$ has been developed with error bounds that decay with similarly high negative powers of k by approximating f with the Hermite interpolating polynomial ϕ. By definition, the Hermite interpolating polynomial satisfies:

$$\phi^{(m)}(x_j) = f^{(m)}(x_j), \quad m = 0, \ldots, \theta_j, \quad j = 1, \ldots, N, \tag{5.6}$$

where $a = x_1, \ldots, x_N = b$ are the quadrature points of multiplicities $\theta_1, \ldots, \theta_N$:

$$\phi(x) = \sum_{j=1}^{N} \sum_{m=1}^{\theta_j} \alpha_{m,j}(x) f^{(m)}(x_j), \tag{5.7}$$

where $\alpha_{m,j}(x)$ is a polynomial of degree $\sum_{j=1}^{N} \theta_j - 1$ satisfying

$$\begin{aligned} \alpha_{m,j}^{(n)}(x_l) &= 0 \text{ for all } n = 0, 1, \ldots, \theta_j - 1, \text{ and } l = 1, 2, \ldots, N \\ \alpha_{m,j}^{(j)}(x_m) &= 1 \end{aligned}$$

For $s = \min\{\theta_1, \theta_N\}$, convergence of asymptotic order $O(k^{-s-1})$ is proved in [64] when the phase-function g has no stationary points and of order $O(k^{-s-1/(r+1)})$ when g has a stationary point of order r, provided a quadrature point coincides with the stationary point. The asymptotic order of convergence is derived in [64] as follows:

- In the case when the oscillating term in the integrand is canonical, $\exp(ikx)$, the integral $I_k^{[a,b]}[f - \phi]$, where ϕ is a Hermite interpolating polynomial, can simply be

integrated by parts s-times,

$$\begin{aligned} I_k^{[a,b]}[f-\phi] &:= \sum_{m=0}^{s} \frac{(-1)^m}{(ik)^{m+1}} \Big((f^{(m)} - \phi^m)(b) \exp(ikb) \\ &- (f^{(m)} - \phi^{(m)})(a) \exp(ika) \Big) + O\left(\frac{1}{k^{s+1}}\right), \end{aligned}$$

provided f is sufficiently smooth. The first s terms of the resulting expansion then vanish by construction, see (5.6) with $s = \min\{\theta_1, \theta_N\}$. The error of such Filon-type method is then of order: $I_k^{[a,b]}[f-\phi] \backsim O(k^{-s-1})$ as $k \to \infty$.

- In the case when the oscillating term in the integrand is non-canonical, $\exp(ikg(x))$, and a stationary point $x = \xi$ is present in the domain of integration (we assume $\xi \neq a, b$), the following technique is used for *removing* the singularity originating from the stationary point,

$$\begin{aligned} I_k^{g,[a,b]}[f] &:= \int_a^b f(x) \exp(ikg(x)) dx \qquad (5.8) \\ &= \int_a^b f(\xi) \exp(ikg(x)) dx + \int_a^b (f(x) - f(\xi)) \exp(ikg(x)) dx \\ &= f(\xi) \int_a^b \exp(ikg(x)) dx \\ &+ \frac{1}{ik} \left\{ \frac{\exp(ikg(b))}{g'(b)} [f(b) - f(\xi)] - \frac{\exp(ikg(a))}{g'(a)} [f(a) - f(\xi)] \right\} \\ &- \frac{1}{ik} \int_a^b \left(\frac{f(x) - f(\xi)}{g'(x)} \right)' \exp(ikg(x)) dx, \qquad (5.9) \end{aligned}$$

where $[f(\cdot) - f(\xi)]/g'(\cdot)$ is a smooth function. Iterating this procedure on the latter integral leads to an asymptotic expansion of the integral Then, provided the quadrature points include the stationary point: $x_j = \xi$, the first s terms of the asymptotic expansion of $I_k^{g,[a,b]}[f-\phi]$ vanish by construction. Moreover, the first moment satisfies

$$\mu_0(k) := \int_a^b \exp(ikg(x)) dx = O\left(k^{-1/(r+1)}\right), \qquad (5.10)$$

hence the asymptotic order of the method follows. To see (5.10), we expand the function g in Taylor series around ξ. The first r terms in the expansion vanish since ξ is a stationary point of order r. Hence we obtain,

$$\mu_0(k) := \int_a^b \exp\left(ik \frac{g^{(r+1)}(\tilde{\xi})}{(r+1)!} (x - \tilde{\xi})^{r+1} \right) dx.$$

Now, by making the change of variables from x to t as

$$t = g^{(r+1)}(\tilde{\xi})(x-\tilde{\xi})^{r+1}/(r+1)!,$$

we deduce

$$\begin{aligned} |\mu_0(k)| &\leq \left| \frac{1}{r+1} \int_{-\infty}^{\infty} \left(\frac{|g^{(r+1)}(\tilde{\xi})|}{(r+1)!|t|} \right)^{\frac{1}{r+1}-1} \exp(ikt)dt \right| \\ &\leq \frac{2|g^{(r+1)}(\tilde{\xi})|}{(r+1)!(r+1)} \left| \int_0^{\infty} \frac{1}{t^{\frac{1}{r+1}-1}} \exp(ikt)dt \right| \\ &\leq c \left(\frac{1}{k} \right)^{\frac{1}{r+1}}, \end{aligned}$$

where c is a constant independent of r.

The singularity removing technique can be applied to integrals with the oscillators that have any finite number of stationary points by partitioning the interval $[a, b]$ into a number of subintervals with a single stationary point residing in a single subinterval.

In a number of subsequent papers, the error analysis of the Filon-type methods concentrated on their asymptotic order of convergence.

Moment-free Filon-type methods

Several authors addressed the problem of the computation of moments: when the oscillating term is of the form $\exp(ikg(x))$, it is not always possible to compute the moments $\mu_n(k)$ defined in (5.4) exactly. In Olver 2007 [83] a moment-free Filon-type method has been developed for approximating

$$I_k^{g,[-1,1]}[f] = \int_{-1}^{1} f(x) \exp(ikg(x))dx, \tag{5.11}$$

where the phase-function g has a stationary point of order r at $x = 0$. A set of basis functions $\{\phi_1, \ldots, \phi_M\}$ is constructed here that interpolates f in (5.8) and a sufficient number of its derivatives at the stationary point. Moreover, the moments of the resulting Filon-type method are guaranteed to be known. The basis functions $\{\phi_1, \ldots, \phi_M\}$ are defined in [83, Lemma 1] for $x \in [-1, 1]$.

Lemma 5.1. *[83, Lemma 1]*
Let

$$\phi_n(x) := D_{r,n}(x) \frac{k^{-\frac{1+n}{r}}}{r} e^{-ikg(x)+\frac{1+n}{2r}i\pi} \left[\Gamma\left(\frac{1+n}{r}, -ikg(x) \right) - \Gamma\left(\frac{1+n}{r}, 0 \right) \right],$$

where

$$D_{r,n}(x) = \begin{cases} (-1)^k, & \text{sgn } x < 0, \quad \textit{and } r \textit{ even} \\ (-1)^k e^{-\frac{1+n}{r} i\pi}, & \text{sgn } x < 0, \quad \textit{and } r \textit{ odd} \\ -1, & \textit{otherwise} \end{cases}$$

Then $\phi_n \in C^\infty[-1,1]$ *and, for* $\mathcal{L}[F] := F' + iwg'F$,

$$\mathcal{L}[\phi_n] := \text{sgn } (x)^{r+n+1} \frac{|g(x)|^{\frac{1+n}{r}} g'(x)}{r}.$$

Furthermore, $\mathcal{L}[\phi_n] \in C^\infty[-1,1]$. *Finally,*

$$I_k^{g,[-1,1]}[\mathcal{L}[\phi_n]] = \phi_n(1)e^{ikg(1)} - \phi_n(-1)e^{ikg(-1)}. \tag{5.12}$$

Let us define a smooth function ψ as follows,

$$\psi(x) := \sum_{n=1}^{M} c_n \mathcal{L}[\phi_n](x),$$

where the coefficients c_n can be determined by solving a system

$$\psi^{(m)}(x_j) = f^{(m)}(x_j), \quad m = 0, \ldots, \theta_j - 1, \quad j = 1, \ldots, N, \quad M = \sum_{j=1}^{N} \theta_j.$$

Then we can approximate $I_k^{g,[-1,1]}[f]$ using Filon-type quadrature that can be found explicitly:

$$Q_{k,N}^{g,[-1,1]}[f] := I_k^{g,[-1,1]}[\psi] = \sum_{n=1}^{N} c_n I_k^{g,[-1,1]}[\mathcal{L}[\phi_n]],$$

where $I_k^{g,[-1,1]}[\mathcal{L}[\phi_n]]$ are known from (5.12). The asymptotic order of the error in approximating (5.11) of this method is $O(k^{-s-1/r})$, where $s = \min\{\theta_1, \theta_N\}$.

Another moment-free method has been developed by Xiang 2007 [102] that has similar asymptotic convergence properties and also uses information on higher derivatives of f. Here the authors propose a change of variables $y = g(x)$ which is valid provided the function g is monotone in all of $[a,b]$, so that

$$\begin{aligned} I_k^{g,[a,b]}[f] &:= \int_a^b f(x) \exp(ikg(x)) \\ &= \int_{g(a)}^{g(b)} \frac{f(g^{-1}(y))}{|g'(g^{-1}(y))|} \exp(iky) dy \simeq \int_{g(a)}^{g(b)} \phi(y) \exp(iky) dy, \end{aligned} \tag{5.13}$$

where ϕ is Hermite interpolating polynomial defined in (5.7). The oscillating term in (5.13) is canonical, i.e. $\exp(iky)$ and the moments of the Filon-type method can be computed

analytically as we have demonstrated in Chapter 4. The asymptotic rate of convergence in this case is $O(k^{-s-1})$, where $s = \min\{\theta_1, \theta_N\}$. On the other hand, if $g(x)$ has a stationary point, ξ, of order r, the following change of variables is employed in [102],

$$y^{r+1} = g(x) - g(\xi). \tag{5.14}$$

Then, the highly-oscillatory integral can be written as

$$\begin{aligned} I_k^{g,[a,b]}[f] &:= \exp(ikg(\xi)) \left(\int_a^{\xi} + \int_{\xi}^{b} \right) f(x) \exp(ik\,[g(x) - g(\xi)])dx \\ &= \exp(ikg(\xi)) \left[\int_0^{(g(b)-g(\xi))^{\frac{1}{r+1}}} f_2(y) \exp(iky^{r+1})dy \right. \\ &- \left. \int_0^{(g(a)-g(\xi))^{\frac{1}{r+1}}} f_2(y) \exp(iky^{r+1})dy \right], \end{aligned} \tag{5.15}$$

where functions f_1 and f_2 containing the Jacobian of the transformation (5.14) are smooth functions, provided f and g are smooth. Then, the moments of the Hermite-based Filon-type quadrature applied to (5.15) can be computed using incomplete Gamma functions [3]. The asymptotic order of convergence of such method is $O(k^{-s-1/(r+1)})$.

It is evident that Filon-type methods can achieve arbitrarily high asymptotic orders of convergence for $k \to \infty$ provided the derivatives of the integrand function f are available at a certain set of points. It is often the case however, that the derivatives of the function f can not be easily obtained (for example if f is a composition of several complicated functions). This is the case in the integrals we aim to compute in Chapter 4.

Levin-type methods and the steepest descent method

There are also a number of other methods that are designed to compute the integrals of the form (5.1) for $k \to \infty$.

Levin-type methods [70] reduce the problem of approximating the integral (5.1) with a simpler problem of finding an antiderivative function $F(x)$ such that,

$$\begin{aligned} I_k^{g,[a,b]}[f] := \int_a^b f(x) \exp(ikg(x))dx &= \int_a^b \frac{d}{dx} \left(F(x) \exp(ikg(x)) \right) \\ &= F(b) \exp(ikg(b)) - F(a) \exp(ikg(a)). \end{aligned}$$

We seek a solution F of the ordinary differential equation

$$\mathcal{L}[F] = f,$$

with no boundary conditions prescribed [70], where the differential operator $\mathcal{L}$ is defined

as

$$\mathcal{L}[F](x) = F'(x) + ikg'(x)F(x). \tag{5.16}$$

The function F is typically approximated using collocation method. Let $\{\psi_1, \ldots, \psi_N\}$ be linearly independent basis functions on $[a, b]$. A collocation approximation to the solution F is defined as

$$v(x) := \sum_{m=1}^{N} c_m \psi_m(x),$$

where the coefficients c_m are determined from a system of linear equations,

$$\mathcal{L}[v](x_1) = f(x_1), \quad \ldots, \quad \mathcal{L}[v](x_N) = f(x_N). \tag{5.17}$$

where $x_1, \ldots, x_N$ is a set of collocation nodes. Then, the integral $I_k^{g,[a,b]}$ can be approximated by

$$Q_{k,N}^{Levin,g,[a,b]} = v(b)\exp(ikg(b)) - v(a)\exp(ikg(a)). \tag{5.18}$$

Levin-type methods become more accurate as $k \to \infty$ [71] and as the equation (5.16) is solved more accurately. Levin-type methods do not require computation of moments. The asymptotic order of convergence of Levin-type methods (for any choice of linearly independent basis functions $\{\psi_1, \ldots, \psi_N\}$) is known to be of order $O(k^{-2})$ [71] provided no stationary points are present and collocation points include the end points of integration.

In Olver 2010 [84] a GMRES-Levin-type method was developed for highly-oscillatory integrals with or without stationary points. The collocation method for solving the differential equation (5.16) leads to a system of equations (5.17). The author proposes a stable method for solving this system using the GMRES algorithm [91]. The rate of convergence of the GMRES algorithm is shown in [84] to increase as $k \to \infty$. The resulting Levin-type method has an asymptotic order of convergence $O(k^{-2})$ and the computational cost decreases with k as $k \to \infty$ to maintain the accuracy. Furthermore, higher order of asymptotic convergence of this method can be achieved if the derivative information of f is used.

In Huybrechs and Vandewalle 2006 [60] a *numerical steepest descent method* was developed where the integration over the real interval $[a, b]$ in $I_k^{g,[a,b]}[f]$ is converted into a path in the complex plane so that the oscillations of the integrand are essentially removed, provided f is analytic. For example, let us consider the integral with a canonical oscillator: $I_k^{[a,b]}[f]$ where f is analytic. Since the value of a line integral along a path between two points in the complex plane does not depend on a path taken provided that the integrand is analytic (Cauchy's Theorem), we can write,

$$I_k^{[a,b]}[f] := \int_a^b f(x)\exp(ikx)dx = \int_L f(x)\exp(ikx)dx,$$

where L is a new path of integration consisting of three lines in the complex plane: $L = L_1 \cup L_2 \cup L_3$, where

L_1 is a vertical path represented by $h_a(p) := a + ip$, $p \in [0, P]$,

L_2 is a vertical path represented by $h_b(p) := b + ip$, $p \in [0, P]$,

L_3 is a horizontal path connecting $h_a(P)$ and $h_b(P)$.

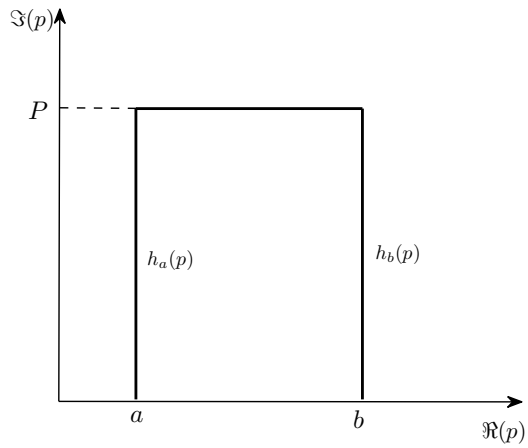


Figure 5.1: *The deformed contour of integration in the complex plane.*

Thus, setting $P \to \infty$, the integral over L_3 vanishes since for all $x \in L_3$, we can write $x = t + iP$, where $t \in [a, b] \in \mathbb{R}$, and

$$\exp(ikx) = \exp(ikt)\exp(-kP) \to 0, \quad \text{as} \quad P \to \infty.$$

With $h_a(p) = a + ip$, $p \in [0, \infty)$, the integral over L_1 is of the form,

$$\int_{L_1} f(x)\exp(ikx)dx = e^{ika}\int_0^\infty f(a + ip)\exp(-kp)dp.$$

Similarly, the integral over L_2 can be written as a non-oscillatory integral. Thus, we obtain,

$$\begin{aligned} I_k^{[a,b]}[f] &= \int_{L_1} f(x)\exp(ikx)dx + \int_{L_2} f(x)\exp(ikx)dx \\ &= e^{ika}\int_0^\infty f(a + ip)\exp(-kp)dp + e^{ikb}\int_0^\infty f(b + ip)\exp(-kp)dp. \end{aligned}$$

Both integrals on the right hand side can be efficiently approximated using Gauss-Laguerre quadrature.

For more general integrals, i.e. with non-canonical oscillators, $I_k^{g,[a,b]}[f]$, the new path $h_x(p)$, $p \in [0, \infty)$ of integration in the complex plane can be found using inverse of g,

provided there are no stationary points in $[a, b]$: $h_x(p) = g^{-1}(g(x) + ip)$, so that

$$\text{Im } g(h_x(p)) > 0.$$

Then, over the new path of integration, the integrand has the desired decay properties: for fixed x,

$$\exp(ikg(h_x(p))) := \exp(ikg(x)) \exp(-kp).$$

The resulting integrals are non-oscillatory and well behaved provided f and g are analytic and f does not grow exponentially large in the complex plane. Integrals can be efficiently approximated using Gauss-Laguerre quadrature. The asymptotic order of the error of this method is $O(k^{-2n-1})$ when n-point Gauss-Laguerre quadrature is used.

The case when stationary points in $[a, b]$ of order r are present in g is also discussed in [60] and the error is derived to be of asymptotic order $O(k^{-2n-1+r/(r+1)})$. The method however is not applicable to non-smooth integrand functions f.

Subsequently, in Huybrechs and Vandewalle 2007 [61], the numerical steepest descent method was applied in the context of boundary integral methods for high frequency scattering problems. The collocation approach for the discretization of the boundary integral equation (2.15) was used.

Numerical experiments in [61] indicate that method has decreasing error with growing wavenumber k.

Quadrature rules for singular integrals. In this Chapter, we will propose a numerical method for computing (5.1) for the case when the function f has algebraic singularities. These types of integrals often arise in methods for solving scattering problems.

Techniques for the numerical computation of integrals with singular integrands include:

- *singularity removing transformations*: using a change of variables, the singularity in the function $f(x)$ can be removed. The disadvantage of such technique applied in the context of Filon-type integration is that the resulting moments $\mu_n(k)$ can not generally be analytically computed.

  In Hascelik 2009 [56] Filon-type methods for computing integrals with algebraic singularities have been considered for a special case, when the integrand is of the form: $x^\alpha f(x) \exp(ikx^{-r})$ for $x \in [0, 1]$, $r > 0$ and $r + \alpha > -1$. The function f (assumed to be smooth), is replaced by Hermite interpolating polynomial of degree N. The moments of the resulting Filon-type method are computed using extended exponential integrals Ei, defined in (5.1.2) [3]. The error bound for the Filon quadrature is given in terms of k, α, N and derivatives of f and inverse powers of r;

- *graded meshes*: instead of applying a quadrature rule on the whole interval of integration, a graded mesh is created. The quadrature rule is then applied on each subinterval of the graded mesh. Graded meshes are constructed so that the length of the subintervals becomes smaller towards the singularity, so that the errors of approximations on each subinterval are uniform and small [74]. For a general idea of graded meshes technique applied to singular integrals without oscillations and their analysis, we refer the reader to [76].

In our target application, the most straightforward technique for computing (5.1) with singular f is to apply the graded meshes technique with Filon-type methods, see Section 5.4. In this chapter, we will extend the error analysis of Filon-type methods to obtain explicit error bounds in terms of wavenumber k, the number of quadrature points N and regularity of f.

5.1.3 Motivation for the chapter

We will consider Filon-Clenshaw-Curtis quadrature for the approximation of the integrals in the form

$$I_k^{[-1,1]}[f] := \int_{-1}^{1} f(x)\exp(ikx)dx \tag{5.19}$$

with a view to implementing this quadrature in numerical methods for boundary integral equations for scattering problems as introduced in Chapter 2 and 4. There are a number of reasons for our choice.

Firstly, the Filon-Clenshaw-Curtis quadrature inherits the attractive features of the Clenshaw-Curtis method (1960) [31]. Clenshaw-Curtis quadrature is a classical method that approximates an integral of the form:

$$I^{[-1,1]}[f] = \int_{-1}^{1} f(x)dx, \qquad \text{with} \qquad Q_N^{[-1,1]}[f] = \int_{-1}^{1} P_N f(x)dx,$$

where $P_N f(x)$ is a polynomial that interpolates $f(x)$ at Chebyshev points:

$$x_{j,N} = \cos\left(\frac{j\pi}{N}\right), \quad j = 0, 1, \ldots, N.$$

The error bounds of the Clenshaw-Curtis quadrature were derived in [99, Theorem 5.1] to be of the form,

Theorem 5.2. *Let Clenshaw-Curtis quadrature $Q_N^{[-1,1]}[f]$ be applied to a function $f \in C([-1,1])$. If f, f',...,$f^{(m+1)}$ are absolutely continuous on $[-1,1]$ for some $m \geq 1$, then*

for all sufficiently large N,

$$\left|I^{[-1,1]}[f] - Q_N^{[-1,1]}[f]\right| \leq \frac{32}{15\pi m}\frac{1}{(2N+1-m)^m}\int_{-1}^{1}\frac{\left|f^{(m+1)}(x)\right|}{\sqrt{1-x^2}}dx. \tag{5.20}$$

From Theorem 5.2, we deduce that if in addition to the requirements of the theorem, $m \leq N+1$, then the following holds

$$\left|I^{[-1,1]}[f] - Q_N^{[-1,1]}[f]\right| \leq C\left(\frac{1}{N}\right)^m\int_{-1}^{1}\frac{\left|f^{(m+1)}(x)\right|}{\sqrt{1-x^2}}dx. \tag{5.21}$$

One of the features of the Clenshaw-Curtis quadrature is that its weights can be precomputed using FFT allowing the quadrature to be applied to many different functions with minimal additional computations. Another feature of the quadrature is that it is "nested", i.e. if the quadrature is computed for $N+1$ points, then computing the quadrature for $2N+1$ points only requires N additional evaluations of f.

The second reason for our choice of the Filon-Clenshaw-Curtis quadrature is that the quadrature does not require the evaluation of the derivatives of f. In our target application, the derivatives of the integrand cannot be easily obtained. As we discussed in Chapter 4, we are required to compute the integral (5.19) with a function $f(x)$ of the form (4.83) that is in itself is an integral with a rather complicated integrand.

Finally, as we will prove later in this chapter, Filon-Clenshaw-Curtis quadrature applied on graded meshes can accurately approximate the integral (5.19) when $f(x)$ has algebraic singularities. We will derive an error bound of the Filon-Clenshaw-Curtis quadrature that is explicit in the wavenumber k, the number of quadrature points N, the number of subintervals in the graded mesh M, the grading parameter q and the regularity of f. This is a novel contribution of this Chapter.

Remark 5.3. *It is worth mentioning that Classical Clenshaw-Curtis quadratures have also been considered for the computation of the integrals of the form:*

$$I_k[f] = \int_{-1}^{1} f(x)w(x)dx,$$

with the function $w(x)$ allowed to contain singularities or oscillations. These methods are often reffered to as modified Clenshaw-Curtis rules or product integration rules with Clenshaw-Curtis points and were developed for example in [94, 50, 89, 88, 90] (1972-1984). However, when $w(x) = \exp(ikx)$, error bounds that are explicit in k and the regularity of f were not presented in any of these references.

In the context of product integration rules with Clenshaw-Curtis points, a number of authors have also addressed the problem of the computation of the moments $\mu_n(k)$, $n =$

$0, 1, \ldots, N$:

$$\mu_n(k) = \int_{-1}^{1} T_n(x) \exp(ikx) dx, \tag{5.22}$$

where $T_n(x)$ denotes the Chebyshev polynomial of degree n that we discuss in detail in Section 5.2.1. For an overview of these methods, for the case when $w(x) = \exp(ikx)$, we refer the reader to Evans and Webster 1999 [44]. However, these methods were proved to be stable only for certain values of k and n. Notably, the moments for the modified Clenshaw-Curtis rule with $w(x) = \exp(ikx)$, can be obtained by a recurrence relation that is derived from the forward recurrence relation for the Chebyshev polynomials. This method of obtaining the moments is only stable for $N < k$. Recently, however, a method for the computation of the moments has been presented in Dominguez et al 2010 [41] that is proved to be stable for all N and k. In Section 5.2.1, we will discuss this method in detail. A public domain Matlab code which implements this algorithm is available at [38].

5.2 Filon-Clenshaw-Curtis quadrature

In this section, we will describe the Filon-Clenshaw-Curtis (FCC) quadrature and discuss the properties of the Chebyshev interpolating polynomial. Also, we will discuss the computation of the weights of the FCC quadrature as presented in [41].

The FCC quadrature $Q_{k,N}^{[-1,1]}[f]$ approximates the integral of the form,

$$I_k^{[-1,1]}[f] := \int_{-1}^{1} f(x)\exp(ikx)dx,$$

by replacing the function f with Chebyshev interpolating polynomial $(P_N f)$ of degree N, see Definition 5.6. Then, we obtain

$$Q_{k,N}^{[-1,1]}[f] := \int_{-1}^{1} (P_N f)(x)\exp(ikx)dx. \tag{5.23}$$

Notation 5.4. *In the remainder of this section, for convenience, we will replace the notation $I_k^{[-1,1]}[f]$ with $I_k[f]$ and similarly, $Q_{k,N}^{[-1,1]}[f]$ with $Q_{k,N}[f]$.*

5.2.1 Chebyshev interpolating polynomial

Definition 5.5. *The Chebyshev polynomial of the first kind of degree n is a function $T_n(x)$, $x \in [-1,1]$, defined by*

$$T_n(x) = \cos(n\arccos(x)). \tag{5.24}$$

Chebyshev polynomials of the first kind can also be defined as functions that satisfy the recurrence relation:

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_n(x) &= 2xT_{n-1}(x) - T_{n-2}(x), \quad n \geq 1, \end{aligned} \tag{5.25}$$

Therefore $T_n(x)$ is a polynomial of degree n. Additionally, Chebyshev polynomials satisfy the recurrence relation involving their derivatives:

$$2T_n(x) = \frac{1}{n+1}T'_{n+1}(x) - \frac{1}{n-1}T'_{n-1}(x). \tag{5.26}$$

In the following Definition, we will define the **Chebyshev interpolating polynomial**.

Definition 5.6. *The Chebyshev interpolating polynomial $(P_N f)$ for the function f is*

defined as follows:

$$(P_N f)(x_j) = f(x_j), \quad at \quad x_j = \cos\left(\frac{j\pi}{N}\right), \quad j = 0, ..., N, \tag{5.27}$$

where x_j are called the Chebyshev points.

Lemma 5.7. *The Chebyshev interpolating polynomial can be written as follows,*

$$(P_N f)(x) = \frac{1}{2}c_0 + c_1 T_1(x) + ... + c_{N-1}T_{N-1}(x) + \frac{1}{2}c_N T_N(x), \tag{5.28}$$

where the coefficients c_j, $j = 0, ..., N$ are given as:

$$\begin{aligned} c_j &:= c_j[f] = \frac{2}{N}\sum_{m=0}^{N}{}'' f(x_m)T_j(x_m) \\ &= \frac{1}{N}\left[f(-1)(-1)^j + 2\sum_{m=1}^{N-1} f(x_m)\cos\left(\frac{mj\pi}{N}\right) + f(1)\right], \end{aligned} \tag{5.29}$$

where $\sum''$ denotes the sum with the first and the last terms halved.

Proof. We use the orthogonality relation

$$\sum_{n=0}^{N}{}'' \cos\left(\frac{nm\pi}{N}\right)\cos\left(\frac{nj\pi}{N}\right) = \begin{cases} 0 & : j \neq m \\ N & : j = m = 0 \text{ or } N, , \\ N/2 & : j = m \neq 0 \text{ or } N. \end{cases}$$

to deduce

$$\begin{aligned} (P_N f)(x_j) &= \frac{2}{N}\sum_{n=0}^{N}{}''\left(\sum_{m=0}^{N}{}'' f(x_m)\cos\left(\frac{nm\pi}{N}\right)\right)\cos\left(\frac{nj\pi}{N}\right) \\ &= \frac{2}{N}\sum_{m=0}^{N}{}'' f(x_m)\left[\sum_{n=0}^{N}{}''\cos\left(\frac{nm\pi}{N}\right)\cos\left(\frac{nj\pi}{N}\right)\right] \\ &= f(x_j). \end{aligned}$$

Hence the result follows. □

Chebyshev interpolating polynomials are related to *trigonometric interpolating polynomials* as we will see in Section 5.2.2, which in turn are related to truncated Fourier series. We will use these equivalences in the error analysis of the Filon-Clenshaw-Curtis quadrature given in Section 5.2.3.

5.2.2 Trigonometric interpolating polynomials

Definition 5.8. *For $g : [-1,1] \to \mathbb{R}$, we denote g_c as the cosine mapping of the function g,*

$$g_c(\theta) = g(\cos\theta). \tag{5.30}$$

Definition 5.9. *For $f : [a,b] \to \mathbb{R}$, we define $\tilde{f} : [-1,1] \to \mathbb{R}$ by*

$$\tilde{f}(t) = f\left(\frac{b+a}{2} + \frac{b-a}{2}t\right). \tag{5.31}$$

So $\tilde{f} = f$ if $[a,b] = [-1,1]$. We denote the cosine mapping of the function $\tilde{f}$ as $\tilde{f}_c$. So $\tilde{f}_c = f_c$ if $[a,b] = [-1,1]$.

The cosine mapping of the interpolating Chebyshev polynomial $(P_N f)$ defined in (5.28) is

$$(P_N f)_c(\theta) = \frac{1}{2}c_0 + c_1\cos(\theta) + ... + c_{N-1}\cos((N-1)\theta) + \frac{1}{2}c_N\cos(N\theta). \tag{5.32}$$

Hence $(P_N f)_c \in \operatorname{span}\{1, \cos(\theta), ..., \cos(N\theta)\}$. Thus, it is a *trigonometric polynomial of degree N* and it interpolates $f_c(\theta)$ at $N+1$ equally spaced points

$$\theta_j := j\pi/N, \quad j = 0, 1, ..., N.$$

In Theorem 5.12, we will present the error of the trigonometric polynomial interpolant at equally spaced points in the periodic Sobolev space H^m_{per} which we define in Definition 5.10. We will use Theorem 5.12 to analyse the error when f_c is approximated by $(P_N f)_c$ when proving k-dependent error bounds of the FCC quadrature, see Theorem 5.14.

Recall that for $\phi \in L^2[-\pi,\pi]$ we define the corresponding Fourier series as follows:

$$\phi(\theta) = \sum_{\mu=-\infty}^{+\infty} \hat{\phi}(\mu)\exp(i\mu\theta), \quad \text{where} \quad \hat{\phi}(\mu) := \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(\theta)\exp(-i\mu\theta)d\theta. \tag{5.33}$$

Definition 5.10. *For $m \geq 0$, the 2π-periodic Sobolev space H^m_{per} is the space of all functions $\phi \in L^2[-\pi,\pi]$ such that $\|\phi\|^2_{H^m_{per}} < \infty$, where*

$$\|\phi\|^2_{H^m_{per}} := |\hat{\phi}(0)|^2 + \sum_{\mu\in\mathbb{Z},\mu\neq 0} |\mu|^{2m}|\hat{\phi}(\mu)|^2. \tag{5.34}$$

H^m_{per} is a Hilbert space with respect to the scalar product

$$(\phi,\varphi)_{H^m_{per}} = \hat{\phi}(0)\overline{\hat{\varphi}(0)} + \sum_{\mu\in\mathbb{Z},\mu\neq 0} |\mu|^{2m}\hat{\phi}(\mu)\overline{\hat{\varphi}(\mu)}.$$

Now, let $\mathcal{Q}_N u$, be the interpolating trigonometric polynomial of degree N that satisfies the following conditions:

$$\begin{aligned} \mathcal{Q}_N u &\in \mathbb{T}_N := \operatorname{span}\{\exp(i\mu\theta),\ \mu = -N, \dots, N\}, \\ (\mathcal{Q}_N u)(\theta_j) &= u(\theta_j),\ \text{where } \theta_j = \frac{\pi j}{N},\quad j = -N, \dots, N. \end{aligned} \tag{5.35}$$

Since the function f_c is even, the operator $\mathcal{Q}_N$ applied to f_c is only composed in terms of cosines,

$$\begin{aligned} \mathcal{Q}_N f_c &\in \operatorname{span}\{\cos(\mu\theta),\ \mu = 0, \dots, N\}, \\ (\mathcal{Q}_N f_c)(\theta_j) &= f_c(\theta_j),\ \text{where } \theta_j = \frac{\pi j}{N},\quad j = -N, \dots, N. \end{aligned} \tag{5.36}$$

Lemma 5.11. *For all $f \in C([-1,1])$,*

$$(P_N f)_c = \mathcal{Q}_N f_c. \tag{5.37}$$

Proof. $P_N f$ is a polynomial of degree $\leq N$. It follows that

$$(P_N f)_c \in \operatorname{span}\{\cos(\mu\theta) : \mu = 0, ..., N\}.$$

Moreover,

$$P_N f(x_j) = f(x_j),$$

with x_j given as in (5.27). Therefore,

$$(P_N f)_c\left(\frac{\pi j}{N}\right) = f_c\left(\frac{\pi j}{N}\right), \qquad j = 0, ..., N.$$

Therefore, by uniquesnes, (5.37) follows. □

The following theorem provides an error estimate for the interpolating trigonometric polynomial $\mathcal{Q}_N$.

Theorem 5.12. ***[Error of the trigonometric polynomial approximation, [92, Theorem 8.3.1]]*** *For $u \in H^{\mu}_{per}$, $\mu \in \mathbb{R}$, $\mu > 1/2$,*

$$\|u - \mathcal{Q}_N u\|_{H^{\lambda}_{per}} \leq \gamma_\mu \left(\frac{1}{N}\right)^{\mu-\lambda} \|u\|_{H^{\mu}_{per}}, \quad (0 \leq \lambda \leq \mu), \tag{5.38}$$

where

$$\gamma_\mu := \left(1 + \sum_{j=1}^{\infty} \frac{1}{j^{2\mu}}\right)^{1/2} < \infty.$$

5.2.3 The moments of the Filon-Clenshaw-Curtis quadrature

In this section, we will discuss the practical computation of the weights and moments of the FCC quadrature defined below in (5.40) and (5.41) respectively. As we have discussed earlier in Section 5.1, the stable computation of moments, in general, can be a challenging problem associated with Filon-type methods. In this case, the moments $\omega_n(k)$, $n = 0, \ldots, N$ can be found recursively using forward recurrence relation for the Chebyshev polynomials (5.26). This algorithm however is stable only when $N \leq k$.

In [41], a two-phase method is proved to be stable, using the forward recurrence relation for Chebyshev polynomials to compute moments (5.41) for $n \leq \min\{k, N\}$ and an algorithm that solves a tridiagonal system of size about $N - k$ to compute the remaining moments (Oliver's algorithm [79]), for $k < n \leq N$. We will discuss this method in more detail in this section.

We begin by deriving the FCC quadrature.

Lemma 5.13. *The Filon-Clenshaw-Curtis quadrature defined in (5.23) (see also Notation 5.4) can be written as follows:*

$$Q_{k,N}[f] := \sum_{m=0}^{N}{}'' W_m(k) f(x_m), \tag{5.39}$$

where the weights $W_m(k)$, $m = 0, 1, \ldots, N$ are defined as:

$$W_m(k) = \frac{2}{N} \sum_{n=0}^{N}{}'' \omega_n(k) \cos\left(\frac{mn\pi}{N}\right), \tag{5.40}$$

and the moments $\omega_n(k)$, $n = 0, 1, \ldots, N$ are:

$$\omega_n(k) = \int_{-1}^{1} T_n(x) \exp(ikx) dx, \tag{5.41}$$

Proof. We can verify equations (5.40) and (5.41) by substituting the Chebyshev interpolating polynomial $(P_N f)$ defined in (5.28) and (5.29) into the integral $I_k[(P_N f)]$,

$$\begin{aligned} Q_{k,N}[f] &= \int_{-1}^{1} (P_N f)(x) \exp(ikx) dx \\ &= \sum_{j=0}^{N}{}'' c_j \int_{-1}^{1} T_j(x) \exp(ikx) dx \\ &= \sum_{m=0}^{N}{}'' \left(\frac{2}{N} \sum_{j=0}^{N}{}'' \omega_j(k) \cos\left(\frac{mj\pi}{N}\right) \right) f(x_m). \end{aligned} \tag{5.42}$$

Therefore, the result follows. □

Two-phase method for computing moments in the FCC quadrature

Depending on N and k, the moments ω_n defined in (5.41) for $n = 0, 1, \ldots, N$ can be computed using the two-phase method described below. If $N \leq k$, then all the moments ω_n, $n = 0, 1, \ldots, N$ can be computed using the algorithm described in the first phase. On the other hand, if $N > k$, then ω_n for $0 \leq n < k$ are computed using the first phase and the remaining moments: ω_n for $k < n \leq N$ are computed using the second phase.

First phase. Using the classical recurrence relation for Chebyshev polynomials (5.26), we obtain:

$$2\omega_n(k) = \rho_{n+1}(k) - \rho_{n-1}(k), \quad n \geq 2, \tag{5.43}$$

where

$$\rho_n(k) := \frac{1}{n} \int_{-1}^{1} T_n{}'(x) \exp(ikx) dx, \quad n \geq 1. \tag{5.44}$$

Integrating formula (5.41) for $\omega_n(k)$ by parts, we obtain,

$$\omega_0(k) = \gamma_0(k) \;\text{ and }\; \omega_n(k) := \gamma_n(k) - \frac{n}{ik}\rho_n(k), \quad n \geq 1, \tag{5.45}$$

where

$$\begin{aligned} \gamma_n(k) &:= \frac{1}{ik}\, T_n(x) \exp(ikx)|_{x=-1}^{x=1} = \frac{1}{ik}\left[\exp(ik) - (-1)^n \exp(-ik)\right] \\ &= \begin{cases} \frac{2\sin k}{k}, & \text{for even } n, \\ \frac{2\cos k}{k}, & \text{for odd } n. \end{cases} \end{aligned} \tag{5.46}$$

Combining the equations (5.43) and (5.45), we obtain

$$2\gamma_n(k) - \frac{2n}{ik}\rho_n(k) = \rho_{n+1}(k) - \rho_{n-1}(k), \tag{5.47}$$

with

$$\rho_1(k) := \gamma_0(k), \tag{5.48}$$

$$\rho_2(k) := 2\gamma_1(k) - \frac{2}{ik}\gamma_0(k). \tag{5.49}$$

Thus, to evaluate the moments $\omega_n(k)$ for $n \leq \min\{N, k\}$, we use the recurrence relation (5.47)-(5.49) to compute $\rho_n(k)$. Then $\omega_n(k)$ can be computed using (5.45). This algorithm is proven to be stable for $n < N \leq k$ in Theorem 5.1 and Corollary 5.2 in [41].

Second phase. When $N \geq k$, additional algorithm must be added where $\rho_n(k)$ for $k \leq n \leq N$ can be computed stably. This can be achieved using Oliver's algorithm [79]. Let $n_0 = \lceil k \rceil$, where $\lceil k \rceil$ denotes the smallest integer $\geq x$, and take $M \geq \max\{n_0/2, N/2\}$.

We can find $\rho_n(k)$, $n = n_0, \ldots, 2M-1$, by solving the tridiagonal system for $\boldsymbol{x} \in \mathbb{R}^{2M-n_0}$,

$$A_M(k)\boldsymbol{x} = \boldsymbol{b}_M(k), \tag{5.50}$$

where

$$A_M(k) := \begin{bmatrix} \frac{2n_0}{ik} & 1 & & & \\ -1 & \frac{2(n_0+1)}{ik} & 1 & & \\ & -1 & \frac{2(n_0+1)}{ik} & 1 & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & \frac{2(2M-1)}{ik} \end{bmatrix}, \quad \boldsymbol{b}_M(k) := \begin{bmatrix} 2\gamma_{n_0}(k) + \rho_{n_0-1}(k) \\ 2\gamma_{n_0+1}(k) \\ 2\gamma_{n_0+2}(k) \\ \\ 2\gamma_{2M-1}(k) - \rho_{2M}(k) \end{bmatrix}. \tag{5.51}$$

The solution of (5.50) is a vector $\boldsymbol{\rho}_M(k)$:

$$\boldsymbol{\rho}_M(k) := [\rho_{n_0}(k), \rho_{n_0+1}(k), \rho_{n_0+2}(k), \ldots, \rho_{2M-1}(k)]^T.$$

The components of the right-hand side vector $\boldsymbol{b}_M(k)$ can be obtained using the following: equation (5.46) for $\gamma_n(k)$, $n = n_0, \ldots, 2M-1$, and the value of $\rho_{n_0-1}(k)$ can be obtained from the first phase. Finally, $\rho_{2M}(k)$ can be obtained, for sufficiently large M, using an asymptotic argument described in Theorem 3.1 [41].

Computing the Filon-Clenshaw-Curtis quadrature using FFT.
The FCC quadrature rule can be implemented in $O(N \log N)$ operations. The first and second phases of the algorithm for computing weights require $O(N)$ operations and $Q_{k,N}[f]$ in (5.39) can be computed in $O(N \log N)$ operations using FFT.

To see this, define the vectors $\boldsymbol{\omega}_N$ and $\boldsymbol{f}_N$ of length N as follows:

$$\begin{aligned} \boldsymbol{\omega}_N &= [\omega_0(k)/2, \omega_1(k), \ldots, \omega_{N-1}(k), \omega_N(k)/2]^T \\ \boldsymbol{f}_N &= [f(x_0)/2, f(x_1), \ldots, f(x_{N-1}), f(x_N)/2]^T. \end{aligned}$$

and define an $N \times N$ matrix $\boldsymbol{T}$ as

$$(\boldsymbol{T})_{j,m} = \cos\left(\frac{jm\pi}{N}\right), \quad j, m = 0, \ldots, N.$$

Then, the quadrature is obtained by straightforward dot product:

$$Q_{k,N}[f] = \boldsymbol{\omega}_N^T \boldsymbol{T} \boldsymbol{f}_N = (\boldsymbol{T}\boldsymbol{\omega}_N)^T \boldsymbol{f}_N.$$

The matrix-vector product $\boldsymbol{T}\boldsymbol{\omega}_N$ is a discrete cosine mapping that can be precomputed

in $O(N \log N)$ operations. Then the quadrature $Q_{k,N}[f]$ can be applied to many different functions without any additional computations of the weights.

5.3 Error of the Filon-Clenshaw-Curtis quadrature

In this section, we will state and prove two theorems which present error estimates for the Filon-Clenshaw-Curtis quadrature. These error estimates are explicit in k, N and regularity of f. In the first theorem, the error is bounded by a periodic Sobolev norm of the cosine mapping of f. The cosine transform of the function is defined in Definition 5.8. The second theorem is analogous to the first but the bound involves a Chebyshev weighted norm of the derivatives of f.

The reasons why we need two error estimates for the Filon-Clenshaw-Curtis quadrature will be explained in Remark 5.16.

5.3.1 Error estimate in terms of the periodic Sobolev norm of f_c

The following theorem is proved in [41].

Theorem 5.14. ***[Error of the Filon-Clenshaw-Curtis quadrature in terms of Sobolev norm of the cosine mapping of*** *f] For $s = 0, 1, 2$ and for all $m \geq m_0 > \max\{1/2, \rho(s)\}$, there exist a constant $C_{m_0} > 0$ such that*

$$|I_k[f] - Q_{k,N}[f]| \leq C_{m_0} \left(\frac{1}{k}\right)^s \left(\frac{1}{N}\right)^{m-\rho(s)} \|f_c\|_{H^m_{per}}, \tag{5.52}$$

where $\rho(s)$ is

$$\rho(s) = \begin{cases} 0, & \text{if } s = 0, \\ 1, & \text{if } s = 1, \\ 7/2 & \text{if } s = 2. \end{cases}$$

Proof. We prove the three cases separately: when $s = 0$, when $s = 1$ and when $s = 2$. We begin by defining the error function $e : [-1, 1] \to \mathbb{R}$ as follows,

$$e(x) := (f - P_N f)(x).$$

The cosine mapping of e is of the form,

$$e_c(\theta) := e(\cos\theta) = (f - P_N f)(\cos\theta) = (f_c - \mathcal{Q}_N f_c)(\theta),$$

where $\mathcal{Q}_N f_c \in \mathbb{T}_N$ is the cosine mapping of $P_N f$, see Lemma 5.11. Therefore,

$$e_c'(\theta) := -e'(\cos\theta)\sin\theta \tag{5.53}$$

$$e_c''(\theta) := -e''(\cos\theta)\sin^2\theta - e'(\cos\theta)\cos\theta. \tag{5.54}$$

From Theorem 5.12 we obtain the bounds on e_c and its derivative,

$$\|e_c\|_{L^2([-\pi,\pi])} \leq c_1 \|e_c\|_{H^0_{per}} \leq C_m \frac{1}{N^m} \|f_c\|_{H^m_{per}}, \tag{5.55}$$

$$\|e_c^{'}\|_{L^2([-\pi,\pi])} \leq c_2 \|e_c\|_{H^1_{per}} \leq C_m \frac{1}{N^{m-1}} \|f_c\|_{H^m_{per}}, \tag{5.56}$$

and for $m > n$,

$$\|e_c^{(n)}\|_{L^2([-\pi,\pi])} \leq c_2 \|e_c\|_{H^n_{per}} \leq C_m \frac{1}{N^{m-n}} \|f_c\|_{H^m_{per}}. \tag{5.57}$$

<u>Case $s = 0$.</u> Using (5.55)

$$\begin{aligned} |I_k[f] - Q_{k,N}[f]| &\leq \left| \int_{-1}^{1} (f(x) - P_N f(x)) \exp(ikx) dx \right| \\ &\leq \int_0^{\pi} |(f(\cos\theta) - P_N f(\cos\theta))| \sin\theta d\theta \\ &\leq \frac{1}{2} \int_{-\pi}^{\pi} |e_c(\theta)| \, d\theta \leq \frac{1}{2} \left(\int_{-\pi}^{\pi} |e_c(\theta)|^2 \, d\theta \right)^{1/2} \left(\int_{-\pi}^{\pi} 1^2 d\theta \right)^{1/2} \\ &\leq \sqrt{\frac{\pi}{2}} \|e_c\|_{L^2([-\pi,\pi])} \leq C_m \frac{1}{N^m} \|f_c\|_{H^m_{per}}. \end{aligned} \tag{5.58}$$

<u>Case $s = 1$.</u> We integrate $I^k[f] - Q_{k,N}[f]$ by parts once and obtain:

$$I_k[f] - Q_{k,N}[f] = \frac{1}{ik} \left[(f(x) - P_N f(x)) \exp(ikx) \right]_{x=-1}^{x=1} - \frac{1}{ik} \int_{-1}^{1} (f - P_N f)' (x) \exp(ikx) dx.$$

Since $P_N f(\pm 1) = f(\pm 1)$, the first term vanishes. By making the substitution $x = \cos\theta$ and using (5.53) we obtain a bound for the second term:

$$\begin{aligned} |I_k[f] - Q_{k,N}[f]| &= \frac{1}{k} \left| \int_0^{\pi} (f_c - \mathcal{Q}_N f_c)' (\theta) \exp(ik\cos\theta) d\theta \right| \\ &\leq \frac{\sqrt{\pi}}{k} \left\| (e_c)' \right\|_{L^2,[0,\pi]} \leq C \frac{1}{k} \|e_c\|_{H^1_{per}} \\ &\leq C_m \frac{1}{k} \frac{1}{N^{m-1}} \|f_c\|_{H^m_{per}} \qquad \text{using (5.56).} \end{aligned} \tag{5.59}$$

<u>Case $s = 2$.</u> Define the function $\phi_N(\theta)$:

$$\phi_N(\theta) := \frac{(e_c)'(\theta)}{\sin\theta} = -(f - P_N f)'(\cos\theta) = -e'(\cos\theta). \tag{5.60}$$

Then integrating (5.59) by parts again, we obtain:

$$\begin{aligned}
|I_k[f] - Q_{k,N}[f]| &= \left|\frac{1}{ik}\int_0^\pi (e_c)'(\theta)\exp(ik\cos\theta)d\theta\right| \\
&= \left|\frac{1}{ik}\int_0^\pi \phi_N(\theta)\exp(ik\cos\theta)\sin\theta d\theta\right| \\
&\leq \left|\frac{1}{k^2}\left(\phi_N(\pi)e^{-ik} - \phi_N(0)e^{ik}\right)\right| + \left|\frac{1}{k^2}\int_0^\pi \phi_N'(\theta)\exp(ik\cos\theta)d\theta\right| \\
&= \frac{1}{k^2}|E_1| + \frac{1}{k^2}|E_2|,
\end{aligned}$$

where we denote

$$\begin{aligned}
E_1 &:= \phi_N(\pi)e^{-ik} - \phi_N(0)e^{ik}, \\
E_2 &:= \int_0^\pi \phi_N'(\theta)\exp(ik\cos\theta)d\theta.
\end{aligned}$$

From (5.54) and (5.60), we deduce

$$\begin{aligned}
(e_c)''(0) &= -e'(\cos 0) = \phi_N(0) \\
(e_c)''(\pi) &= e'(\cos\pi) = -\phi_N(\pi).
\end{aligned}$$

Then for all $m \geq 3$, using the Sobolev embedding theorem, we deduce

$$\begin{aligned}
|E_1| \leq |\phi_N(0)| + |\phi_N(\pi)| &= \left|(e_c)''(0)\right| + \left|(e_c)''(\pi)\right| \\
&\leq C\|e_c\|_{H^3_{per}} \leq C_m\frac{1}{N^{m-3}}\|f_c\|_{H^m_{per}}, \qquad \text{by (5.57).}
\end{aligned}$$

Now, it remains to find a bound on $|E_2|$. Since e_c is a even function, we can write it as a cosine series (see Lemma 5.18 later):

$$e_c(\theta) = \widehat{e_c}(0) + 2\sum_{m=1}^{\infty}\widehat{e_c}(m)\cos m\theta, \tag{5.61}$$

where

$$\widehat{e_c}(m) := \frac{1}{\pi}\int_0^\pi e_c(\theta)\cos(m\theta)d\theta, \qquad m \geq 0.$$

From (5.61) it follows,

$$(e_c)'(\theta) = -2\sum_{m=1}^{\infty} m\widehat{e_c}(m)\sin m\theta. \tag{5.62}$$

Substituting (5.62) into (5.60), we obtain

$$\phi_N(\theta) = -2\sum_{m=1}^{\infty} m\widehat{e_c}(m)\frac{\sin m\theta}{\sin\theta}. \tag{5.63}$$

Denote $\sigma(\theta) := (\sin\theta)/\theta$. For $\theta \in [-\pi/2, \pi/2]$, $\sigma(\theta) \geq 2/\pi$ and so for $\theta \in [0, \pi/2]$, and $m \geq 1$,

$$\left|\left(\frac{\sin m\theta}{\sin\theta}\right)'\right| = \left|m\left(\frac{\sigma(m\theta)}{\sigma(\theta)}\right)'\right| = \left|m^2\frac{\sigma'(m\theta)}{\sigma(\theta)} - m\frac{\sigma(m\theta)\sigma'(\theta)}{\sigma^2(\theta)}\right| \leq Cm^2. \tag{5.64}$$

Moreover, writing

$$\frac{\sin m\theta}{\sin\theta} = (-1)^{m-1}\frac{\sin m(\theta-\pi)}{\sin(\theta-\pi)},$$

allows us to extend (5.64) to $\theta \in [0, \pi]$. Therefore,

$$|E_2| \leq \pi\|\phi_N'\|_{L^\infty([0,\pi])} \leq C\sum_{m=1}^{\infty} m^3|\widehat{e}_c(m)|. \tag{5.65}$$

To complete the estimate on E_2, we use the elementary estimates

$$\sum_{m=1}^{N} m^6 < \frac{(N+1)^7}{7} \qquad \text{and} \qquad \sum_{m=N+1}^{\infty}\frac{1}{m^{1+\alpha}} < \frac{1}{\alpha N^\alpha}, \quad (\alpha > 0).$$

Splitting the sum (5.65) for $m \leq N$ and $m > N$, using Cauchy-Schwartz inequality, we deduce for $s \geq s_0 > 7/2$

$$\begin{aligned}
|E_2| &\leq C\left\{\left[\sum_{m=1}^{N} m^6\right]^{1/2}\left[\sum_{m=1}^{N}|\widehat{e}_c(m)|^2\right]^{1/2} + \left[\sum_{m=N+1}^{\infty}\frac{1}{m^{2s-6}}\right]^{1/2}\left[\sum_{m=N+1}^{\infty} m^{2s}|\widehat{e}_c(m)|^2\right]^{1/2}\right\} \\
&\leq C\left\{\left(\frac{1}{N}\right)^{-7/2}\|e_c\|_{H^0_{per}} + \left(\frac{1}{N}\right)^{s-7/2}\|e_c\|_{H^s_{per}}\right\} \\
&\leq C_{s_0}\left(\frac{1}{N}\right)^{s-7/2}\|f_c\|_{H^s_{per}}
\end{aligned}$$

□

The error estimate presented in Theorem 5.14 is not optimal when k is small. The estimate (5.52) implies that as $k \to 0$, the error of the FCC quadrature may blow up with a rate of order k^{-s}, $s = 1, 2$. In the following Corollary, this result is extended to include all possible k.

Corollary 5.15. *Under the conditions of the Theorem 5.14, for $s = 0, 1, 2$ and for all $m \geq m_0 > \max\{1/2, \rho(s)\}$, there exist a constant $C_{m_0} > 0$ such that*

$$|I_k[f] - Q_{k,N}[f]| \leq C_{m_0}\min\left\{1, \left(\frac{1}{k}\right)^s\right\}\left(\frac{1}{N}\right)^{m-\rho(s)}\|f_c\|_{H^m_{per}}. \tag{5.66}$$

Proof. For $k \geq 1$, the result follows from Theorem 5.12. For $k \leq 1$, the result follows from

(5.58). □

Remark 5.16. *Theorem 5.14 provides the error bound of the Filon-Clenshaw-Curtis quadrature in terms of the periodic Sobolev norm of the cosine mapping of f.*

One of the ways of computing integrals (5.1) when f has algebraic singularities is to apply the quadrature on a mesh graded towards the singularity. The idea is that the error of the quadrature can then be made uniform on subintervals and small overall. Investigating the convergence of the Filon-Clenshaw-Curtis quadrature applied on graded meshes requires the error bounds that are explicit in terms of derivatives of f and the length of the interval the quadrature is applied to (see later Corollary 5.25 where we consider integrals over $[a, b]$ rather than $[-1, 1]$). Therefore, we need an alternative to the error bound (5.52). We present this result in Theorem 5.17 below.

5.3.2 Error estimate in terms of Chebyshev weighted norm of $f^{(m)}$

In this section, we will state and prove explicit error estimates for the FCC quadrature in terms of the Chebyshev weighted norm of the derivatives of f.

Theorem 5.17. ***[Error of the Filon-Clenshaw-Curtis quadrature in terms of Chebyshev weighted norm of f^m]*** *Let $f \in C^m([-1, 1])$ be m-times continuously differentiable function. For $s = 0, 1, 2$ let $\rho(s)$ be defined as in Theorem 5.14, and for all m such that $0 \leq m \leq N + 1$, there exist constants [1] $\sigma_{m,N} > 0$, such that*

$$|I_k[f] - Q_{k,N}[f]| \leq \sigma_{m,N} \left(\frac{1}{N}\right)^{m-\rho(s)} \left(\frac{1}{k}\right)^{s} \left[\int_{-1}^{1} \frac{\left|f^{(m)}(t)\right|^2}{\sqrt{1-t^2}} dt\right]^{1/2}. \tag{5.67}$$

The proof of this theorem requires a few definitions and lemmas.

Lemma 5.18. *Consider an even function $\phi \in L^2([-\pi, \pi])$, i.e. $\phi(\theta) = \phi(-\theta)$, $\theta \in [-\pi, \pi]$. The Fourier series for the function ϕ can be written as follows:*

$$\phi(\theta) = \hat{\phi}(0) + 2\sum_{\mu=1}^{\infty} \hat{\phi}(\mu)\cos(\mu\theta), \quad \textit{where} \quad \hat{\phi}(\mu) := \frac{1}{\pi}\int_0^{\pi} \phi(\theta)\cos(\mu\theta)d\theta. \tag{5.68}$$

Furthermore,

$$\|\phi\|^2_{H^m_{per}} = |\hat{\phi}(0)|^2 + 2\sum_{\mu=1}^{\infty} \mu^{2m}|\hat{\phi}(\mu)|^2. \tag{5.69}$$

[1] The size of the constant $\sigma_{m,N}$ does not play role in the error estimates of the composite FCC which we aim to obtain in this chapter and therefore is not displayed explicitly in the error estimate. However, the explicit expression for the constant can be found in Remark 5.22.

Proof. From the definition of Fourier series for general functions in $L^2([-\pi,\pi])$ (5.33), we deduce

$$\begin{aligned}
\hat{\phi}(\mu) &:= \frac{1}{2\pi}\int_{-\pi}^{\pi}\phi(\theta)\exp(i\mu\theta)d\theta \\
&= \frac{1}{2\pi}\int_{-\pi}^{0}\phi(\theta)\exp(i\mu\theta)d\theta + \frac{1}{2\pi}\int_{0}^{\pi}\phi(\theta)\exp(i\mu\theta)d\theta \\
&= \frac{1}{2\pi}\int_{0}^{\pi}\phi(-\theta)\exp(-i\mu\theta)d\theta + \frac{1}{2\pi}\int_{0}^{\pi}\phi(\theta)\exp(i\mu\theta)d\theta \\
&= \frac{1}{\pi}\int_{0}^{\pi}\phi(\theta)\cos(\mu\theta)d\theta. \qquad (5.70)
\end{aligned}$$

Moreover,

$$\hat{\phi}(-\mu) := \frac{1}{\pi}\int_{0}^{\pi}\phi(-\theta)\cos(-\mu\theta)d\theta = \frac{1}{\pi}\int_{0}^{\pi}\phi(\theta)\cos(\mu\theta)d\theta = \hat{\phi}(\mu). \qquad (5.71)$$

Therefore,

$$\begin{aligned}
\phi(\theta) &= \sum_{\mu=-\infty}^{-1}\hat{\phi}(\mu)\exp(i\mu\theta) + \hat{\phi}(0) + \sum_{\mu=1}^{\infty}\hat{\phi}(\mu)\exp(i\mu\theta) \\
&= \hat{\phi}(0) + \sum_{\mu=1}^{\infty}\hat{\phi}(\mu)\left(\exp(i\mu\theta) + \exp(-i\mu\theta)\right) \\
&= \hat{\phi}(0) + 2\sum_{\mu=1}^{\infty}\hat{\phi}(\mu)\cos(\mu\theta).
\end{aligned}$$

Hence, (5.68) follows. Similarly, (5.69) follows by the definition of the 2π-periodic Sobolev space H^m_{per} given in (5.34):

$$\begin{aligned}
\|\phi\|^2_{H^m_{per}} &= \sum_{\mu=-\infty}^{-1}|\mu|^2 m|\hat{\phi}(\mu)|^2 + |\hat{\phi}(0)|^2 + \sum_{\mu=1}^{\infty}|\mu|^2 m|\hat{\phi}(\mu)|^2 \\
&= |\hat{\phi}(0)|^2 + 2\sum_{\mu=1}^{\infty}|\mu|^2 m|\hat{\phi}(\mu)|^2, \qquad \text{using (5.71).}
\end{aligned}$$

□

In Lemma 5.19, we will describe the property of the Fourier coefficients for the cosine mapping function g_c.

Lemma 5.19. *Let g_c be the cosine mapping of the function $g \in C([-1,1])$. Then g_c is an even function in $L^2([-\pi,\pi])$. The coefficients of the Fourier cosine series of g_c satisfy the*

following property: for $\mu \in \mathbb{N}$, $\mu \neq 0$,

$$\widehat{g_c}(\mu) = \frac{1}{2\mu}\left[\widehat{(g')_c}(\mu-1) - \widehat{(g')_c}(\mu+1)\right]. \tag{5.72}$$

Proof. In order to show this, we use Lemma 5.18 for the definition of $\widehat{g_c}(\mu)$ and integrate by parts:

$$\begin{aligned}\widehat{g_c}(\mu) &= \frac{1}{\pi}\int_0^\pi g_c(\theta)\cos(\mu\theta)d\theta \\ &= \frac{1}{\pi\mu}\left[g_c(\theta)\sin(\mu\theta)\right]_{\theta=0}^{\theta=\pi} - \frac{1}{\pi\mu}\int_0^\pi (g_c)'(\theta)\sin(\mu\theta)d\theta.\end{aligned}$$

The first term vanishes since $\sin(\mu\pi) = \sin(0) = 0$ for $\mu \in \mathbb{Z}$. Then,

$$\begin{aligned}\widehat{g_c}(\mu) &= \frac{1}{\mu\pi}\int_0^\pi g'(\cos\theta)\sin\theta\sin(\mu\theta)d\theta \\ &= \frac{1}{2\mu\pi}\int_0^\pi (g')_c(\theta)\left[\cos((\mu-1)\theta) - \cos((\mu+1)\theta)\right]d\theta \\ &= \frac{1}{2\mu}\left(\widehat{(g')_c}(\mu-1) - \widehat{(g')_c}(\mu+1)\right).\end{aligned}$$

Hence the result follows. □

Definition 5.20. *We define the operator $\mathcal{S}_N : L^2([-\pi,\pi]) \to \mathbb{T}_N$, where $\mathbb{T}_N$ is the space of trigonometric polynomials of degree N (5.35), as follows:*

$$(\mathcal{S}_N\phi)(\theta) = \sum_{\mu=-N}^{\mu=N} \hat{\phi}(\mu)\exp(i\mu\theta), \quad \theta \in [-\pi,\pi], \quad N \geq 0, \tag{5.73}$$

where $\hat{\phi}(\mu)$ is defined in (5.33). Moreover, the operator $\mathcal{S}_N$ applied to an even function ϕ, yields,

$$(\mathcal{S}_N\phi)(\theta) = \hat{\phi}(0) + 2\sum_{\mu=1}^{\mu=N} \hat{\phi}(\mu)\cos(\mu\theta), \tag{5.74}$$

and $\hat{\phi}(\mu)$ may be written in the form (5.68).

The operator $\mathcal{S}_N$ is often referred to as a *truncated* Fourier operator.

Using Lemma 5.19, we prove the following important result for the truncated Fourier operator $\mathcal{S}_N$ applied to f_c.

Lemma 5.21. *For all $0 \leq m \leq N+1$, there exist constants $\sigma_{m,N} > 0$ such that*

$$\|(I - \mathcal{S}_N)f_c\|_{H^m_{per}} \leq \sigma_{m,N}\|(f^{(m)})_c\|_{H^0_{per}}. \tag{5.75}$$

Proof. Expanding the function $(I - \mathcal{S}_N) f_c$ in Fourier series, we obtain:

$$\begin{aligned}(I - \mathcal{S}_N)\, f_c(\theta) &= 2\sum_{\mu=1}^{\infty} \widehat{f}_c(\mu)\cos(\mu\theta) - 2\sum_{\mu=1}^{N} \widehat{f}_c(\mu)\cos(\mu\theta) \\ &= 2\sum_{\mu=N+1}^{\infty} \widehat{f}_c(\mu)\cos(\mu\theta). \end{aligned} \tag{5.76}$$

Let

$$F(\theta) := \sum_{\mu=N+1}^{\infty} \widehat{f}_c(\mu)\cos(\mu\theta).$$

Then, the function F is even and

$$\begin{aligned}\hat{F}(\nu) &= \frac{1}{\pi}\int_0^{\pi} F(\theta)\cos(\nu\theta)d\theta \\ &= \frac{1}{\pi}\int_0^{\pi}\left(\sum_{\mu=N+1}^{\infty} \widehat{f}_c(\mu)\cos(\mu\theta)\right)\cos(\nu\theta)d\theta \\ &= \frac{1}{\pi}\sum_{\mu=N+1}^{\infty} \widehat{f}_c(\mu)\int_0^{\pi}\cos(\nu\theta)\cos(\mu\theta)d\theta.\end{aligned}$$

Since

$$\int_0^{\pi}\cos(\nu\theta)\cos(\mu\theta)d\theta = \begin{cases} \frac{\pi}{2}, \text{ if } \nu = \mu, \\ 0, \text{ otherwise,}\end{cases}$$

we obtain,

$$\hat{F}(\nu) = \begin{cases} 0, \quad \nu = 0, ..., N, \\ \frac{1}{2}\widehat{f}_c(\nu), \text{ if } \nu \geq N+1.\end{cases} \tag{5.77}$$

Then, by (5.69) and (5.76),

$$\begin{aligned}\|(I - \mathcal{S}_N)f_c\|^2_{H^m_{per}} &= 2\|F\|^2_{H^m_{per}} := 2\left(|\hat{F}(0)|^2 + 2\sum_{\mu=1}^{\infty}\mu^{2m}|\hat{F}(\mu)|^2\right) \\ &= \sum_{\mu=N+1}^{\infty}\mu^{2m}\left|\widehat{f}_c(\mu)\right|^2, \qquad \text{using (5.77).}\end{aligned} \tag{5.78}$$

Now, using Lemma 5.19, we obtain:

$$\begin{aligned}
\|(I-\mathcal{S}_N)\, f_c\|^2_{H^m_{per}} &\leq \frac{1}{4}\sum_{\mu\geq N+1}\mu^{2m-2}\left|\widehat{(f')_c}(\mu-1)-\widehat{(f')_c}(\mu+1)\right|^2 \\
&\leq \frac{1}{2}\Bigg[\sum_{\mu\geq N+1}\mu^{2m-2}\left|\widehat{(f')_c}(\mu-1)\right|^2 \\
&\qquad + \sum_{\mu\geq N+1}\mu^{2m-2}\left|\widehat{(f')_c}(\mu+1)\right|^2\Bigg], \qquad (5.79)
\end{aligned}$$

since $(a-b)^2 \leq 2a^2+2b^2$. Then, using the following elementary inequality,

$$\mu+1 \leq \frac{N+1}{N}\mu, \qquad \text{for } \mu \geq N,$$

we obtain from (5.79),

$$\begin{aligned}
2\,\|(I-\mathcal{S}_N)\, f_c\|^2_{H^m_{per}} &\leq \sum_{\mu\geq N}(\mu+1)^{2m-2}\left|\widehat{(f')_c}(\mu)\right|^2 + \sum_{\mu\geq N+2}(\mu-1)^{2m-2}\left|\widehat{(f')_c}(\mu)\right|^2 \\
&\leq \left(\frac{N+1}{N}\right)^{2m-2}\sum_{\mu\geq N}\mu^{2m-2}\left|\widehat{(f')_c}(\mu)\right|^2 + \sum_{\mu\geq N+2}\mu^{2m-2}\left|\widehat{(f')_c}(\mu)\right|^2 \\
&\leq 2\left(\frac{N+1}{N}\right)^{2m-2}\sum_{\mu\geq N}\mu^{2m-2}\left|\widehat{(f')_c}(\mu)\right|^2. \qquad (5.80)
\end{aligned}$$

Analogously to (5.78), we can show that,

$$\|(I-\mathcal{S}_{N-1})(f')_c\|^2_{H^{m-1}_{per}} = \sum_{\mu=N}^{\infty}\mu^{2(m-1)}\left|\widehat{(f')_c}(\mu)\right|^2.$$

Then, dividing both sides of (5.80) by 2, we deduce,

$$\|(I-\mathcal{S}_N)\, f_c\|_{H^m_{per}} \leq \left(\frac{N+1}{N}\right)^{m-1}\left\|(I-\mathcal{S}_{N-1})\,(f')_c\right\|_{H^{m-1}_{per}}. \qquad (5.81)$$

Now, since $m-1 \leq N$, using (5.81) iteratively $m-1$ times, we obtain,

$$\begin{aligned}
\|(I-\mathcal{S}_N)\, f_c\|_{H^m_{per}} &\leq \left(\frac{N+1}{N}\right)^{m-1}\left(\frac{N}{N-1}\right)^{m-2}\cdots\left(\frac{N-m+3}{N-m+2}\right) \\
&\times \|\,(I-\mathcal{S}_{N-m+1})\,(f^{(m-1)})_c\|_{H^1_{per}} \\
&\leq \frac{(N+1)^m(N-m+1)!}{N!}\|\,(I-\mathcal{S}_{N-m+1})\,(f^{(m-1)})_c\|_{H^1_{per}} \qquad (5.82)
\end{aligned}$$

When $m < N+1$, one additional iteration of (5.82) together with the fact that

$$\| \left(I - \mathcal{S}_{N-m}\right) (f^{(m)})_c\|^2_{H^0_{per}} \leq \|(f^{(m)})_c\|^2_{H^0_{per}},$$

yields the desired result,

$$\|(I - \mathcal{S}_N)\, f_c\|_{H^m_{per}} \leq \sigma_{m,N}\ \|(f^{(m)})_c\|_{H^0_{per}},$$

where

$$\sigma_{m,N} := \frac{(N+1)^m (N-m+1)!}{N!}.$$

On the other hand, if $m = N+1$, the right hand side of the equation (5.82), can be bounded using (5.79) as follows,

$$\begin{aligned}
\| \left(I - \mathcal{S}_0\right) (f^{(m-1)})_c\|^2_{H^1_{per}} &\leq \frac{1}{2}\left[\sum_{\mu=0}^{\infty} \left|\widehat{(f^{(m)})}_c(\mu)\right|^2 + \sum_{\mu=2}^{\infty} \left|\widehat{(f^{(m)})}_c(\mu)\right|^2\right] \\
&\leq \left|\widehat{(f^{(m)})}_c(0)\right|^2 + 2\sum_{\mu=1}^{\infty} \left|\widehat{(f^{(m)})}_c(\mu)\right|^2 \\
&=: \|(f^{(m)})_c\|^2_{H^0_{per}}.
\end{aligned}$$

Hence the result (5.75) follows. □

Remark 5.22. *Note that the constants $\sigma_{m,N}$ are bounded as $N \to \infty$. This can be shown by noticing that $\Gamma(N+1) = N!$ and using the following formula (see [3, (6.1.46)]),*

$$\left(\lim_{N\to\infty} \sigma_{m,N} =\right) \qquad \lim_{N\to\infty} N^{m-1} \frac{\Gamma(N-m)}{\Gamma(N+1)} = 1.$$

Finally, using Theorem 5.14 and Lemma 5.21, we can prove Theorem 5.17.

Proof of Theorem 5.17

Proof. We use Theorem 5.14, to determine the error bound of the FCC quadrature in terms of the cosine mapping of the integrand function. Denote $p(x)$ as

$$p(x) = \widehat{f}_c(0) + 2\sum_{\mu=1}^{\mu=N} \widehat{f}_c(\mu) T_\mu(x), \qquad x \in [-1,1], \tag{5.83}$$

where $f_c : [0,\pi] \to \mathbb{R}$ is the cosine mapping, defined in (5.30), of f. Then the cosine mapping of p satisfies,

$$p_c(\theta) = \left(\mathcal{S}_N f_c\right)(\theta). \tag{5.84}$$

Note that the quadrature $Q_{k,N}$ defined in (5.23) is exact for all polynomials of degree up

to N, and so

$$I_k[p] = Q_{k,N}[p].$$

Then, we deduce, for all $m \geq 0$,

$$\begin{aligned}
|I_k[f] - Q_{k,N}[f]| &= |I_k[f-p] - Q_{k,N}[f-p]| \\
&\leq C_{m_0}\left(\frac{1}{k}\right)^s \left(\frac{1}{N}\right)^{m-\rho(s)} \|(f-p)_c\|_{H^m_{per}}, \qquad \text{by Theorem 5.14} \\
&\leq C_{m_0}\left(\frac{1}{k}\right)^s \left(\frac{1}{N}\right)^{m-\rho(s)} \|f_c - (\mathcal{S}_N f_c)\|_{H^m_{per}}, \qquad \text{by (5.84).}
\end{aligned}$$

Then, using Lemma 5.21, for $0 \leq m \leq N+1$,

$$|I_k[f] - Q_{k,N}[f]| \leq \sigma_{m,N}\left(\frac{1}{k}\right)^s \left(\frac{1}{N}\right)^{m-\rho(s)} \|(f^{(m)})_c\|_{H^0_{per}},$$

which completes the proof. □

5.3.3 Extension to the integrals over $[a, b]$

Let us now consider an integral over $[a, b]$:

$$I_k^{[a,b]}[f] := \int_a^b f(x)\exp(ikx)dx.$$

To apply the FCC quadrature, we first need to transform the integral using the following linear change of variables:

$$x = m + ht, \quad t \in [-1,1], \qquad \text{where} \quad m := \frac{b+a}{2}, \quad \text{and} \quad h := \frac{b-a}{2}, \tag{5.85}$$

into an integral over $[-1, 1]$,

$$\begin{aligned}
I_k^{[a,b]}[f] &= h\exp(ikm)\int_{-1}^{1} f(m+ht)\exp(ikht)\,dt \\
&= h\exp(ikm)\int_{-1}^{1} \tilde{f}(t)\exp(i\tilde{k}t)dt \\
&= h\exp(ikm)\, I_{\tilde{k}}^{[-1,1]}[\tilde{f}],
\end{aligned} \tag{5.86}$$

where

$$\tilde{k} = hk, \tag{5.87}$$

and $\tilde{f} : [-1, 1] \to \mathbb{R}$ is defined as in (5.31): $\tilde{f}(t) = f(m + ht)$, $t \in [-1, 1]$.

Now, we can approximate $I_k^{[a,b]}[f]$by applying the FCC quadrature to approximate the

integral $I_{\tilde{k}}^{[-1,1]}[\tilde{f}]$ in (5.86). The result is derived as follows.

Notation 5.23. *We denote $Q_{k,N}^{[a,b]}[f]$ as an approximation to the integral $I_k^{[a,b]}[f]$, defined as:*

$$Q_{k,N}^{[a,b]}[f] := h \exp(ikm)\, Q_{\tilde{k},N}^{[-1,1]}[\tilde{f}], \tag{5.88}$$

The following corollary is an extension of the Theorem 5.14 to the case of integration over $[a,b]$.

Corollary 5.24. ***[Error of the FCC quadrature on $[a,b]$ in terms of the Sobolev norm of the cosine mapping of the integrand function]*** *Let $f \in C^m[a,b]$ be m-times continuously differentiable function. For $s = 0,1,2$ and for all $m \geq m_0 > \max\{1/2, \rho(s)\}$, there exist a constant $C_{m_0} > 0$ such that*

$$\left| I_k^{[a,b]}[f] - Q_{k,N}^{[a,b]}[f] \right| \leq C_{m_0} \left(\frac{b-a}{2}\right)^{1-s} \left(\frac{1}{k}\right)^{s} \left(\frac{1}{N}\right)^{m-\rho(s)} \|(\tilde{f})_c\|_{H^m_{per}}, \tag{5.89}$$

where $\tilde{f}$ is defined in (5.31) and $\rho(s)$ is defined in (5.14).

Proof. From Theorem 5.14, we obtain the following estimate

$$\begin{aligned} \left| I_k^{[a,b]}[f] - Q_{k,N}^{[a,b]}[f] \right| &= h \left| I_{\tilde{k}}^{[-1,1]}[\tilde{f}] - Q_{\tilde{k},N}^{[-1,1]}[\tilde{f}] \right| \\ &\leq C_{m_0} h \frac{1}{\tilde{k}^s} \left(\frac{1}{N}\right)^{m-\rho(s)} \|(\tilde{f})_c\|_{H^m}, \end{aligned} \tag{5.90}$$

where $h := (b-a)/2$. Substituting equation (5.87) into the estimate we obtain the result. □

The following corollary is more useful extension of the Theorem 5.17 to the cases of integration over $[a,b]$.

Since from now on, we will be considering composite Filon-Clenshaw-Curtis rules with N fixed and $b-a \to 0$, we do not any more keep the negative power of N term in (5.89) explicitly in the analysis.

Corollary 5.25. ***[Error of the FCC quadrature on $[a,b]$ in terms of the Chebyshev weighted norm of the derivatives of the integrand function]*** *Let $f \in C^m[a,b]$ be m-times continuously differentiable function. For $s = 0,1,2$ and $\rho(s)$ defined as in Theorem 5.17, and for all m such that $0 \leq m \leq N+1$, there exist constants $\sigma_{m,N} > 0$,*

$$\left| I_k^{[a,b]}[f] - Q_{k,N}^{[a,b]}[f] \right| \leq \left(\frac{b-a}{2}\right)^{m+1-s} \sigma_{m,N} \left(\frac{1}{k}\right)^{s} \left[\int_a^b \frac{\left|f^{(m)}(x)\right|^2}{\sqrt{(b-x)(x-a)}} dx \right]^{1/2}. \tag{5.91}$$

Proof. From the definition of $\tilde{f}$, see (5.31), we deduce,

$$\tilde{f}^{(m)}(t) = h^m f^{(m)}\left(m + ht\right), \tag{5.92}$$

where $h = (b-a)/2$. Thus,

$$\begin{aligned} \int_{-1}^{1} \frac{\left|\tilde{f}^{(m)}(t)\right|^2}{\sqrt{1-t^2}} dt &= h^{2m} \int_{-1}^{1} \frac{\left|\tilde{f}^{(m)}\left(\frac{b+a}{2} + \frac{b-a}{2}t\right)\right|^2}{\sqrt{1-t^2}} dt \\ &= h^{2m} \int_{a}^{b} \frac{\left|f^{(m)}(x)\right|^2}{\sqrt{(b-x)(x-a)}} dx. \end{aligned} \tag{5.93}$$

From Theorem 5.17, we obtain the following estimate

$$\begin{aligned} \left|I_k^{[a,b]}[f] - Q_{k,N}^{[a,b]}[f]\right| &= h\left|I_{\tilde{k}}^{[-1,1]}[\tilde{f}] - Q_{\tilde{k},N}^{[-1,1]}[\tilde{f}]\right| \\ &\leq \sigma_{m,N} h \left(\frac{1}{\tilde{k}}\right)^s \left[\int_{-1}^{1} \frac{\left|\tilde{f}^{(m)}(t)\right|^2}{\sqrt{1-t^2}} dt\right]^{1/2}. \end{aligned} \tag{5.94}$$

Hence substituting (5.93) into (5.94), the result follows. □

5.4 The application of the Filon-Clenshaw-Curtis rule to integrals with algebraic singularities

Let us consider an integral of the form:

$$I_k^{[a,b]}[f] := \int_a^b f(x)\exp(ikx)dx, \tag{5.95}$$

where the integrand function f has an algebraic singularity at a point c in the interval $[a,b]$. Quadrature rules applied to the whole interval $[a,b]$ may perform badly. In order to ensure better convergence rates, in this section the integral is split into a sum of two integrals with the singularity c at the end point of the interval of integration. Then quadratures are applied on a mesh graded toward the singularity c, to the two integrals separately.

Similarly, if the integrand in (5.95) is singular at points s_j, $j = 1, \ldots, J$ in the interval $[a,b]$, then the integral is split into integrals with singularities confined to the end points,

$$I_k^{[a,b]}[f] := I_k^{[a,s_1]}[f] + \sum_{j=2}^{J-1} I_k^{[s_{j+1},s_j]}[f] + I_k^{[s_J,b]}[f]. \tag{5.96}$$

In this section, without the loss of generality, we consider integrals of the form,

$$I_k^{[0,1]}[f] := \int_0^1 f(x)\exp(ikx)dx, \tag{5.97}$$

where the function f has an algebraic singularity at $x = 0$. With an elementary linear transformation, the integrals in (5.96) can each be transformed into the integral in (5.97) with the singularity confined to the origin.

The technique for the computation of (5.97) consist of applying Filon-Clenshaw-Curtis quadrature on the subintervals of a mesh which is graded towards the origin.

We define the space of all functions that have an algebraic singularity of order β as follows:

Definition 5.26. *Let $m \in \mathbb{N}$, $0 < \beta \leq 1$. Denote $C_\beta^m[0,1]$ as the space of all functions $v : [0,1] \to \mathbb{C}$ that have m continuous derivatives on $(0,1)$ and for which $||v||_{m,\beta} < \infty$ where*

$$||v||_{m,\beta,[0,1]} = \max\left\{ \sup_{x\in[0,1]} |v(x)|, \sup_{x\in[0,1]} \left| x^{j-\beta} D^j v(x) \right|, \quad j = 1, ..., m. \right\}. \tag{5.98}$$

To approximate such functions accurately, we introduce a mesh graded towards the origin.

Definition 5.27. $\Pi_{M,q}$ *is a set of points on* $[0,1]$ *defined as*

$$\Pi_{M,q} := \left\{ x_j : \ j = 0, 1, \ldots, M, \quad x_j := \left(\frac{j}{M}\right)^q \right\}. \tag{5.99}$$

We write the integral $I_k^{[0,1]}[f]$ as a sum of integrals over intervals of the graded mesh $\Pi_{M,q}$. We approximate each integral using FCC quadrature as follows:

- In the **first subinterval**, for convenience of the analysis, we rewrite the integral as a sum of a known constant and a new integral with the integrand function that vanishes at 0:

  $$\int_0^{x_1} f(x)dx = x_1 f(0) + \int_0^{x_1} (f(x) - f(0))\, dx = x_1 f(0) + \int_0^{x_1} \tilde{f}(x)dx.$$

  The latter integral is then approximated using FCC quadrature.

- In the **remaining subintervals**, we apply FCC quadrature directly to the integral.

Then, the total error of the approximation is bounded as a sum of errors on each subinterval:

$$\begin{aligned} E(k, M, q, N, f) \quad &:= \quad \left| I_k^{[x_0,x_1]}[\tilde{f}] - Q_{k,N}^{[x_0,x_1]}[\tilde{f}] \right| + \sum_{j=0}^{M-1} \left| I_k^{[x_j,x_{j+1}]}[f] - Q_{k,N}^{[x_j,x_{j+1}]}[f] \right| \\ &\leq \quad |\tilde{e}_0| + \sum_{j=1}^{M-1} |e_j|, \end{aligned} \tag{5.100}$$

where e_j is the error on the j-th subinterval.

To analyse the total error of the composite FCC quadrature, we estimate the individual error of the approximation on each subinterval. We find these estimates using Theorem 5.17 for all subintervals excluding the first subinterval containing the singularity. In the first subinterval, we estimate the error using Corollary 5.24.

Error estimates on each subinterval $[x_j, x_{j+1}]$, $j \neq 0$, are presented below in Theorem 5.28 for the j-th subinterval. In Theorem 5.31 the error estimate for the first subinterval, $[0, x_1]$ is presented. To prove Theorem 5.31, we require Lemma 5.29 and Theorem 5.30.

Lemma 5.28 (**Error in the j-th subinterval**). *For* $0 \leq m \leq N+1$, *and all* $j = 1, ..., M-1$, *and for* $s = 0, 1, 2$, *there exist* $\sigma_m > 0$ *such that*

$$|e_j| \leq \sigma_m \left(\frac{1}{k}\right)^s h_j^{m+1-s} \max_{x \in [x_j, x_{j+1}]} |f^{(m)}(x)|, \tag{5.101}$$

where $h_j = x_{j+1} - x_j$.

Proof. The proof is a consequence of Corollary 5.25 and the following inequality,

$$\int_a^b \frac{\left|f^{(m)}(x)\right|^2}{\sqrt{(b-x)(x-a)}}dx \leq C \max_{x\in[a,b]} \left|f^{(m)}(x)\right|^2.$$

□

Using Definition 5.8, we determine the cosine mapping of the function $f \in C^m_\beta[0,\chi]$, $\chi \in (0,1)$.

In the following lemma, we determine which Sobolev space the cosine mapping of the function x^β belongs to.

Lemma 5.29. *Let g be a function defined as*

$$g(x) = x^\beta, \quad x \in [0,\chi],$$

where $\chi \in (0,1)$ and $\beta \in (0,1)$. Following (5.30) and (5.31), the cosine mapping of g is

$$\tilde{g}_c(\theta) = g\left(\frac{\chi}{2}(1+\cos\theta)\right) = g\left(\chi\cos^2\frac{\theta}{2}\right), \qquad \theta \in [0,\pi]. \tag{5.102}$$

Then, for any $\varepsilon > 0$,

$$\tilde{g}_c \in H^{\frac{1}{2}+2\beta-\varepsilon}_{per}. \tag{5.103}$$

Furthermore, there exist a constant $C > 0$, such that

$$\|\tilde{g}_c\|_{H^{\frac{1}{2}+2\beta-\varepsilon}_{per}} \leq C\chi^\beta. \tag{5.104}$$

Proof. We want to show that the Sobolev norm of $\tilde{g}_c$ in $H^{\frac{1}{2}+2\beta-\varepsilon}_{per}$ is bounded. For this, we require the estimates of the Fourier coefficients of the cosine transform of the function $\tilde{g}_c$. We will use the following notation for the Fourier coefficients of $\tilde{g}_c$,

$$(\tilde{g}_c)\hat{}\,(m) = \text{ Fourier Transform of } \tilde{g}_c.$$

Following (5.69), we write,

$$\|\tilde{g}_c\|^2_{H^s_{per}} := |(\tilde{g}_c)\hat{}\,(0)|^2 + 2\sum_{m=1}^{\infty} m^{2s}\,|(\tilde{g}_c)\hat{}\,(m)|^2, \tag{5.105}$$

where the Fourier coefficients are defined as

$$(\tilde{g}_c)\hat{}\,(m) := \frac{1}{\pi}\int_0^\pi \tilde{g}_c(\theta)\cos(m\theta)d\theta.$$

Now, we estimate the Fourier coefficients for $\tilde{g}_c$ by firstly integrating by parts,

$$\begin{aligned}
(\tilde{g}_c)\widehat{\ }(m) &:= \frac{1}{\pi}\int_0^\pi \tilde{g}_c(\theta)\cos(m\theta)d\theta, \\
&= \frac{1}{m\pi}\tilde{g}_c(\theta)\,\sin(m\theta)|_{\theta=0}^{\theta=\pi} - \frac{1}{m\pi}\int_0^\pi (\tilde{g}_c)'(\theta)\sin(m\theta)d\theta \\
&= \frac{\chi}{m\pi}\int_0^\pi g'\left(\chi\cos^2\left(\frac{\theta}{2}\right)\right)\cos\left(\frac{\theta}{2}\right)\sin\left(\frac{\theta}{2}\right)\sin(m\theta)d\theta.
\end{aligned}$$

Therefore,

$$\begin{aligned}
(\tilde{g}_c)\widehat{\ }(m) &= \frac{\beta\chi}{m\pi}\int_0^\pi \left(\chi\cos^2\left(\frac{\theta}{2}\right)\right)^{\beta-1}\cos\left(\frac{\theta}{2}\right)\sin\left(\frac{\theta}{2}\right)\sin(m\theta)d\theta \\
&= \frac{\beta\chi^\beta}{m\pi}\int_0^\pi \cos^{2\beta-1}\left(\frac{\theta}{2}\right)\sin\left(\frac{\theta}{2}\right)\sin(m\theta)d\theta \\
&= \frac{2\beta\chi^\beta}{m\pi}\int_0^{\frac{\pi}{2}} \cos^{2\beta-1}(t)\sin(t)\sin(2mt)dt.
\end{aligned}$$

The latter integral can be found in [53, (3.633.1)] ,

$$|(\tilde{g}_c)\widehat{\ }(m)| = C\pi\frac{\chi^\beta}{m\pi}\frac{m}{B\left(\frac{2\beta+2m}{2}+1, \frac{2\beta-2m}{2}+1\right)}, \tag{5.106}$$

where the constant C is independent of m and B is the Beta function defined, for example in [53, (8.384.1)] as follows:

$$\mathcal{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \tag{5.107}$$

where Γ is the Gamma function [53, Section 8.31]. For large values of its variable, Γ can be approximated using Stirling's formula [3, (6.1.37)]. This leads to the following property of the Beta functions: for $x, y \to \infty$

$$B(x,y) \cong \sqrt{2\pi}\frac{x^{x-\frac{1}{2}}y^{y-\frac{1}{2}}}{(x+y)^{x+y-\frac{1}{2}}}. \tag{5.108}$$

Then, for $m \to \infty$,

$$\begin{aligned}
|B(\beta+m+1, \beta-m+1)| &\leq c_\beta\,|\beta+m+1|^{\beta+m+\frac{1}{2}}\,|\beta-m+1|^{\beta-m+\frac{1}{2}} \\
&\leq C_\beta m^{2\beta+1},
\end{aligned}$$

where c_β and C_β are constants independent of m. Finally, for $m \to \infty$, we obtain from (5.106),

$$|(\tilde{g}_c)\widehat{\ }(m)| \leq C\chi^\beta m^{-2\beta-1}, \tag{5.109}$$

where C is a constant independent of m. Substituting this estimate into the definition of the Sobolev norm (5.105), we obtain

$$\begin{aligned} \|\tilde{g}_c\|^2_{H^s_{per}} &:= |\,(\tilde{g}_c)\hat{}\,(0)|^2 + 2\sum_{m=1}^{\infty} m^{2s}\,|(\tilde{g}_c)\hat{}\,(m)|^2 \\ &\leq |\,(\tilde{g}_c)\hat{}\,(0)|^2 + C\chi^{2\beta}\sum_{m=1}^{\infty} m^{2s} m^{-4\beta-2}. \end{aligned} \tag{5.110}$$

The sum on the right hand side is convergent if and only if $2s - 4\beta - 2 < -1$. Thus, $g_c \in H^{2\beta+\frac{1}{2}-\varepsilon}$.

Furthermore, observe that

$$\begin{aligned} (\tilde{g}_c)\hat{}\,(0) &:= \frac{1}{\pi}\int_0^{\pi}\left(\chi\cos^2\frac{\theta}{2}\right)^{\beta} d\theta \\ &= \frac{\chi^{\beta}}{\pi}\int_0^{\pi}\left(\cos^{\beta}\frac{\theta}{2}\right)^{2} d\theta \\ &\leq \frac{\chi^{\beta}}{\pi}\int_0^{\pi}(1)^2\, d\theta \\ &= \chi^{\beta}, \end{aligned} \tag{5.111}$$

Then, by substituting (5.111) into (5.110), we obtain (5.104). □

Corollary 5.30. *For $f \in C^3_{\beta}[0,\chi]$, where $\chi < 1$, and $\beta \in (0,1)$, the following holds:*

$$\tilde{f}_c \in H^{\frac{1}{2}+2\beta-\varepsilon}_{per}. \tag{5.112}$$

If, in addition, f vanishes at 0, then,

$$\left\|\tilde{f}_c\right\|_{H^{\frac{1}{2}+2\beta-\varepsilon}_{per}} \leq C\chi^{\beta}\|f\|_{1,\beta,[0,\chi]}. \tag{5.113}$$

Proof. We will prove the corollary for the case when $\beta \leq 1/4$. For $\beta \geq 1/4$ the proof follows similarly. We aim to show $\tilde{f}_c \in H^s$, where $s = 1/2 + 2\beta - \varepsilon < 1$. Following (5.69), we write,

$$\left\|\tilde{f}_c\right\|^2_{H^s_{per}} := \left|\left(\tilde{f}_c\right)\hat{}\,(0)\right|^2 + 2\sum_{m=1}^{\infty} m^{2s}\left|\left(\tilde{f}_c\right)\hat{}\,(m)\right|^2,$$

is equivalent to the Sobolev-Slobodetskiy norm, $\|\cdot\|_s$, (see [66, Theorem 8.5]) for $0 < s < 1$:

$$\left\|\tilde{f}_c\right\|^2_s := \int_0^{\pi}|\tilde{f}_c(t)|^2 dt + \int_0^{\pi}\int_0^{\pi}\frac{|\tilde{f}_c(t) - \tilde{f}_c(\tau)|^2}{\left|\sin\frac{t-\tau}{2}\right|^{2s+1}} d\tau dt. \tag{5.114}$$

Note that $\tilde{g}'(t) \geq 0$ for all $t \in [-1,1]$, where $g(x) = x^\beta$. Then, for $t, \tau \in [0,\pi]$,

$$\begin{aligned}
\left|\tilde{f}_c(t) - \tilde{f}_c(\tau)\right| &= \left|\int_\tau^t \left(\tilde{f}_c\right)'(\theta)d\theta\right| \\
&\leq \int_\tau^t \left|\tilde{f}'(\cos\theta)\sin(\theta)\right| d\theta \\
&\leq \int_\tau^t \left|\tilde{f}'(\cos\theta)\right| \sin(\theta)d\theta \qquad \text{since} \quad \sin(\theta) \geq 0, \\
&\leq \|f\|_{1,\beta,[0,\chi]} \int_\tau^t \tilde{g}'(\cos\theta)\sin(\theta)d\theta = \|f\|_{1,\beta,[0,\chi]} \int_\tau^t (\tilde{g}_c)'(\theta)d\theta \\
&= \|f\|_{1,\beta,[0,\chi]} \left|\tilde{g}_c(t) - \tilde{g}_c(\tau)\right|. \qquad (5.115)
\end{aligned}$$

Furthermore, we observe the following, using (5.115) and the fact that $\tilde{g}_c(\pi) = \tilde{g}(-1) = g(0) = 0$,

$$\begin{aligned}
\int_0^\pi |\tilde{f}_c(t)|^2 dt &\leq \int_0^\pi |\tilde{f}_c(\pi)|^2 dt + \int_0^\pi |\tilde{f}_c(t) - \tilde{f}_c(\pi)|^2 dt \\
&\leq \pi|\tilde{f}_c(\pi)|^2 + ||f||_{1,\beta,[0,\chi]} \int_0^\pi |\tilde{g}_c(t)|^2 dt. \qquad (5.116)
\end{aligned}$$

Therefore, using (5.116) to bound the first term in (5.114) and using (5.115) to bound the second term, we obtain

$$\begin{aligned}
\left\|\tilde{f}_c\right\|_s^2 &\leq \pi|\tilde{f}_c(\pi)|^2 + ||f||_{1,\beta,[0,\chi]} \int_0^\pi |\tilde{g}_c(t)|^2 dt \\
&+ ||f||_{1,\beta,[0,\chi]} \int_0^{2\pi} \int_0^{2\pi} \frac{|\tilde{g}_c(t) - \tilde{g}_c(\tau)|^2}{\left|\sin\frac{t-\tau}{2}\right|^{2s+1}} d\tau dt \\
&= \pi|\tilde{f}_c(\pi)|^2 + ||f||_{1,\beta,[0,\chi]} \left(\int_0^\pi |\tilde{g}_c(t)|^2 dt + \int_0^\pi \int_0^\pi \frac{|\tilde{g}_c(t) - \tilde{g}_c(\tau)|^2}{\left|\sin\frac{t-\tau}{2}\right|^{2s+1}} d\tau dt\right) \\
&= \pi|\tilde{f}_c(\pi)|^2 + ||f||_{1,\beta,[0,\chi]} \|\tilde{g}_c\|^2_{H^s_{per}}. \qquad (5.117)
\end{aligned}$$

Thus (5.112) follows from Lemma 5.29.

Note that, if $f(0) = 0$, then $\tilde{f}_c(\pi) = 0$ and the first term in (5.117) vanishes. Then, the bound (5.113) follows from Lemma 5.29.

When $\beta > 1/4$, the proof follows analogously by using [66, Corollary 8.6]: for $s = n + q$, $0 < q < 1$, $n \in \mathbb{N}$,

$$\left\|\tilde{f}_c\right\|^2_{H^s_{per}} = \|\tilde{f}_c\|^2_{H^0_{per}} + \left\|\left(\tilde{f}_c\right)^{(n)}\right\|_q^2, \qquad (5.118)$$

with $n = 1$, when $1/4 < \beta < 3/4$ and $n = 2$, when $\beta > 3/4$. □

In the following proposition, we obtain an estimate for $\tilde{e}_0$ using Theorem 5.14.

Theorem 5.31 (**Error in the first subinterval**). *The error of the FCC approximation of* $I_k^{[0,x_1]}\left[\tilde{f}\right]$ *is bounded as follows: for* $s = 0, 1, 2$,

$$|\tilde{e}_0| \leq C \left(\frac{1}{k}\right)^s \left(\frac{1}{M}\right)^{q(1+\beta-s)} \|f\|_{1,\beta,[0,x_1]}, \tag{5.119}$$

where C *is independent of* k, M *and* f.

Proof. The estimate for $\tilde{e}_0$ can be obtained from Corollary 5.24 (extension of the Theorem 5.14 to $[a, b]$). Since $\tilde{f}(0) = 0$, by (5.113)

$$|\tilde{e}_0| \leq C x_1^{\beta} h_0^{1-s} \left(\frac{1}{k}\right)^s \left(\frac{1}{N}\right)^{1/2+2\beta-s-\varepsilon} \left\|\tilde{f}_c\right\|_{H_{per}^{2\beta+\frac{1}{2}-\varepsilon}}, \tag{5.120}$$

where $h_0 = x_1 - 0 = (1/M)^q$. Thus the result follows. □

Finally, combining the results of Theorem 5.30 and Theorem 5.28, we deduce the composite error of the FCC quadrature applied on graded meshes.

5.4.1 Error estimate of the composite Filon-Clenshaw-Curtis quadrature

In this section we state and prove the main result of this chapter: the error, $E(k, M, q, N, f)$, of the Filon-Clenshaw-Curtis quadrature applied on graded meshes for the computation of integrals with algebraic singularities. Recall, the error is bounded as a sum of errors on each subinterval:

$$\begin{aligned} E(k, M, q, N, f) &:= \left| I_k^{[x_0,x_1]}[\tilde{f}] - Q_{k,N}^{[x_0,x_1]}[\tilde{f}] \right| + \sum_{j=0}^{M-1} \left| I_k^{[x_j,x_{j+1}]}[f] - Q_{k,N}^{[x_j,x_{j+1}]}[f] \right| \\ &\leq |\tilde{e}_0| + \sum_{j=1}^{M-1} |e_j|, \end{aligned} \tag{5.121}$$

where e_j is the error on the j-th subinterval.

Theorem 5.32. ***[Error of the composite Filon-Clenshaw-Curtis quadrature for singular integrals]*** *Let $f \in C_\beta^{N+1}[0,1]$, $\beta \in (0,1)$ and let $\Pi_{M,q}$ be as in (5.99). Choose*

$$q \geq \frac{N+1-s}{1+\beta-s}, \tag{5.122}$$

then for $s = 0, 1$,

$$E(k, M, q, N, f) \leq C \left(\frac{1}{k}\right)^s \left(\frac{1}{M}\right)^{N+1-s} \|f\|_{N+1,\beta,[0,1]}, \tag{5.123}$$

where the constant C depends on N, β and s.

Proof. In this proof, we denote C as a generic constant that may depend on N, β, q and s, but not on M and k. We seek the estimates on the sum of individual errors e_j, $j = 0, \ldots, M-1$, see (5.121). The error on the first subinterval is bounded as in Theorem 5.31 and on the j-th subinterval as in Lemma 5.28.

We begin by estimating the sum in (5.121) using Lemma 5.28,

$$\begin{aligned} \sum_{j=1}^{M-1} |e_j| &\leq C \left(\frac{1}{k}\right)^s \sum_{j=1}^{M-1} h_j^{N+2-s} \max_{x\in[x_j,x_{j+1}]} \left| f^{(N+1)}(x) \right| \\ &\leq C \left(\frac{1}{k}\right)^s \|f\|_{N+1,\beta,[x_1,1]} \sum_{j=1}^{M-1} h_j^{N+2-s} x_j^{\beta-N-1}. \end{aligned} \tag{5.124}$$

In (5.124) the constant σ_m appearing in Lemma 5.28 has been absorbed in the constant C.

For each h_j, $j = 1, \ldots, M-1$, h_j can be bounded using the Mean Value Theorem,

$$x_j^{(1-1/q)} \frac{q}{M} \leq h_j \leq x_{j+1}^{(1-1/q)} \frac{q}{M}. \tag{5.125}$$

Then, using the second inequality in (5.125), the sum of errors in (5.124) is bounded as follows,

$$\sum_{j=1}^{M-1} |e_j| \leq C \left(\frac{1}{k}\right)^s \|f\|_{N+1,\beta,[x_1,1]} \frac{1}{M^{N+2-s}} \sum_{j=1}^{M-1} x_j^{(1-1/q)(N+2-s)+\beta-N-1}, \tag{5.126}$$

where the constant C depends on q. We can find an upper bound for the sum on the right hand side of (5.126) as follows, for $q\alpha > -1$,

$$\sum_{j=1}^{M-1} x_j^{\alpha} = \sum_{j=1}^{M-1} \left(\frac{j}{M}\right)^{q\alpha} \leq CM. \tag{5.127}$$

Then, taking $\alpha = (1-1/q)(N+2-s) + \beta - N - 1$, for $q\alpha > -1$,

$$\sum_{j=1}^{M-1} x_j^{(1-1/q)(N+2-s)+\beta-N-1} \leq CM,$$

and

$$\sum_{j=1}^{M-1} |e_j| \leq C \left(\frac{1}{k}\right)^s \left(\frac{1}{M}\right)^{N+1-s} \|f\|_{N+1,\beta,[x_1,1]}. \tag{5.128}$$

The condition $q\alpha > -1$ is equivalent to (5.122). On the other hand, using Lemma 5.31, we deduce that the first term in (5.121) is bounded as follows,

$$\begin{aligned} |e_0| &\leq C \left(\frac{1}{k}\right)^s \left(\frac{1}{M}\right)^{q(1+\beta-s)} \|f\|_{N+1,\beta,[0,x_1]} \\ &\leq C \left(\frac{1}{k}\right)^s \left(\frac{1}{M}\right)^{N+1-s} \|f\|_{N+1,\beta,[0,x_1]}. \end{aligned} \tag{5.129}$$

Then, the total error, including the first subinterval, is bounded as in (5.123) This concludes the proof. □

5.5 Numerical examples

In this section we carry out numerical experiments where we compute integrals with algebraic singularities using the composite Filon-Clenshaw-Curtis quadrature to verify the theoretical error estimates presented in Section 5.4.

Example 1.

In this example, we investigate the error of the (non-composite) Filon-Clenshaw-Curtis quadrature *without* grading applied to a singular integral presented below: for $\beta > 0$,

$$I_k^{[-1,1]}[f_\beta] = \int_{-1}^{1} f_\beta(x)\exp(ikx)dx, \qquad \text{where} \quad f_\beta(x) := \left(\frac{1+x}{2}\right)^\beta. \tag{5.130}$$

From Corollary 5.30, we know that the function f_β in (5.130) satisfies,

$$(f_\beta)_c \in H_{per}^{\frac{1}{2}+2\beta-\varepsilon}.$$

The purpose of this experiment is to examine the rate of decay of the error as $k \to \infty$ for different choices of β for fixed N.

verify that Theorem 5.14 is sharp with respect to parameter β.

	$\beta = 1/8$		$\beta = 1/4$		$\beta = 1/2$		$\beta = 7/8$		$\beta = 3/2$	
k_i	$E_{k_i}(24)$	$r(k_i)$	$E_{k_i}(24)$	$r(k_i)$	$E_{k_i}(24)$	$r(k_i)$	$E_{k_i}(24)$	$r(k_i)$	$E_{k_i}(24)$	$r(k_i)$
100	2.8e-003		1.1e-003		1.7e-004		4.2e-006		2.4e-007	
200	1.8e-003	0.62	6.9e-004	0.69	9.5e-005	0.81	2.1e-006	0.99	1.0e-007	1.23
400	9.2e-004	0.96	3.4e-004	1.02	4.3e-005	1.13	8.8e-007	1.26	3.8e-008	1.45
800	4.4e-004	1.05	1.6e-004	1.13	1.8e-005	1.25	3.3e-007	1.40	1.3e-008	1.60
1600	2.1e-004	1.09	6.9e-005	1.17	7.3e-006	1.32	1.2e-007	1.50	3.9e-009	1.70
3200	9.7e-005	1.11	3.0e-005	1.20	2.8e-006	1.38	3.9e-008	1.58	1.1e-009	1.79
6400	4.5e-005	1.11	1.3e-005	1.22	1.1e-006	1.41	1.3e-008	1.63	3.1e-010	1.85
12800	2.1e-005	1.12	5.5e-006	1.23	3.9e-007	1.44	4.0e-009	1.67	8.4e-011	1.89
25600	9.5e-006	1.12	2.3e-006	1.24	1.4e-007	1.46	1.2e-009	1.71	2.2e-011	1.93
51200	4.3e-006	1.12	9.8e-007	1.24	5.1e-008	1.47	3.7e-010	1.73	5.7e-012	1.94

Table 5.1: *The table illustrates the errors $E_k(N)$ defined in (5.131) and the ratios of errors, $r(k)$, defined in (5.132).*

In Table 5.1 we tabulate the errors: $E_k(N)$ defined as

$$E_k(N) = \left| I_k^{[-1,1]}[f_\beta] - Q_{k,N}^{[-1,1]}[f_\beta] \right|, \tag{5.131}$$

for $\beta \in \{1/4, 1/2, 7/8, 3/2\}$ and for $k \in \{100 \times 2^i,\ i = 0, ..., 9\}$, and fixed $N = 24$. That

is, the integral $I_k^{[-1,1]}[f_\beta]$ is approximated with 25-point Filon-Clenshaw-Curtis rule on $[-1,1]$.

In the table, we also display the ratios,

$$r(k) := \log_2 \frac{E_{k/2}(N)}{E_k(N)}, \tag{5.132}$$

that indicate the rate, s, at which the error $E_k(N)$ decays with respect to k, i.e. $E_k(N) \backsim O(k^{-s})$ and

$$\begin{aligned} r(k) &= \log_2 \left(\left(\frac{k}{2} \right)^{-s} k^s \right) \\ &= s. \end{aligned}$$

From the Theorem 5.14, we expect to see $O(k^{-1})$ convergence for $\beta > 1/4$ and $O(k^{-2})$ for $\beta > 3/2$.

We clearly observe $O(k^{-1})$ convergence for $\beta = 1/8$ and $O(k^{-5/4})$ convergence for $\beta = 1/4$ which is better than theory predicts. On the other hand, for $\beta = 1/2$, the error converges with order close to $O(k^{-3/2})$, while for $\beta = 7/8$ the convergence rate is $O(k^{-7/4})$ and for $\beta = 3/2$ the convergence rate is $O(k^{-2})$ indicating the sharpness of the estimate in Theorem 5.14.

The results of this example are consistent with a similar experiment conducted in [41].

Example 2.

In this example, we again consider the integral (5.130). Here, we compute $I_k^{[-1,1]}[f_\beta]$ using the *composite* Filon-Clenshaw-Curtis quadrature. The purpose of this experiment is to verify the theoretical predictions given by Theorem 5.32 for fixed M as $k \to \infty$ and as the parameter β changes. Furthermore, the example demonstrates the advantages of using the composite quadrature over the non-composite FCC applied to singular integrals.

We compute the integral $I_k^{[-1,1]}[f_\beta]$ by applying FCC quadrature on the subintervals of the graded mesh on $[-1,1]$:

$$\Pi_{M,q} := \left\{ x_j : \ j = 0, 1, \ldots, M, \quad x_j := 2 \left(\frac{j}{M} \right)^q - 1 \right\}, \tag{5.133}$$

with fixed

$$M = 6 \quad \text{and} \quad q = 12. \tag{5.134}$$

	$\beta = 1/8$		$\beta = 1/4$		$\beta = 1/2$		$\beta = 7/8$		$\beta = 3/2$	
k_i	$E_{k_i}(6,12)$	$r(k_i)$	$E_{k_i}(6,12)$	$r(k_i)$	$E_{k_i}(6,12)$	$r(k_i)$	$E_{k_i}(6,12)$	$r(k_i)$	$E_{k_i}(6,12)$	$r(k_i)$
200	1.5e-004		1.1e-004		2.3e-005		7.2e-007		1.9e-007	
400	8.2e-005	0.90	5.6e-005	0.94	1.2e-005	1.02	2.9e-007	1.33	4.7e-008	2.05
800	3.4e-005	1.29	2.3e-005	1.30	4.6e-006	1.32	1.2e-007	1.24	1.5e-008	1.69
1600	1.2e-005	1.54	7.8e-006	1.54	1.6e-006	1.57	4.0e-008	1.59	3.9e-009	1.91
3200	3.6e-006	1.67	2.4e-006	1.68	4.8e-007	1.71	1.2e-008	1.75	1.1e-009	1.79
6400	1.4e-006	1.39	7.9e-007	1.63	1.4e-007	1.80	3.2e-009	1.89	2.8e-010	1.99
12800	1.9e-007	2.84	1.7e-007	2.22	3.6e-008	1.93	8.5e-010	1.93	7.1e-011	2.00

Table 5.2: *The table illustrates the errors $E_k(M,q)$ defined in (5.135) and the ratios of errors, $r(k)$, defined in (5.136).*

In Table 5.2 we tabulate the errors: $E_k(M,q)$ defined as

$$E_k(M,q) = \left| I_k^{[-1,1]}[f_\beta] - \sum_{j=0}^{M-1} Q_{N,k}^{[x_j,x_{j+1}]}[f_\beta] \right|, \quad \text{with} \quad N = 4, \tag{5.135}$$

for $\beta \in \{1/4, 1/2, 7/8, 3/2\}$ and for $k \in \{100 \times 2^i,\ i = 0, ..., 9\}$.

To verify the convergence rates we also display the ratios,

$$r(k) := \log_2 \frac{E_{k/2}(M,q)}{E_k(M,q)}, \tag{5.136}$$

which indicate the rate s at which the error $E_k(M,q)$ decays with respect to k, i.e. $E_k(M,q) \backsim O(k^{-s})$.

Note that with our choice of parameters M and N, the total number of function evaluations in the interval $[-1,1]$ is equal to 25 which is the same number as in the previous example. Comparing Table 5.2 and Table 5.1, we see that for fixed k, the errors of the composite FCC quadrature are smaller that non-composite FCC (i.e. FCC applied to the whole interval $[-1,1]$) although the same number of function evaluations were used in both experiments.

With our choice of parameters q and N (see condition (5.122) in Theorem 5.32), we expect the errors to be decaying with respect to k as follows for $\beta = 1/8$ and $\beta = 1/4$, the errors should decay at the following rate $E_k(M,q) = O(k^0)$; while for $\beta = 1/2$ and $\beta = 7/8$, we expect $E_k(M,q) = O(k^{-1})$.

In Table 5.2, we observe even better convergence results: For $\beta = 1/8$ and $\beta = 1/4$, $E_k(M,q)$ decays with k at a rate between $O(k^{-1})$ and $O(k^{-2})$. On the other hand, for $\beta = 1/2$ and $\beta = 7/8$, $E_k(M,q)$ decays with k at a rate close to $O(k^{-2})$.

	FCC quadrature $E_{N_i}(1,1)$			composite FCC $E_{N_i}(6,12)$	
N	$k=400$	$k=1600$	$M\times N$	$k=400$	$k=1600$
24	9.2e-004	4.5e-005	6×4	1.5e-005	1.0e-006
48	5.9e-004	4.4e-005	6×8	8.4e-007	2.3e-007
96	1.8e-004	4.2e-005	6×16	1.5e-008	1.5e-008
192	9.7e-005	2.6e-005	6×32	5.5e-012	3.3e-009

Table 5.3: *The table illustrates the errors $E_N(M,q)$ defined in (5.137).*

In Table 5.3, we display the errors $E_{N_i}(M,q)$, defined as,

$$E_N(M,q) = \left| I_k^{[-1,1]}[f_\beta] - \sum_{j=0}^{M-1} Q_{N,k}^{[x_j,x_{j+1}]}[f_\beta] \right|, \tag{5.137}$$

with x_j, $j=0,...,M$ defined in (5.133). The values of M and q are both equal to 1 when non-composite FCC is applied, and are as given in (5.134) for the composite version.

The purpose of this experiment is to compare the performance of the composite FCC rule with the non-composite FCC quadrature when the number of quadrature points, N, is increased. For the non-composite rule, we display the errors of the approximation with $N_i = \{24\times 2^i, i=0,1,2,3\}$. For the composite rule, we fix parameters $q=12$, $M=6$ and take $N_i = \{4\times 2^i, i=0,1,2,3\}$. The total number of function evaluations in both cases is the same.

From the Table 5.3, we observe very slow convergence with respect to N for the case of non-composite FCC. For the composite FCC, on the other hand, we observe convergence of order $O(N^{-1})$ at worst.

Example 3.

In this experiment we compute the integral

$$I_k\left[\sqrt{x}\right] = \int_0^1 \sqrt{x}\exp(ikx)dx,$$

for fixed values of N and q and varying M and k. Here, we can think of the parameter β featuring in the previous examples, as $1/2$.

In Table 5.4, we display the error $E_k(M,q,N)$ defined as,

$$E_k(M,q,N) := \left| I_k^{[-1,1]}\left[\sqrt{x}\right] - \sum_{j=0}^{M-1} Q_{N,k}^{[x_j,x_{j+1}]}\left[\sqrt{x}\right] \right|, \tag{5.138}$$

	k	$M=4$	$M=8$	$M=16$	$M=32$
$N=8, q=6$					
	10	2.4e-006	4.6e-009	9.1e-012	4.7e-013
	100	4.0e-006	5.3e-009	1.4e-011	2.1e-014
	1000	1.8e-006	5.8e-009	1.7e-011	1.1e-014
	10000	2.0e-006	4.6e-009	1.3e-011	2.7e-014
$N=10, q=8$					
	10	7.7e-008	3.7e-011	1.9e-014	4.9e-016
	100	2.2e-007	1.8e-009	5.6e-013	2.0e-016
	1000	1.5e-007	1.2e-009	3.6e-013	2.1e-016
	10000	1.2e-007	3.5e-010	3.3e-013	1.4e-016

Table 5.4: *The table displays the errors $E_k(M,q,N)$ defined in (5.138) for fixed N and q and varying M and k.*

where we choose q and N so that the condition (5.122) holds with $s=0$:

$$q \geq \frac{N+1}{1+\beta}.$$

The error $E_k(M,q,N)$ is equal to $E(k,M,q,N,f)$ defined in (5.121).

Therefore, from Theorem 5.32, we expect no convergence with respect to k and order $O(M^{-N-1})$ convergence with respect to growing M. We show the values for q and N in the first column of Table 5.4. We observe no convergence with respect to k as expected.

In Table 5.5, we display the ratios

$$r_k(M,q,N) := \log_2 \frac{E_k(M,q,N)}{E_k(2M,q,N)}, \tag{5.139}$$

which indicate the rate, s, at which the error $E_k(M,q,N)$ decays with respect to M, i.e. $E_k(M,q,N) \backsim O(M^{-s})$.

We observe that $r_k(M,q,N)$ tends to the value of $N+1$ as M increases which indicates the sharpness of the error bound in Theorem 5.32 with respect to M.

	$r_k(4,q,N)$	$r_k(8,q,N)$	$r_k(16,q,N)$	expected ratio
N = 6, q = 5, k = 1000	7.7	8.1	7.3	7
N = 8, q = 6, k = 1000	8.2	8.0	10.6	9
N = 10, q = 8, k = 1000	8.4	10.1	11.2	11

Table 5.5: *The table displays the ratios $r_k(M,q,N)$ of errors defined in (5.139) for $k=1000$ to determine the convergence rate of the composite FCC with respect to M.*

In Table 5.6, we display the error $E_k(M,q,N)$ with N and q satisfying the inequality (5.122) with $s=1$:

$$q \geq \frac{N}{\beta}.$$

Therefore, from Theorem 5.32, we expect $O(k^{-1})$ convergence with respect to k and order $O(M^{-N})$ convergence with respect to growing M.

Also in Table 5.6, we display the ratios

$$q_k(M,q,N) := log_{10}\frac{E_k(M,q,N)}{E_{10k}(M,q,N)}, \tag{5.140}$$

which indicate the rate at which the error $E_k(M,q,N)$ decays with respect to k. We expect the ratios $q_k(M,q,N)$ to tend to 1. In Table 5.6, we observe that the ratios indeed tend to the value of 1.

		$M=4$		$M=8$	
		$E_k(M,q,N)$	$q_k(M,q,N)$	$E_k(M,q,N)$	$q_k(M,q,N)$
$N=8, q=16$	k				
	100	6.8e-005		6.3e-007	
	1000	1.2e-006	1.7	4.0e-008	1.2
	10000	2.5e-007	0.7	4.9e-009	0.9
	100000	6.1e-009	1.6	5.4e-010	1.0
$N=10, q=20$					
	100	1.3e-004		8.3e-007	
	1000	2.9e-006	1.7	9.6e-008	1.0
	10000	1.7e-007	1.2	1.0e-008	1.0
	100000	1.2e-008	1.2	2.2e-010	1.7

Table 5.6: *The table displays the errors $E_k(M,q,N)$ and their ratios $q_k(M,q,N)$ that are defined in (5.140) to determine the convergence rate with respect to k.*

Example 4

In this example, we consider approximating integrals of the form

$$I_k[\log] := \int_0^1 \log(x)\exp(ikx)dx, \quad \text{and} \quad I_k\left[x^\beta\right] := \int_0^1 x^\beta \exp(ikx)dx, \quad \text{with} \quad -1<\beta<0,$$

using the composite Filon-Clenshaw-Curtis quadrature.

The integral $I_k[\log]$ is a model integral typically arising in boundary integral equations for high-frequency Helmholtz equation in two dimensions.

To avoid evaluation of log-function and x^β, $\beta<0$, at $x=0$, we compute the integrals

$$I_k[\log] := \int_a^1 \log(x)\exp(ikx)dx, \quad \text{and} \quad I_k[x^\beta] := \int_a^1 x^\beta \exp(ikx)dx, \tag{5.141}$$

with $a=10^{-20}$. We aim to investigate what values of parameters M and N are required

k	log	$\beta = -1/8$	$\beta = -1/4$	$\beta = -1/2$
10	3.36e-001	2.28e-001	2.28e-001	5.94e-001
100	5.42e-002	1.05e-002	2.89e-002	1.68e-001
1000	7.65e-003	2.38e-003	6.76e-003	5.62e-002
10000	9.91e-004	4.35e-004	1.30e-003	1.78e-002

Table 5.7: *The absolute values of the integrals in (5.141)*

to achieve a certain accuracy for varying k.

In the first experiment, we apply the composite FCC quadrature with fixed parameters $N = 16$ and $q = 12$ and increasing M. We start with $M = 2$ and compute the values of each integral with increasing M until the difference between the exact value and the computed value is less than $TOL = 1.0e - 012$. The results are displayed in Table 5.8 for the integrals in (5.141) with $\beta = -1/8$, $\beta = -1/4$ and $\beta = -1/2$. The absolute values of the exact solutions are displayed in Table 5.7.

	log		$\beta = -1/8$		$\beta = -1/4$		$\beta = -1/2$	
k	$E(k, M, q, N, \log)$	M	$E(k, M, q, N, f_\beta)$	M	$E(k, M, q, N, f_\beta)$	M	$E(k, M, q, N, f_\beta)$	M
10	4.7e-013	15	7.2e-013	15	7.6e-013	18	7.5e-013	46
100	4.7e-013	15	7.2e-013	15	7.6e-013	18	7.4e-013	46
1000	2.7e-013	14	7.0e-013	15	7.6e-013	18	7.5e-013	46
10000	3.1e-013	15	6.5e-013	15	7.7e-013	18	7.5e-013	46

Table 5.8: *The table displays the values of the parameter M required for the composite FCC to achieve an absolute accuracy of $TOL = 1.0e - 012$ for varying k as well as the error of the FCC $E(k, M, q, N, f)$, defined in (5.122). The parameters N and q are fixed at $N = 16$, $q = 12$.*

We observe from Table 5.8 that to maintain the same level of accuracy with increasing k, the total number of function evaluations of the composite FCC does not increase. However, as the singularity becomes more severe, the number of subintervals M has to grow rapidly.

In the second experiment, we apply the composite FCC quadrature with fixed parameters $N = 16$ and $q = 12$ to the integral:

$$I_k[x^{-1/4}] = \int_a^1 \frac{1}{x^{1/4}} \exp(ikx)dx.$$

In this experiment, the values of TOL vary and we study the values of parameter M required to achieve the absolute accuracy TOL. The results are displayed in Table 5.9.

k	$TOL = 1.0e-006$	$TOL = 1.0e-009$	$TOL = 1.0e-012$	$TOL = 1.0e-015$
10	5	9	18	34
100	6	9	18	33
1000	6	9	18	33
10000	6	10	18	33
ratio		14.8	10.4	11.22

Table 5.9: *The table of values M required for the composite FCC to achieve an accuracy of TOL with fixed $N = 16$, $q = 12$ and $\beta = -1/4$. In the bottom line of the table, the ratios defined in (5.142), averaged over k, are displayed.*

In the last line of the Table 5.9, we display the ratios:

$$ratio = \frac{\log_{10} \frac{TOL_{previous}}{TOL_{current}}}{\log_{10} \frac{M_{previous}}{M_{current}}}, \tag{5.142}$$

averaged over all k. The $ratio$ determines at which rate, s, the error of the composite FCC converges with respect to M, i.e. $E(k, M, q, N, f) \backsim O(M^{-s})$.

From Table 5.9, we observe that M remains largely fixed for increasing k.

Finally, in the third experiment in this example, we apply the composite FCC quadrature with fixed parameters $M = 10$ and $q = 12$ and increasing N. We start with $N = 8$ and compute the values of the integral with doubling N until the difference between the exact solution and the computed solution is less than TOL. We then record the value of N. The results are displayed in Table 5.10.

k	$TOL = 1.0e-006$	$TOL = 1.0e-009$	$TOL = 1.0e-012$	$TOL = 1.0e-015$
10	16	32	64	256
100	16	32	64	256
1000	16	32	64	256
10000	16	32	64	256

Table 5.10: *The table of values N required for the composite FCC to achieve an accuracy of TOL with fixed $M = 10$, $q = 12$, $\beta = -1/4$.*

From Table 5.10 we observe that N remains fixed for increasing k.

Chapter 6

Error of the fully-discrete Galerkin method

6.1 Introduction

In this chapter, we investigate the error of the fully-discrete Galerkin method, described in Chapter 2. Recall that the standard combined integral equation can be written in the form

$$\mathcal{R}_k v = f_k, \quad \text{on} \quad \Gamma, \tag{6.1}$$

where the operator $\mathcal{R}_k$ and the function f_k are defined in (2.15) and (2.16) respectively. We will consider them in more detail later.

The boundary Γ is parametrised using a 2π-periodic parametrisation $\gamma(s)$, $s \in [0, 2\pi]$. As before, we denote $(v, w)_{L^2(\Lambda)}$ as the ususal L^2-inner product of complex-valued functions, where $\Lambda \in [0, 2\pi]$. We denote $\|v\|^2_{L^2(\Lambda)} := (v, v)_{L^2(\Lambda)}$. When $\Lambda \subseteq [0, 2\pi]$, we denote the inner product and the norm simply by (v, w) and $\|v\|$.

The variational formulation of (6.1) seeks $v \in L^2([0, 2\pi])$ such that,

$$a_k(v, w) := (\mathcal{R}_k v, w) = (f_k, w), \qquad \text{for all} \;\; w \in L^2([0, 2\pi]). \tag{6.2}$$

The plan for this chapter is as follows. In Section 6.2, we describe and derive the error estimates of the semi-discrete Galerkin approximation to the solution of (6.2), closely following [40]. In Section 6.3, we describe the entries of the *stiffness* matrices arising from the Galerkin discretization of (6.2). Moreover, the entries of the stiffness matrix arising via the star-combined formulation will also be described. In practice, these entries can not be computed exactly. The numerical integration method described in Chapter 4 can be utilised to approximate the entries of the stiffness matrix. The numerical integration method uses the composite Filon-Clenshaw-Curtis quadrature, for which error estimates were derived in Chapter 5. In Section 6.4 we use the Strang Lemma to derive the error bounds of the fully- discrete Galerkin method. These error bounds incorporate the error of the inexact approximation of the entries of the stiffness matrix. Finally, in Section 6.5, we discuss the results of numerical experiments.

6.2 Error of the semi-discrete Galerkin approximation

Recall the approximation space $\mathbb{V}_k^{\boldsymbol{d}}$ for the Galerkin discretization is constructed using the ansatz (3.2),

$$v(s) = kV(s,k)\exp(ik\boldsymbol{\gamma}(s)\cdot\boldsymbol{a}). \tag{6.3}$$

The slowly-varying function $V(s,k)$ has different asymptotic behaviour in different parts of the boundary (see Chapter 3): the illuminated part (that we denote as Λ_2), transition parts (Λ_1 and Λ_3) and the shadow part of the boundary (Λ_4). In order to capture this behaviour, we introduce the partition of unity, see (2.35). In this thesis, we choose the partition of unity $\{\chi_j\}$, $j = 1, ..., 4$, with χ_j as characteristic functions:

$$\chi_j(s) = \begin{cases} 1, & \text{if } s \in \Lambda_j, \\ 0 & \text{if } s \notin \Lambda_j. \end{cases} \tag{6.4}$$

Then the approximation space $\mathbb{V}_k^{\boldsymbol{d}}$ is defined as follows,

$$\mathbb{V}_k^{\boldsymbol{d}} = \oplus_{j=1}^{3}\mathbb{V}_k^j, \tag{6.5}$$

where $\mathbb{V}_k^2$ is an approximation subspace for the illuminated part and $\mathbb{V}_k^1$ and $\mathbb{V}_k^3$ for transition parts. Since the solution in the shadow part of the boundary is exponentially small, see Theorem 3.3, we approximate the solution by zero in the shadow.

The subspaces $\mathbb{V}_k^j$ in (6.5) are defined as follows,

$$\mathbb{V}_k^j = \operatorname{span}\left\{\phi_{j,m}, \quad m = 0, ..., d_j : \quad \phi_{j,m}(s) = kP_m(s)\exp(ik\boldsymbol{\gamma}(s)\cdot\boldsymbol{a}) \ , \ s \in \Lambda_j\right\}, \tag{6.6}$$

where $\{P_m\}_{m=0}^{d_j}$ is the Chebyshev polynomial basis:

$$P_m(s) = T_m\left(\frac{2s-(b_j+a_j)}{b_j-a_j}\right), \qquad s \in [a_j, b_j] =: \Lambda_j, \tag{6.7}$$

and $T_m(x)$, $m = 0, 1, ...$, $x \in [-1, 1]$, are Chebyshev polynomials,

$$T_n(x) = \cos(n\arccos(x)).$$

The points a_j and b_j denote the end points of the interval Λ_j, $j = 1, ..., 3$.

We seek a **semi-discrete Galerkin solution** $\tilde{v} \in \mathbb{V}_k^{\boldsymbol{d}}$ which satisfies the following system of equations:

$$a_k(\tilde{v}, w) = (f_k, w), \quad \text{for all } w \in \mathbb{V}_k^{\boldsymbol{d}}. \tag{6.8}$$

The theorem below provides the error bound for the semi-discrete Galerkin approximation.

Theorem 6.1 (***Cea's Lemma***). *Suppose that a_k satisfies for all $v, w \in L^2([0, 2\pi])$ the two assumptions*

$$\begin{aligned} \textit{continuity} \quad |a_k(v,w)| &\leq B_k \|v\| \|w\|, \qquad B_k > 0 \qquad (6.9) \\ \textit{coercivity} \quad |a_k(v,v)| &\geq \alpha_k \|v\|^2, \qquad \alpha_k > 0. \qquad (6.10) \end{aligned}$$

Then both the weak form (6.2) and the Galerkin approximation (6.8) have unique solutions, $\tilde{v} \in \mathbb{V}_k^{\boldsymbol{d}}$. Moreover,

$$\|v - \tilde{v}\| \leq \left(\frac{B_k}{\alpha_k}\right) \min_{\tilde{w} \in \mathbb{V}_k^{\boldsymbol{d}}} \|v - \tilde{w}\|. \tag{6.11}$$

We can write the solution v as follows,

$$v(s) = \sum_{j=1}^{4} \chi_j v(s) = k \sum_{j=1}^{3} \chi_j(s) \exp(ik\boldsymbol{\gamma}(s) \cdot \boldsymbol{a}) V(s,k) + \chi_4(s) v(s), \tag{6.12}$$

and we write $\tilde{w} \in \mathbb{V}_k^{\boldsymbol{d}}$ as

$$\tilde{w}(s) = k \sum_{j=1}^{3} \chi_j(s) \exp(ik\boldsymbol{\gamma}(s) \cdot \boldsymbol{a}) \; p_j(s) \tag{6.13}$$

for some polynomial $p_j \in \mathbb{P}_{d_j}$ where $\mathbb{P}_{d_j}$ denotes the space of all polynomials of degree d_j, $j = 1, 2, 3$.

Using (6.12) and (6.13) in (6.11) we deduce,

$$\|v - \tilde{v}\| \leq \left(\frac{B_k}{\alpha_k}\right) \left[k \sum_{j=1}^{3} \inf_{p_j \in \mathbb{P}_{d_j}} \|V_j(\cdot, k) - p_j\|_{L^2(\Lambda_j)} + \|v\|_{L^2(\Lambda_4)} \right], \tag{6.14}$$

where

$$V_j(s,k) := V(s,k)\chi_j(s).$$

Hence, in order to estimate the error of the Galerkin approximation we require the error estimates of the best polynomial approximations in V_j, $j = 1, 2, 3$. These estimates can be found using Theorem 3.6 stated in Chapter 3, which provides the asymptotic estimates of the derivatives of V in the illuminated and transition parts of the boundary.

Choosing the intervals Λ_j as follows

$$\begin{aligned} \Lambda_1 = [t_1 - c_1 k^{-1/3}, t_1 + c_2 k^{-2/9}], \; \Lambda_2 = [t_1 + c_2 k^{-2/9}, t_2 - c_2 k^{-2/9}], \\ \Lambda_3 = [t_2 - c_2 k^{-2/9}, t_2 + c_1 k^{-1/3}], \; \Lambda_4 = [t_2 + c_1 k^{-1/3}, 2\pi] \cup [0, t_1 - c_1 k^{-1/3}], \end{aligned} \tag{6.15}$$

ensures that the error of the best polynomial approximation of V_j, $j = 1, 2, 3$ is small in

Λ_j and uniform, i.e. of the same order of convergence in k as $k \to \infty$ in the three intervals Λ_1, Λ_2 and Λ_3.

The error estimates for the best polynomial approximation of V in the illuminated and transition parts of the boundary are presented below in Theorem 6.2 and Theorem 6.3 respectively.

Theorem 6.2. *[40, Theorem 6.3] [**The best polynomial approximation in the illuminated part of the boundary**] For all fixed n, there exists $C_n > 0$ such that, for all $n \leq d+1$ and k sufficiently large,*

$$\inf_{p \in \mathbb{P}^d} \|V_2(\cdot, k) - p_2\|_{L^2(\Lambda_2)} \leq C_n k^{-2/3} \left(\frac{k^{1/9}}{d} \right)^n, \qquad \text{as} \quad d \to \infty. \tag{6.16}$$

Theorem 6.3. *[40, Theorem 6.4][**The best polynomial approximation in the transition part of the boundary**] For all $n \geq 2$ and $n \leq d+1$ there exists C_n independent of k such that, for $j = 1, 3$,*

$$\inf_{p \in \mathbb{P}^d} \|V_j(\cdot, k) - p_j\|_{L^2(\Lambda_j)} \leq C_n k^{-4/9} \left(\frac{k^{1/9}}{d} \right)^n, \qquad \text{as} \quad d \to \infty. \tag{6.17}$$

Moreover, in the shadow part of the boundary, the solution v is known to be exponentially small.

Theorem 6.4. *[40, Theorem 6.5][**The estimate of the solution in the shadow part of the boundary**] There exist positive constants c_0, c_0' such that for all k sufficiently large,*

$$\|v\|_{L^2(\Lambda_4)} \leq c_0' \exp(-c_0 k^\delta). \tag{6.18}$$

Substituting the estimates (6.16), (6.17) and (6.18) in (6.14), we deduce the error of the Galerkin approximation is bounded as follows.

Theorem 6.5. *[40, Theorem 6.7] Let $\tilde{v} \in \mathbb{V}_k^{\boldsymbol{d}}$ be the semi-discrete Galerkin solution of (6.8) with intervals Λ_j chosen as in (6.15). Suppose that polynomials of degree d_I are used in the illuminated zone and d_T in the transition zones. Then for all $n \geq 6$ and $n \leq d_I + 1$ and $n \leq d_T + 1$, there exist a constant $C_n > 0$ such that*

$$\|v - \tilde{v}\| \leq C_n \left(\frac{B_k}{\alpha_k} \right) \left[k \left\{ k^{-2/3} \left(\frac{k^{1/9}}{d_I} \right)^n + k^{-4/3} \left(\frac{k^{1/9}}{d_T} \right)^n \right\} + \exp\left(-c_0 k^\delta\right) \right], \tag{6.19}$$

where $6 \leq n \leq \min\{d_I, d_T\} + 1$, C_n is a constant independent of k.

The error bound (6.19) suggests that in order to maintain the accuracy as $k \to \infty$, the number of degrees of freedom, which is proportional to d, must grow at a rate slightly higher than $k^{1/9}$ as $k \to \infty$.

If $d_I = d_T = d$, then the error bound (6.19) simplifies to the following estimate.

Corollary 6.6. *[40, Corollary 6.8] Under the conditions of Theorem 6.5, suppose also that $d_I = d_T = d$. Then for $n \geq 6$ and $n \leq d+1$,*

$$\|v - \tilde{v}\| \leq \frac{B_k}{\alpha_k} k \left\{ k^{-2/3} \left(\frac{k^{1/9}}{d} \right)^n + \exp\left(-c_0 k^\delta\right) \right\}. \tag{6.20}$$

6.3 Inexact approximation of the entries of the stiffness matrix

The Galerkin method requires the assembly and solution of a system of linear equations (6.8). The coefficients of this system of linear equations are double integrals that are often highly-oscillatory and singular and the integrals cannot be computed exactly as we have seen for the case of the standard-combined formulation in Chapter 4.

In this section, we revisit the standard and star-combined potential integral equations and write out the entries of their stiffness matrices in Section 6.3.1 and Section 6.3.2 respectively. In practice, we use the numerical integration method developed in Chapters 4 and 5 for the approximation of highly-oscillatory singular integrals to calculate the entries of the stiffness matrix. In Section 6.3.3, we recall the error of the numerical integration method, i.e. the error of the approximation of the entries of the stiffness matrix derived in Chapter 5.

6.3.1 Standard formulation

Standard combined integral equation is defined as follows

$$\mathcal{R}_k v = f_k, \quad \text{on} \quad \Gamma \tag{6.21}$$

where the operator $\mathcal{R}_k$ is defined by

$$\mathcal{R}_k := \frac{1}{2} I + \mathcal{D}'_k - ik\mathcal{S}_k, \tag{6.22}$$

and the function f_k is defined as

$$f_k(\boldsymbol{x}) = \partial_{\boldsymbol{n}} u^I(\boldsymbol{x}) - iku^I(\boldsymbol{x}). \tag{6.23}$$

The semi-discrete Galerkin solution, $\tilde{v} \in \mathbb{V}_k^{\boldsymbol{d}}$, satisfies the linear system of equations (6.8) that can be written in the matrix form as follows:

$$\boldsymbol{R}\boldsymbol{V} = \boldsymbol{f}. \tag{6.24}$$

The matrix $\boldsymbol{R}$ is a discretization matrix that consists of 9 subblocks $\boldsymbol{R}_{l,j}$, $l, j = 1, 2, 3$:

$$\begin{pmatrix} R_{[1,1]} & R_{[1,2]} & R_{[1,3]} \\ R_{[2,1]} & R_{[2,2]} & R_{[2,3]} \\ R_{[3,1]} & R_{[3,2]} & R_{[3,3]} \end{pmatrix} \begin{pmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \\ \boldsymbol{V}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \\ \boldsymbol{f}_3 \end{pmatrix}, \tag{6.25}$$

with entries in each subblock of the form,

$$\boldsymbol{R}_{[l,j]}(n,m) = k^2 \frac{1}{2} \left(\int_{\Lambda_j} P_m(s) P_n(s) ds \right) \delta_{j,l} \tag{6.26a}$$

$$+ k^2 \int_{\Lambda_l} \int_{\Lambda_j} P_m(t) P_n(s) \left[\partial_{\boldsymbol{n}(s)} \Phi_k(\boldsymbol{\gamma}(s), \boldsymbol{\gamma}(t)) - ik \Phi_k(\boldsymbol{\gamma}(s), \boldsymbol{\gamma}(t)) \right] \times$$

$$\times \exp\left(ik \boldsymbol{a} \cdot [\gamma(t) - \gamma(s)] \right) |\gamma'(t)| dt ds, \tag{6.26b}$$

The integrals can be written in a more familiar notation from Chapter 4 as follows,

$$\boldsymbol{R}_{[l,j]}(n,m) = k^2 \frac{1}{2} \left(\int_{\Lambda_j} P_m(s) P_n(s) ds \right) \delta_{j,l} + k^2 \int_{\Lambda_l} \int_{\Lambda_j} M(s,t) \exp(ik\Psi(s,t)) dt ds,$$

where $M(s,t)$ is defined as in (4.13),

$$M(s,t) := g(s,t) \left[\partial_{\boldsymbol{n}(s)} \Phi_k(\boldsymbol{\gamma}(s), \boldsymbol{\gamma}(t)) - ik \Phi_k(\boldsymbol{\gamma}(s), \boldsymbol{\gamma}(t)) \right] \exp\left(-ik \left| \boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t) \right| \right), \tag{6.27}$$

and the phase function Ψ is defined as in (4.12),

$$\Psi(s,t) = |\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)| - \boldsymbol{a} \cdot (\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)) . \tag{6.28}$$

The integral (6.26a) is a slowly-varying one-dimensional integral that can be computed using classical quadrature rules (i.e. quadratures based on polynomial approximation of the integrand). On the other hand, the integral (6.26b) is a highly-oscillatory double integral that can be computed using the method described in Chapter 4.

The vector $\boldsymbol{f}$ consists of three subvectors: $\boldsymbol{f}_1$, $\boldsymbol{f}_2$ and $\boldsymbol{f}_3$. The entries of the subvector $\boldsymbol{f}_j$, $j = 1, 2, 3$, are defined as

$$\begin{aligned} \boldsymbol{f}_j(m) &:= (f_k, \phi_{j,m}) := \int_{\Lambda_j} \left(\partial_{\boldsymbol{n}} u^I - iku^I \right)(s) \overline{\phi_{j,m}(s)} ds \\ &= ik^2 \int_{\Lambda_j} \left(\boldsymbol{a} \cdot \boldsymbol{n}(s) - 1 \right) P_m(s) ds. \end{aligned} \tag{6.29}$$

The integrals (6.29) are slowly-varying and can be computed using classical quadratures.

6.3.2 Star-combined formulation

The star-combined integral equation [96] is defined as

$$\mathcal{A}_k v = g_k, \tag{6.30}$$

where the operator $\mathcal{A}_k$ is

$$\mathcal{A}_k := (\boldsymbol{x} \cdot \boldsymbol{n}(\boldsymbol{x}))\left(\frac{1}{2} + \mathcal{D}'_k\right) + (\boldsymbol{x} \cdot \boldsymbol{t}(\boldsymbol{x}))\frac{\partial}{\partial \boldsymbol{t}}\mathcal{S}_k + \eta_k(\boldsymbol{x})\mathcal{S}_k, \tag{6.31}$$

and the function g_k is given as

$$g_k(\boldsymbol{x}) := \boldsymbol{x} \cdot \nabla u^I(\boldsymbol{x}) + \eta_k(\boldsymbol{x}) u^I(\boldsymbol{x}), \tag{6.32}$$

where η_k is a "coupling" function,

$$\eta_k(\boldsymbol{x}) = \left(\frac{1}{2} - ik|\boldsymbol{x}|\right).$$

Having parameterised the boundary Γ using the 2π-periodic parametrisation $\gamma(s)$, $s \in [0, 2\pi]$, we seek the discrete Galerkin solution, $v \in \mathbb{V}_k^{\boldsymbol{d}}$ of the equation (6.30) by solving the system of equations:

$$(\mathcal{A}_k v, w) = (g_k, w) \qquad \text{for all} \quad w \in \mathbb{V}_k^{\boldsymbol{d}},$$

that can be written in the matrix form as follows:

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{g}. \tag{6.33}$$

Again, the matrix $\boldsymbol{A}$ is a discretization matrix that consists of 9 subblocks $\boldsymbol{A}_{l,j}$, $l, j = 1, 2, 3$:

$$\begin{pmatrix} A_{[1,1]} & A_{[1,2]} & A_{[1,3]} \\ A_{[2,1]} & A_{[2,2]} & A_{[2,3]} \\ A_{[3,1]} & A_{[3,2]} & A_{[3,3]} \end{pmatrix} \begin{pmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \\ \boldsymbol{V}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{g}_1 \\ \boldsymbol{g}_2 \\ \boldsymbol{g}_3 \end{pmatrix}. \tag{6.34}$$

Before we write out the entries of the matrix $\boldsymbol{A}$, let us consider the inner product arising of each term in the star-combined operator (6.31) applied to $\phi_{j,m}$ with $\phi_{l,n}$. We recall $\phi_{l,n}$ is defined in (6.6). For convenience let us denote $\mathcal{A}_{1,k}$, $\mathcal{A}_{2,k}$ and $\mathcal{A}_{3,k}$ as follows:

$$\mathcal{A}_k =: \mathcal{A}_{1,k} + \mathcal{A}_{2,k} + \mathcal{A}_{3,k},$$

where

$$\begin{aligned}
\mathcal{A}_{1,k} &= (\boldsymbol{\gamma}\cdot\boldsymbol{n})(s)\left(\frac{1}{2}+\mathcal{D}'_k\right),\\
\mathcal{A}_{2,k} &= (\boldsymbol{\gamma}\cdot\boldsymbol{t})(s)\frac{\partial}{\partial\boldsymbol{t}}\mathcal{S}_k,\\
\mathcal{A}_{3,k} &= \left(\frac{1}{2}-ik|\boldsymbol{\gamma}(s)|\right)\mathcal{S}_k.
\end{aligned}$$

The inner product of $\mathcal{A}_{1,k}\phi_{j,m}(s)$ with $\phi_{l,n}(s)$ is of the form,

$$\begin{aligned}
(\mathcal{A}_{1,k}\phi_{j,m}(s),\phi_{l,n}(s)) &= \int_{\Lambda_j}(\boldsymbol{\gamma}\cdot\boldsymbol{n})(s)\left(\frac{1}{2}+\mathcal{D}'_k\right)\phi_{j,m}(s)\overline{\phi_{l,n}(s)}ds\\
&= k^2\frac{1}{2}\left(\int_{\Lambda_j}(\boldsymbol{\gamma}\cdot\boldsymbol{n})(s)P_n(s)P_m(s)ds\right)\delta_{j,l}\\
+\quad k^2\int_{\Lambda_l}\int_{\Lambda_j}&(\boldsymbol{\gamma}\cdot\boldsymbol{n})(s)P_m(t)P_n(s)\frac{\partial\Phi_k(\boldsymbol{\gamma}(s),\boldsymbol{\gamma}(t))}{\partial\boldsymbol{n}(s)}\exp\left(ik\boldsymbol{a}\cdot[\boldsymbol{\gamma}(t)-\boldsymbol{\gamma}(s)]\right)|\boldsymbol{\gamma}'(t)|dtds.
\end{aligned}$$

We can write the double integral in the form (4.11),

$$\begin{aligned}
(\mathcal{A}_{1,k}\phi_{j,m}(s),\phi_{l,n}(s)) &= k^2\frac{1}{2}\left(\int_{\Lambda_j}(\boldsymbol{\gamma}\cdot\boldsymbol{n})(s)P_n(s)P_m(s)ds\right)\delta_{j,l} &(6.35)\\
&+ k^2\int_{\Lambda_l}\int_{\Lambda_j}M_1(s,t)\exp\left(ik\Psi(s,t)\right)dtds, &(6.36)
\end{aligned}$$

where

$$M_1(s,t) := (\boldsymbol{\gamma}\cdot\boldsymbol{n})(s)P_m(t)P_n(s)\frac{\partial\Phi_k(\boldsymbol{\gamma}(s),\boldsymbol{\gamma}(t))}{\partial\boldsymbol{n}(s)}\exp\left(-ik\left|\boldsymbol{\gamma}(t)-\boldsymbol{\gamma}(s)\right|\right)|\boldsymbol{\gamma}'(t)| \quad (6.37)$$

and the phase function Ψ is defined as in (6.28).

The single integral (6.35) is slowly-varying and can be computed using classical quadratures. On the other hand, the integral (6.36) is highly-oscillatory double integral that can be computed using methodology described in Chapters 4 and 5.

The inner product of $\mathcal{A}_{2,k}\phi_{j,m}(s)$ with $\phi_{l,n}(s)$ is more complicated: we use integration by parts to write

$$\begin{aligned}
(\mathcal{A}_{2,k}\phi_{j,m}(s),\phi_{l,n}(s)) &:= \int_{\Lambda_l}(\boldsymbol{\gamma}\cdot\boldsymbol{t})(s)\frac{\partial}{\partial s}\mathcal{S}_k\phi_{j,m}(s)\overline{\phi_{l,n}(s)}ds\\
&= \left[(\boldsymbol{\gamma}\cdot\boldsymbol{t})(s)\overline{\phi_{l,n}(s)}\mathcal{S}_k\phi_{j,m}(s)\right]_{s=a_l}^{b_l} - \int_{\Lambda_l}\left[(\boldsymbol{\gamma}\cdot\boldsymbol{t})(s)\overline{\phi_{l,n}(s)}\right]'\mathcal{S}_k\phi_{j,m}(s)ds, \quad (6.38)
\end{aligned}$$

where a_l and b_l denote the end points of the interval Λ_l and $\overline{y}$ denote a complex conjugate of y.

The first term on the right hand side of (6.38) is a difference between two one-dimensional integrals. For example at $s = a_l$,

$$\begin{aligned}
&[(\boldsymbol{\gamma}\cdot\boldsymbol{t})(a_l)]\,\overline{\phi_{l,n}(a_l)}\mathcal{S}_k\phi_{j,m}(a_l) := \\
&k^2(\boldsymbol{\gamma}\cdot\boldsymbol{t})(a_l)P_n(a_l)\exp(-ik\boldsymbol{\gamma}(a_l)\cdot\boldsymbol{a})\int_{\Lambda_j}\Phi_k(\boldsymbol{\gamma}(a_l),\boldsymbol{\gamma}(t))P_m(t)\exp(ik\boldsymbol{a}\cdot\boldsymbol{\gamma}(t))\,|\boldsymbol{\gamma}'(t)|dt \\
= \;& k^2(\boldsymbol{\gamma}\cdot\boldsymbol{t})(a_l)\int_{\Lambda_j}\Phi_k(\boldsymbol{\gamma}(a_l),\boldsymbol{\gamma}(t))P_n(a_l)P_m(t)\exp(ik\boldsymbol{a}\cdot(\boldsymbol{\gamma}(t)-\boldsymbol{\gamma}(a_l)))\,|\boldsymbol{\gamma}'(t)|dt. \quad (6.39)
\end{aligned}$$

The integral in (6.39) can be written as follows:

$$[(\boldsymbol{\gamma}\cdot\boldsymbol{t})(a_l)]\,\overline{\phi_{l,n}(a_l)}\mathcal{S}_k\phi_{j,m}(a_l) = k^2\int_{\Lambda_j} f(a_l,t)\exp(ik\Psi(a_l,t))\,dt, \quad (6.40)$$

where

$$f(s,t) = (\boldsymbol{\gamma}\cdot\boldsymbol{t})(s)\Phi_k(\boldsymbol{\gamma}(s),\boldsymbol{\gamma}(t))\exp(-ik|\boldsymbol{\gamma}(s)-\boldsymbol{\gamma}(t)|)\,P_n(s)P_m(t)|\boldsymbol{\gamma}'(t)| \quad (6.41)$$

is a slowly-varying function and the phase function Ψ is defined in (6.28).

The integral in (6.40) is a non-canonical highly-oscillatory integral that can be transformed into a canonical integral by the method described in Section 4.2 of Chapter 4. Then the Filon-Clenshaw-Curtis quadrature can be applied to compute the resulting integral. Moreover, if the integrand is singular, the composite version of the Filon-Clenshaw-Curtis quadrature should be applied.

The second term on the right-hand side of (6.38), on the other hand, is a double integral of the form

$$\begin{aligned}
&\int_{\Lambda_l}\left[(\boldsymbol{\gamma}\cdot\boldsymbol{t})(s)\overline{\phi_{l,n}(s)}\right]'\mathcal{S}_k\phi_{j,m}(s)ds \\
&= k^2\int_{\Lambda_l}\int_{\Lambda_j}\left[(\boldsymbol{\gamma}\cdot\boldsymbol{t})'(s)P_n(s) - ik(\boldsymbol{\gamma}'(s)\cdot\boldsymbol{a})(\boldsymbol{x}\cdot\boldsymbol{t})(s)P_n(s) + (\boldsymbol{x}\cdot\boldsymbol{t})(s)P_n'(s)\right]\times \\
&\qquad\times\Phi_k(\boldsymbol{\gamma}(s),\boldsymbol{\gamma}(t))P_m(t)\exp(ik\boldsymbol{a}\cdot[\boldsymbol{\gamma}(t)-\boldsymbol{\gamma}(s)])\,|\gamma'(t)|dtds.
\end{aligned}$$

The double integral can be written in the form,

$$\int_{\Lambda_l}\left[(\boldsymbol{\gamma}\cdot\boldsymbol{t})(s)\overline{\phi_{l,n}(s)}\right]'\mathcal{S}_k\phi_{j,m}(s)ds = k^2\int_{\Lambda_l}\int_{\Lambda_j} M_2(s,t)\exp(ik\Psi(s,t))\,dtds,$$

where

$$\begin{aligned}
M_2(s,t) \;:=\; &\left[(\boldsymbol{\gamma}\cdot\boldsymbol{t})'(s)P_n(s) - ik(\boldsymbol{\gamma}'(s)\cdot\boldsymbol{a})(\boldsymbol{x}\cdot\boldsymbol{t})(s)P_n(s) + (\boldsymbol{x}\cdot\boldsymbol{t})(s)P_n'(s)\right]P_m(t)\times \\
&\times\Phi_k(\boldsymbol{\gamma}(s),\boldsymbol{\gamma}(t))\exp(-ik\,|\boldsymbol{\gamma}(t)-\boldsymbol{\gamma}(s)|)\,|\boldsymbol{\gamma}'(t)|, \quad (6.42)
\end{aligned}$$

and the phase function Ψ is defined as in (6.28).

Finally, the inner product of $\mathcal{A}_{3,k}\phi_{j,m}(s)$ with $\phi_{l,n}(s)$ is of the form

$$\begin{aligned}
&(\mathcal{A}_{3,k}\phi_{j,m}(s), \phi_{l,n}(s)) = \int_{\Lambda_j} \left(\frac{1}{2} - ik|\boldsymbol{\gamma}(s)|\right) \mathcal{S}_k\phi_{j,m}(s)\overline{\phi_{l,n}(s)}ds \\
&= k^2 \int_{\Lambda_l}\int_{\Lambda_j} P_m(t)P_n(s) \left(\frac{1}{2} - ik|\boldsymbol{\gamma}(s)|\right) \Phi_k(s,t) \exp\left(ik\boldsymbol{a}\cdot[\boldsymbol{\gamma}(t)-\boldsymbol{\gamma}(s)]\right)|\boldsymbol{\gamma}'(t)|dtds \\
&= k^2 \int_{\Lambda_l}\int_{\Lambda_j} M_3(s,t)\exp\left(ik\Psi(s,t)\right)dsdt,
\end{aligned}$$

with

$$M_3(s,t) = P_m(t)P_n(s)\left(\frac{1}{2} - ik|\boldsymbol{\gamma}(s)|\right)\Phi_k(s,t)\exp\left(-ik\left|\boldsymbol{\gamma}(t)-\boldsymbol{\gamma}(s)\right|\right)|\boldsymbol{\gamma}'(t)|, \tag{6.43}$$

and the phase function Ψ is defined as in (6.28).

Now, we can write out the (n,m)-th entry of the (l,j)-th subblock of the matrix $\boldsymbol{A}$ as follows:

$$\boldsymbol{A}_{[l,j]}(n,m) = \frac{k^2}{2}\left(\int_{\Lambda_j}(\gamma\cdot\boldsymbol{n})(s)P_n(s)P_m(s)ds\right)\delta_{j,l} \tag{6.44a}$$

$$+\, k^2\int_{\Lambda_l}\int_{\Lambda_j} M_1(s,t)\exp\left(ik\Psi(s,t)\right)dtds \tag{6.44b}$$

$$+\, k^2\int_{\Lambda_j} f(b_l,t)\exp(ik\Psi(b_l,t))dt \tag{6.44c}$$

$$-\, k^2\int_{\Lambda_j} f(a_l,t)\exp(ik\Psi(a_l,t))dt \tag{6.44d}$$

$$+\, k^2\int_{\Lambda_l}\int_{\Lambda_j} M_2(s,t)\exp\left(ik\Psi(s,t)\right)dtds \tag{6.44e}$$

$$+\, k^2\int_{\Lambda_l}\int_{\Lambda_j} M_3(s,t)\exp\left(ik\Psi(s,t)\right)dtds, \tag{6.44f}$$

where M_1, M_2 and M_3 are defined in (6.37), (6.42) and (6.43) respectively, the function f is defined in (6.41) and Ψ is defined in (6.28).

Integrals (6.44a) and (6.44b) correspond to the inner product $\mathcal{A}_{1,k}\phi_{j,m}(s)$ with $\phi_{l,n}(s)$. Integrals (6.44c), (6.44d) and (6.44e) correspond to the inner product $\mathcal{A}_{2,k}\phi_{j,m}(s)$ with $\phi_{l,n}(s)$. Finally, the remaining integrals (6.44f) correspond to the inner product $\mathcal{A}_{3,k}\phi_{j,m}(s)$ with $\phi_{l,n}(s)$.

Computing each entry of the discretization matrix $\boldsymbol{A}$ consists of computing highly-oscillatory double integrals (6.44b), (6.44f) and (6.44e); a slowly-varying one-dimensional integral (6.44a); and highly-oscillatory one-dimensional integrals (6.44c) and (6.44d).

The entries of the vector $\boldsymbol{g}$ on the right hand-side of the matrix equation (6.33) are one-dimensional slowly-varying integrals of the form

$$\begin{aligned}\boldsymbol{g}_j(m) &:= (g_k, \phi_{j,m}) \\ &= ik^2 \int_{\Lambda_j} \left[(\boldsymbol{a}\cdot\boldsymbol{n})(s)(\boldsymbol{\gamma}\cdot\boldsymbol{n})(s) + (\boldsymbol{a}\cdot\boldsymbol{t})(s)(\boldsymbol{\gamma}\cdot\boldsymbol{t})(s) + \left(\frac{1}{2ik} - |\gamma(s)|\right)\right] P_m(s)ds,\end{aligned}$$

that can be computed using classical quadratures.

6.3.3 The error in the inexact approximation of the entries

From Theorem 4.3, we know the Galerkin discretisation matrix for the standard and star-combined formulations consists of single integrals, for example (6.40), as well as double integrals of the form,

$$J_k := \int_{\Lambda_l}\int_{\Lambda_j} M(s,t)\exp\left(ik\Psi(s,t)\right)dtds, \tag{6.45}$$

where the function M is slowly-varying function and has an algebraic singularity and the phase-function Ψ is defined in (4.12). Using the numerical integration method described in Chapter 4, we can write the integral (6.45) in the form

$$J_k = \sum_{j=0}^{\mathcal{J}-1} \int_{\tau_j}^{\tau_{j+1}} F_{j+1}(\tau)\exp(ik\tau)d\tau \tag{6.46}$$

where τ_j, $j = 0,\ldots,\mathcal{J}-1$ are known and F_j, $j = 1,...,\mathcal{J}$ are one-dimensional slowly-varying integrals.

Moreover, the single integrals with a non-canonical oscillators such as integral in (6.40), can be written in the form

$$I_k = \int_{\tau_j}^{\tau_{j+1}} \tilde{f}(\tau)\exp(ik\tau)d\tau,$$

where $\tilde{f}$ may be singular.

The integral I_k as well as integrals on the right-hand side in (6.46) can be efficiently computed using the composite Filon-Clenshaw-Curtis method discussed in Chapter 5. Denote the composite Filon-Clenshaw-Curtis approximation to J_k in (6.46) as $\tilde{J}_k$:

$$\tilde{J}_k := \sum_{j=0}^{\mathcal{J}-1} Q_{k,N}^{[\tau_j,\tau_{j+1}]}[F_{j+1}].$$

Then from Theorem 5.32, we deduce that provided q and N satisfy $q \geq (N+1-s)/(1+$

$\beta - s$), where N is the order of the FCC rule on each subinterval, M is the number of subintervals and q is a parameter of the grading defined in (5.99), and $\beta \in (0,1)$ denotes the "degree" of the algebraic singularity, see Definition 5.26, the following holds,

$$|J_k - \tilde{J}_k| \leq C \left(\frac{1}{k}\right)^s \left(\frac{1}{M}\right)^{N+1-s}, \tag{6.47}$$

where $C > 0$ is a constant independent of k and M.

6.4 Strang's Lemma and the error of the fully-discrete Galerkin method

In practice, we seek the solution of the system of equations (6.8),

$$a_k(\tilde{v}, w) = (f_k, w), \quad \text{for all } w \in \mathbb{V}_k^{\boldsymbol{d}},$$

with the left-hand side approximated rather than known exactly. We denote by $\tilde{a}_k(v, w)$ the approximation of $a_k(v, w)$ for $v, w \in \mathbb{V}_k^{\boldsymbol{d}}$.

We then say that $\tilde{\tilde{v}} \in \mathbb{V}_k^{\boldsymbol{d}}$ is the **fully-discrete Galerkin solution** of the following system of equations,

$$\tilde{a}_k(\tilde{\tilde{v}}, w) = (f_k, w), \quad \text{for all } w \in \mathbb{V}_k^{\boldsymbol{d}}. \tag{6.48}$$

The following theorem, known as the *Strang Lemma*, provides an error estimate for the Galerkin approximation (6.48).

Theorem 6.7. *[30, Theorem 4.2.2][**Strang's Lemma**] Consider the family of discrete problems (6.48) for which the associated sesquilinear form $\tilde{a}_k$ is coercive and continuous:*

1) $\tilde{a}_k(v, v) \geq \tilde{\alpha}_k ||v||^2 \quad \forall v \in \mathbb{V}_k^{\boldsymbol{d}}$,

2) $|\tilde{a}_k(u, v)| \leq \tilde{B}_k ||u|| ||v|| \quad \forall u, v \in \mathbb{V}_k^{\boldsymbol{d}}$.

Then the following holds:

$$\|v - \tilde{\tilde{v}}\| \leq \inf_{z \in \mathbb{V}_k^{\boldsymbol{d}}} \left\{ \left(1 + \frac{\tilde{B}_k}{\tilde{\alpha}_k}\right) \|v - z\| + \frac{1}{\tilde{\alpha}_k} \sup_{w \in \mathbb{V}_k^{\boldsymbol{d}}} \frac{|a_k(z, w) - \tilde{a}_k(z, w)|}{\|w\|} \right\}. \tag{6.49}$$

In order to apply this result to determine the error estimates for the fully-discrete Galerkin method, we need some additional work. Since, in practice, we are using the Chebyshev polynomial basis as the approximation space $\mathbb{V}_k^{\boldsymbol{d}}$, we need to state some properties of the Chebyshev polynomials in Section 6.4.1. Then, in Section 6.4.2, we incorporate the error of the inexact approximation of $a_k(\tilde{v}, w)$ given in (6.47), in the estimate (6.49) given by the Strang Lemma. Finally in Theorem 6.15 we determine the error of the fully-discrete Galerkin method.

6.4.1 Properties of the Chebyshev-weighted norm

In the following lemma we define the Chebyshev weighted norm and determine the Chebyshev weighted norm of the linear combination of Chebyshev polynomials.

Lemma 6.8 (The Parseval's equality for Chebyshev expansions)**.** *Denote $I := [-1, 1]$. Consider a polynomial $p(x)$ defined as*

$$p(x) = \sum_{j=0}^{d} c_j T_j(x), \qquad x \in I,$$

where $T_j(x)$, $j = 0, 1, 2, ...$, $x \in I$ are Chebyshev polynomials defined in (5.24). Let $\|\cdot\|_{L^2(I),\omega}$ be the Chebyshev-weighted norm,

$$\|p\|_{L^2(I),\omega} := \left(\int_{-1}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}} dx \right)^{1/2}.$$

Then p has the following Chebyshev-weighted norm,

$$\|p\|^2_{L^2(I),\omega} = \pi |c_0|^2 + \frac{\pi}{2} \sum_{j=1}^{d} |c_j|^2. \tag{6.50}$$

Proof. Chebyshev polynomials T_j, $j = 0, 1, 2, ...$ satisfy the orthogonality relation:

$$\int_{-1}^{1} T_n(x) T_m(x) \frac{dx}{\sqrt{1-x^2}} = \begin{cases} 0 & : n \neq m \\ \pi & : n = m = 0 \\ \pi/2 & : n = m \neq 0 \end{cases}$$

Therefore,

$$\|p\|^2_{L^2(I),\omega} = \int_{-1}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}} dx = \sum_{j=0}^{d} \sum_{m=0}^{d} c_j c_m \int_{-1}^{1} \frac{T_j(x) T_m(x)}{\sqrt{1-x^2}} dx = \pi |c_0|^2 + \frac{\pi}{2} \sum_{j=1}^{d} |c_j|^2.$$

Thus the result follows. □

The following proposition shows that the L_∞-norm of p is bounded by the Chebyshev-weighted L_2-norm.

Proposition 6.9. *For a polynomial p defined as $p(x) = \sum_{j=0}^{d} c_j T_j(x)$, $x \in I$ the following inequality holds,*

$$||p||^2_{L^\infty(I)} \leq \frac{2(d+1)}{\pi} ||p||^2_{L^2(I),\omega} \tag{6.51}$$

Proof. Since $p(x) = \sum_{j=0}^{d} c_j T_j(x)$, and $|T_j(x)| \leq 1$, for all $|x| \leq 1$, we deduce

$$\begin{aligned}(p(x))^2 = \left(\sum_{j=0}^{d} c_j T_j(x)\right)^2 &\leq \left(\sum_{j=0}^{d} |c_j^2|\right)\left(\sum_{j=0}^{d} |T_j(x)|^2\right) \\ &\leq (d+1)\left(\sum_{j=0}^{d} |c_j|^2\right).\end{aligned} \tag{6.52}$$

On the other hand, using Lemma 6.8, we obtain

$$\begin{aligned}\sum_{j=0}^{d} |c_j|^2 &= \frac{2}{\pi}\left(\|p\|^2_{L^2(I),\omega} - \frac{\pi}{2}|c_0|^2\right) \\ &\leq \frac{2}{\pi}\|p\|^2_{L^2(I),\omega}.\end{aligned} \tag{6.53}$$

Finally, combining (6.53) and (6.53), we obtain the result (6.51),

$$p^2(x) \leq \frac{2(d+1)}{\pi}\|p\|^2_{L^2(I),\omega}.$$

□

In the following lemma, we obtain a bound the Chebyshev-weighted norm of the polynomial p in terms of the L^2-norm.

Lemma 6.10. ***Inverse estimate for the weighted norm.*** *For the polynomial $p(x) = \sum_{j=0}^{d} c_j T_j(x)$, the Chebyshev-weighted L^2-norm of p is bounded by L^2-norm of p as follows:*

$$\|p\|^2_{L^2(I),\omega} \leq Cd\|p\|^2_{L^2(I)},$$

where C is independent of d.

Proof. [93, proof of Theorem 3.96] Set $\varepsilon_{\mu d} = 1 - \cos(1/\mu d)$ for $\mu \geq 1$. We write

$$\int_{-1}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}}dx = \int_{-1}^{-1+\varepsilon_{\mu d}} \frac{(p(x))^2}{\sqrt{1-x^2}}dx + \int_{-1+\varepsilon_{\mu d}}^{1-\varepsilon_{\mu d}} \frac{(p(x))^2}{\sqrt{1-x^2}}dx + \int_{1-\varepsilon_{\mu d}}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}}dx. \tag{6.54}$$

Since the first and the last integrals in (6.54) are equal, we simplify (6.54),

$$\int_{-1}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}}dx = 2\int_{1-\varepsilon_{\mu d}}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}}dx + \int_{-1+\varepsilon_{\mu d}}^{1-\varepsilon_{\mu d}} \frac{(p(x))^2}{\sqrt{1-x^2}}dx. \tag{6.55}$$

We can estimate the integrals on the right hand side of (6.55) by observing that the following inequalities hold:

1.
$$\max_{x\in[1,1-\varepsilon_{\mu p}]} |p(x)| \leq \|p(x)\|_{L_\infty(I)}$$

2.
$$\max_{x\in[-1+\varepsilon_{\mu d},1-\varepsilon_{\mu d}]} \frac{1}{\sqrt{1-x^2}} \leq \frac{1}{\sqrt{1-(1-\varepsilon_{\mu d})^2}} = \frac{1}{\sin(1/\mu d)} \leq 2\mu d$$

3.
$$\int_{1-\varepsilon_{\mu d}}^{1} \frac{1}{\sqrt{1-x^2}} dx = \pi/2 - \arcsin(1-\varepsilon_{\mu d}) = \frac{1}{\mu d}$$

Then applying inequality 1 to the first integral on the right hand side of (6.55), and inequality 2 to the second integral, we deduce,

$$\int_{-1}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}} dx \leq 2\mu d \int_{-1+\varepsilon_{\mu d}}^{1-\varepsilon_{\mu d}} (p(x))^2 dx + 2||p||^2_{L^\infty(I)} \int_{1-\varepsilon_{\mu d}}^{1} \frac{1}{\sqrt{1-x^2}} dx.$$

Finally using inequality 3, we conclude,

$$\int_{-1}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}} dx \leq 2\mu d \int_{-1+\varepsilon_{\mu d}}^{1-\varepsilon_{\mu d}} (p(x))^2 dx + \frac{2}{\mu d} ||p||^2_{L^\infty(I)}.$$

Then, using Proposition 6.9, we deduce,

$$\int_{-1}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}} dx \leq 2\mu d \int_{-1}^{1} (p(x))^2 dx + \frac{4(d+1)}{\pi \mu d} \int_{-1}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}} dx.$$

Therefore,

$$\left(1 - \frac{4(d+1)}{\pi \mu d}\right) \int_{-1}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}} \leq 2\mu d \int_{-1}^{1} (p(x))^2 dx.$$

Selecting sufficiently large μ, we obtain

$$\int_{-1}^{1} \frac{(p(x))^2}{\sqrt{1-x^2}} dx \leq Cd \int_{-1}^{1} (p(x))^2 dx,$$

where $C > 0$ is a constant independent of d. □

6.4.2 Error of the fully-discrete Galerkin method

We write the solution v of (6.1) as in (6.12),

$$v(s) = \sum_{j=1}^{4} \chi_j v(s) = k \sum_{j=1}^{3} \chi_j(s) V(s,k) \exp(ik\boldsymbol{\gamma}(s)\cdot \boldsymbol{a}) + \chi_4(s) v(s), \tag{6.56}$$

and we write $z \in \mathbb{V}_k^{\boldsymbol{d}}$ as

$$z(s) = k \sum_{j=1}^{3} \chi_j(s) p_j(s) \exp(ik\boldsymbol{\gamma}(s) \cdot \boldsymbol{a}). \tag{6.57}$$

Here $p_j \in \mathbb{P}_d$ is defined as

$$p_j(s) := \sum_{n=0}^{d} \alpha_n P_n(s),$$

where P_n, $n = 0, \ldots, d$ are Chebyshev polynomials defined in (6.7).

Remark 6.11. *The Chebyshev-weighted norm of $p_j(s)$, $s \in \Lambda_j := [a_j, b_j]$, is defined as follows,*

$$\|p_j\|^2_{L^2(\Lambda_j),\omega} := \int_{a_j}^{b_j} \frac{(p_j(s))^2}{\sqrt{(s-a_j)(b_j-s)}} ds.$$

The result of Lemma 6.8 is equivalent to

$$\|p_j\|^2_{L^2(\Lambda_j),\omega} = \pi|\alpha_0|^2 + \frac{\pi}{2} \sum_{n=1}^{d} |\alpha_n|^2. \tag{6.58}$$

Furthermore, Lemma 6.10 also implies

$$\|p_j\|^2_{L^2(\Lambda_j),\omega} \leq Cd \|p_j\|^2_{L^2(\Lambda_j)}.$$

For convenience, assume that the approximation space $\mathbb{V}_k^{\boldsymbol{d}}$ has $\boldsymbol{d} = (d, d, d)$. In other words, we assume that the three subspaces of $\mathbb{V}_k^{\boldsymbol{d}}$ are of the same dimension d. We write $w \in \mathbb{V}_k^{\boldsymbol{d}}$ similarly to (6.57) as follows

$$\begin{aligned} w(s) &= k \sum_{j=1}^{3} \chi_j(s) q_j(s) \exp(ik\boldsymbol{\gamma}(s) \cdot \boldsymbol{a}) \qquad \text{with } q_j(s) := \sum_{n=0}^{d} \beta_n P_n(s) \\ &= k \sum_{j=1}^{3} \chi_j(s) \left(\sum_{n=0}^{d} \beta_n P_n(s) \right) \exp(ik\boldsymbol{\gamma}(s) \cdot \boldsymbol{a}) \\ &= \sum_{j=1}^{3} \chi_j(s) \left(\sum_{n=0}^{d} \beta_n \phi_{j,n}(s) \right), \end{aligned} \tag{6.59}$$

where $\phi_{j,n}$ is a basis element of the approximation space $\mathbb{V}_k^{\boldsymbol{d}}$ as defined in (6.6).

In estimate (6.49) given in Theorem 6.7, the second term corresponds to the error contribution of the inexact approximation of the coefficients in (6.48). Our aim is to find an upper bound on the error (of the inexact approximation of coefficients) that would ensure that the total error (6.49) is of the same order as in estimate (6.20). We do this by introducing a parameter ε that denotes the maximum error among the approximations of

all coefficients in (6.48).

Definition 6.12. *We denote ε by*

$$\varepsilon := \max_{j,n,l,m} \left| a_k(\phi_{j,n}, \phi_{l,m}) - \tilde{a}_k(\phi_{j,n}, \phi_{l,m}) \right| . \tag{6.60}$$

In the following lemma, we derive the bound on $|a_k(z,w) - \tilde{a}_k(z,w)|$ for $z, w \in \mathbb{V}_k^{\boldsymbol{d}}$ in terms of ε and d.

Lemma 6.13. *For any z and $w \in \mathbb{V}_k^{\boldsymbol{d}}$, with $\boldsymbol{d} = (d,d,d)^T$,*

$$|a_k(z,w) - \tilde{a}_k(z,w)| \leq C\varepsilon \left(\frac{d^2}{k} \right) \|z\| \|w\|, \tag{6.61}$$

where ε is defined as in (6.60) and where $C > 0$ is a constant independent of k and d.

Proof. For $j = 1,2,3,4$, denote

$$z_j(s) = \chi_j(s) z(s), \quad \text{and} \quad w_j(s) = \chi_j(s) w(s).$$

Then,

$$z(s) = z_1(s) + z_2(s) + z_3(s) + z_4(s), \quad \text{and} \quad w(s) = w_1(s) + w_2(s) + w_3(s) + w_4(s),$$

and[1]

$$|a_k(z,w) - \tilde{a}_k(z,w)| \leq \sum_{j=1}^{4} \sum_{l=1}^{4} |a_k(z_l, w_j) - \tilde{a}_k(z_l, w_j)|. \tag{6.62}$$

Let us now fix l and j and consider $|a_k(z_l, w_j) - \tilde{a}_k(z_l, w_j)|$. For ease of notation, we drop the indices by introducing

$$\mathfrak{z}(s) = z_l(s), \quad \text{and} \quad \mathfrak{w}(s) = w_j(s),$$

[1] By the properties of the sesquilinear form, we have

$$\begin{aligned} |a_k(\alpha z_1 + \beta z_2, w) - \tilde{a}_k(\alpha z_1 + \beta z_2, w)| &= |\alpha a_k(z_1, w) + \beta a_k(z_2, w) - \alpha \tilde{a}_k(z_1, w) - \beta \tilde{a}_k(z_2, w)| \\ &\leq \alpha |a_k(z_1, w) - \tilde{a}_k(z_1, w)| + \beta |a_k(z_2, w) - \tilde{a}_k(z_2, w)|. \end{aligned}$$

Hence (6.62) follows.

Then, $\mathfrak{z}(s)$ and $\mathfrak{w}(s)$ are of the form (6.59),

$$\begin{aligned}\mathfrak{z}(s) &= \sum_{n=0}^{d} \alpha_n \phi_{j,n}(s) \\ \mathfrak{w}(s) &= \sum_{n=0}^{d} \beta_n \phi_{j,n}(s)\end{aligned}$$

Therefore,

$$\begin{aligned}|a_k(\mathfrak{z},\mathfrak{w}) - \tilde{a}_k(\mathfrak{z},\mathfrak{w})| &\leq \left|\sum_{m=0}^{d}\sum_{n=0}^{d} \alpha_m\beta_n \left(a_k(\phi_{j,n},\phi_{l,m}) - \tilde{a}_k(\phi_{j,n},\phi_{l,m})\right)\right| \\ &\leq \sum_{m=0}^{d}\sum_{n=0}^{d} |\alpha_m||\beta_n| \left|a_k(\phi_{j,n},\phi_{l,m}) - \tilde{a}_k(\phi_{j,n},\phi_{l,m})\right| \\ &\leq \varepsilon \left\{\sum_{m=0}^{d} |\alpha_m|\right\}\left\{\sum_{n=0}^{d} |\beta_n|\right\}. \end{aligned} \tag{6.63}$$

Furthermore,

$$\mathfrak{z}(s) = kp(s)\exp(ik\boldsymbol{\gamma}(s)\cdot\boldsymbol{a}), \qquad \text{where } p(s) := \sum_{n=0}^{d} \alpha_n P_n(s), \tag{6.64}$$

$$\mathfrak{w}(s) = kq(s)\exp(ik\boldsymbol{\gamma}(s)\cdot\boldsymbol{a}), \qquad \text{where } q(s) := \sum_{n=0}^{d} \beta_n P_n(s). \tag{6.65}$$

Lemma 6.8 (see equation (6.58)) gives

$$\sum_{m=0}^{d} |\alpha_m|^2 \leq \frac{2}{\pi}\|p\|^2_{L^2(\Lambda_l),\omega}. \tag{6.66}$$

Also note that from (6.64),

$$\frac{1}{k}\|\mathfrak{z}\|_{L^2(\Lambda_l),\omega} = \|p\|_{L^2(\Lambda_l),\omega}. \tag{6.67}$$

Substituting (6.67) into (6.66), we deduce

$$\sum_{m=0}^{d} |\alpha_m|^2 \leq \frac{2}{k\pi}\|\mathfrak{z}\|^2_{L^2(\Lambda_l),\omega}. \tag{6.68}$$

Now using Cauchy-Schwarz inequality and (6.68) we deduce

$$\begin{aligned}\sum_{m=0}^{d}|\alpha_m| &\leq \left\{\sum_{m=0}^{d}|\alpha_m|^2\right\}^{1/2}\left\{\sum_{m=0}^{d}|1|^2\right\}^{1/2}\\ &\leq C\left(\frac{d}{k}\right)^{1/2}\|\mathfrak{z}\|_{L^2(\Lambda_l),\omega}. \end{aligned}\tag{6.69}$$

From (6.63), using Lemma 6.10, we obtain

$$\begin{aligned}|a_k(\mathfrak{z},\mathfrak{w})-\tilde{a}_k(\mathfrak{z},\mathfrak{w})| &\leq C_1\varepsilon\frac{d}{k}\|\mathfrak{z}\|_{L^2(\Lambda_l),\omega}\|\mathfrak{w}\|_{L^2(\Lambda_j),\omega}\\ &\leq C_2\varepsilon\frac{d^2}{k}\|\mathfrak{z}\|_{L^2(\Lambda_l)}\|\mathfrak{w}\|_{L^2(\Lambda_j)}. \end{aligned}\tag{6.70}$$

Finally, since the result (6.70) holds for all l and $j=1,2,3$, by substituting (6.70) in (6.62), we obtain the result of the lemma,

$$|a_k(z,w)-\tilde{a}_k(z,w)| \leq C\varepsilon\frac{d^2}{k}\|z\|\|w\|. \tag{6.71}$$

□

In the following lemma, we determine the bounds on continuity and coercivity constants for the discrete sesquilinear form $\tilde{a}_k$.

Lemma 6.14. *The discrete sesquilinear form $\tilde{a}_k$ is coercive and continuous,*

$$\begin{aligned}\tilde{a}_k(v,v) &\geq \tilde{\alpha}_k||v||^2, \quad \forall v\in\mathbb{V}_k^{\boldsymbol{d}}\\ |\tilde{a}_k(u,v)| &\leq \tilde{B}_k||u||||v||, \quad \forall u,v\in\mathbb{V}_k^{\boldsymbol{d}},\end{aligned}$$

with $\boldsymbol{d}=(d,d,d)^T$ and

$$\tilde{B}_k \leq B_k + C\varepsilon\frac{d^2}{k}, \tag{6.72}$$

$$\tilde{\alpha}_k \geq \alpha_k - C\varepsilon\frac{d^2}{k} > 0, \tag{6.73}$$

for sufficiently large[2] k, where ε satisfies (6.60). The constants α_k and B_k denote the coercivity and continuity constants, respectively, of the sesquilinear form a_k as defined in

[2] in fact we require

$$k \geq C\frac{d^2\varepsilon}{\alpha_k-\delta}, \quad \text{for some } \delta>0.$$

(2.25),

$$\begin{aligned} a_k(v,v) &\geq \alpha_k \|v\|^2 \qquad \forall v \in L^2(\Gamma), \qquad \textit{coercivity}, \\ |a_k(u,v)| &\leq B_k \|u\|\|v\| \qquad \forall u,v \in L^2(\Gamma), \qquad \textit{continuity}. \end{aligned} \tag{6.74}$$

Proof. From Lemma 6.13 we know that for all $z, w \in \mathbb{V}_k^{\boldsymbol{d}}$,

$$|a_k(z,w) - \tilde{a}_k(z,w)| \leq C\varepsilon \frac{d^2}{k} \|z\|\|w\|.$$

Then

$$a_k(z,w) - C\varepsilon \frac{d^2}{k} \|z\|\|w\| \leq \tilde{a}_k(z,w) \leq a(z,w) + C\varepsilon \frac{d^2}{k} \|z\|\|w\|. \tag{6.75}$$

The left-hand inequality in (6.75) yields

$$\tilde{a}_k(z,z) \geq \alpha_k \|z\|^2 - C\varepsilon \frac{d^2}{k} \|z\|^2.$$

This leads to the bound on the coercivity constant for the discrete sesquilinear form, $\tilde{a}_k$. We deduce that from the right-hand inequality in (6.75),

$$|\tilde{a}_k(z,w)| \leq B_k \|z\|\|w\| + C\varepsilon \frac{d^2}{k} \|z\|\|w\|.$$

Thus the result for the continuity follows. □

Theorem 6.15 (***Error of the fully-discrete Galerkin method in terms of*** ε)**.** *Let* $\tilde{\tilde{v}} \in \mathbb{V}_k^{\boldsymbol{d}}$ *be the solution of the fully-discrete Galerkin method with* Λ_j, $j = 1,2,3$ *defined (6.15). Let* ε *be defined as in (6.60). Let* $\boldsymbol{d}$ *in* $\mathbb{V}_k^{\boldsymbol{d}}$ *equal to* $\boldsymbol{d} = (d,d,d)$*. Then, for all* $n \geq 6$ *with* $n \leq \max\{d_I, d_T\} + 1$*,*

$$\begin{aligned} \|v - \tilde{\tilde{v}}\| &\leq C_n \left(1 + \frac{\tilde{B}_k}{\tilde{\alpha}_k} + \frac{1}{\tilde{\alpha}_k} C\varepsilon d^2 k^{-1}\right) \left(k^{1/3} \left(\frac{k^{1/9}}{d}\right)^n + \exp(-c_0 k^\delta)\right) \\ &+ C' \frac{1}{\tilde{\alpha}_k} \varepsilon d^2, \end{aligned} \tag{6.76}$$

where C_n *and* $C' > 0$ *are constants independent of* k*.*

Proof. We will prove the theorem for the case when $d_I = d_T = d$. The proof for other cases follows analogously. Using Lemma 6.13, we simplify the estimate (6.49) as follows,

$$\|v - \tilde{\tilde{v}}\| \leq \inf_{z \in \mathbb{V}_k^{\boldsymbol{d}}} \left\{ \left(1 + \frac{\tilde{B}_k}{\tilde{\alpha}_k}\right) \|v - z\| + \frac{1}{\tilde{\alpha}_k} C\varepsilon d^2 k^{-1} \|z\| \right\}. \tag{6.77}$$

Moreover, since

$$\|z\| \leq \|v - z\| + \|v\|,$$

we write (6.77) as follows,

$$\begin{aligned}\|v-\tilde{\tilde{v}}\| &\leq \inf_{z\in\mathbb{V}_k^{\boldsymbol{d}}}\left\{\left(1+\frac{\tilde{B}_k}{\tilde{\alpha}_k}\right)\|v-z\|+\frac{1}{\tilde{\alpha}_k}C\varepsilon d^2k^{-1}\left(\|v-z\|+\|v\|\right)\right\}\\ &\leq \inf_{z\in\mathbb{V}_k^{\boldsymbol{d}}}\left\{\left(1+\frac{\tilde{B}_k}{\tilde{\alpha}_k}+\frac{1}{\tilde{\alpha}_k}C\varepsilon d^2k^{-1}\right)\|v-z\|+\frac{1}{\tilde{\alpha}_k}C\varepsilon d^2k^{-1}\|v\|\right\}.\end{aligned}$$

Using (6.56) and (6.57), we write

$$\begin{aligned}\|v-z\| &\leq k\sum_{j=1}^{3}\|V_j(\cdot,k)-p_j\|_{L^2(\Lambda_j)}+\|v\|_{L^2(\Lambda_4)},\\ \|v\| &\leq \sum_{j=1}^{4}\|v\|_{L^2(\Lambda_j)},\end{aligned}$$

where $V_j(s,k)=\chi_j(s)V(s,k)$. Then

$$\begin{aligned}\|v-\tilde{\tilde{v}}\| &\leq \left(1+\frac{\tilde{B}_k}{\tilde{\alpha}_k}+\frac{1}{\tilde{\alpha}_k}C\varepsilon d^2k^{-1}\right)\left\{k\sum_{j=1}^{3}\inf_{p_j\in\mathbb{P}_d}\|V_j(\cdot,k)-p_j\|_{L^2(\Lambda_j)}+\|v\|_{L^2(\Lambda_4)}\right\}\\ &+ \frac{1}{\tilde{\alpha}_k}C\varepsilon d^2k^{-1}\sum_{j=1}^{3}\|v\|_{L^2(\Lambda_j)}.\end{aligned}$$

Using the estimates on the best polynomial approximation of V_j, $j=1,2,3$ from Theorem 6.2 and Theorem 6.3, we deduce that for $6\leq n\leq d+1$, there exist k-independent constants $C_n>0$ such that

$$\begin{aligned}\|v-\tilde{\tilde{v}}\| &\leq C_n\left(1+\frac{\tilde{B}_k}{\tilde{\alpha}_k}+\frac{1}{\tilde{\alpha}_k}C\varepsilon d^2k^{-1}\right)\left\{k^{1/3}\left(\frac{k^{1/9}}{d}\right)^n+\|v\|_{L^2(\Lambda_4)}\right\}\\ &+ \frac{1}{\tilde{\alpha}_k}C\varepsilon d^2k^{-1}\sum_{j=1}^{3}\|v\|_{L^2(\Lambda_j)}.\end{aligned}\tag{6.78}$$

Now, note that

$$\|v\|^2_{L^2(\Lambda_j)}:=k\int_{\Lambda_j}|V(s,k)|^2ds.\tag{6.79}$$

From Theorem 3.6, we know that for $j=1,2,3$

$$|V(s,k)|\leq C_0',\quad \text{for all}\ \ s\in\Lambda_j,$$

where $C_0'>0$ is a constant independent of k and s. Thus, we deduce from (6.79), for

$j = 1, 2, 3,$

$$\|v\|_{L^2(\Lambda_j)} \leq C_0 k, \tag{6.80}$$

where $C_0 > 0$ is a constant independent of k. Substituting (6.80) into (6.78), we obtain

$$\|v - \tilde{\tilde{v}}\| \leq C_n \left(1 + \frac{\tilde{B}_k}{\tilde{\alpha}_k} + \frac{1}{\tilde{\alpha}_k} C\varepsilon d^2 k^{-1}\right) \left(k^{1/3} \left(\frac{k^{1/9}}{d}\right)^n + \|v\|_{L^2(\Lambda_4)}\right) + C' \frac{1}{\tilde{\alpha}_k} \varepsilon d^2,$$

where $C' > 0$ is also a constant independent of k.

Finally, using the estimate of $\|v\|_{L^2(\Lambda_4)}$ from Theorem 6.4, we obtain the result (6.82). □

Proposition 6.16. *Under the assumptions of Theorem 6.15, if for* $6 \leq n \leq d + 1$,

$$\varepsilon \leq C B_k k^{1/3} d^{-2} \left(\frac{k^{1/9}}{d}\right)^n, \tag{6.81}$$

then

$$\|v - \tilde{\tilde{v}}\| \leq C_n \left(1 + \frac{B_k}{\alpha_k}\right) \left\{k^{1/3} \left(\frac{k^{1/9}}{d}\right)^n + \exp\left(-c_0 k^\delta\right)\right\}. \tag{6.82}$$

where $C_n > 0$ *is a constant independent of* k.

Proof. We know that

$$\tilde{\alpha}_k \geq \alpha_k - C\varepsilon d^2 k^{-1}. \tag{6.83}$$

For sufficiently large d in (6.81), ε can be made sufficiently small to satisfy

$$\tilde{\alpha_k} \geq c_1 \alpha_k, \tag{6.84}$$

where c_1 is a constant satisfying[3] $1/4 < c_1 < 1/2$.

Moreover, from Lemma 6.14, we deduce

$$1 + \frac{\tilde{B}_k}{\tilde{\alpha}_k} \leq 1 + \frac{B_k + C\varepsilon d^2 k^{-1}}{\alpha_k - C\varepsilon d^2 k^{-1}} \leq \frac{\alpha_k + B_k}{\tilde{\alpha}_k}.$$

[3] In order to satisfy (6.84), the parameter ε must satisfy

$$\varepsilon \leq \frac{\alpha_k (1 - c_1) k}{2Cd^2}, \tag{6.85}$$

where C is a constant appearing in (6.83). Comparing the bound with (6.81), we are satisfied that for sufficiently large n, (6.85) holds.

Now, using (6.81), we deduce

$$\begin{aligned} 1+\frac{\tilde{B}_k}{\tilde{\alpha}_k}+\frac{1}{\tilde{\alpha}_k}C\varepsilon d^2k^{-1} &\leq \frac{\alpha_k+B_k}{\tilde{\alpha}_k}+C\frac{B_k}{\tilde{\alpha}_k}k^{-2/3}\left(\frac{k^{1/9}}{d}\right)^n \\ &\leq \frac{1}{c_1}\left(1+\frac{B_k}{\alpha_k}\right)+C\frac{B_k}{\tilde{\alpha}_k}k^{-2/3}\left(\frac{k^{1/9}}{d}\right)^n \\ &\leq \frac{c_2}{c_1}\left(1+\frac{B_k}{\alpha_k}\right), \end{aligned}$$

with $c_2 > 1$.

Furthermore, we estimate the second term in (6.82) as follows,

$$\begin{aligned} C'\frac{1}{\tilde{\alpha}_k}\varepsilon d^2 &\leq \tilde{C}\frac{B_k}{\tilde{\alpha}_k}k^{1/3}\left(\frac{k^{1/9}}{d}\right)^n \\ &\leq \tilde{C}'\left(1+\frac{B_k}{\alpha_k}\right)k^{1/3}\left(\frac{k^{1/9}}{d}\right)^n. \end{aligned}$$

Hence, the result follows from Theorem 6.15. □

Corollary 6.17. *Let $v \in L^2([0,2\pi])$ be the solution of (6.2) and let $\tilde{\tilde{v}} \in \mathbb{V}_k^{\boldsymbol{d}}$ be the fully-discrete solution of the system of equations (6.48) with the approximation space $\mathbb{V}_k^{\boldsymbol{d}}$ defined in (6.5), with $\boldsymbol{d} = (d,d,d)$ and Λ_j, $j = 1,2,3$ defined in (6.15). Let ε be defined as in (6.60). Provided the parameters of the composite Filon-Clenshaw-Curtis quadrature are chosen to satisfy:*

$$N \geq d+2, \qquad M \geq dk^{-1/9}, \qquad q \geq (N+1)/(1+\beta), \tag{6.86}$$

then for all $n \geq 6$ with $n \leq d+1$, the following holds

$$\|v-\tilde{\tilde{v}}\| \leq C_n\left(1+\frac{B_k}{\alpha_k}\right)\left\{k^{1/3}\left(\frac{k^{1/9}}{d}\right)^n+\exp\left(-c_0k^{\delta}\right)\right\}, \tag{6.87}$$

where δ is defined in Theorem 3.3, equation (3.9).

Proof. From Proposition 6.16, for (6.87) to hold, we require ε to satisfy (6.81)

$$\varepsilon \leq C\frac{k^{1/3}}{d^2}\left(\frac{k^{1/9}}{d}\right)^n = Ck^{1/9}\left(\frac{k^{1/2}}{d}\right)^2\left(\frac{k^{1/9}}{d}\right)^n.$$

From (6.47), it follows that

$$|J_k-\tilde{J}_k| \leq C\left(\frac{1}{M}\right)^{N+1}.$$

Choosing $N > d + 2$ and $M > dk^{-1/9}$ yields

$$\varepsilon \backsim |J_k - \tilde{J}_k| \leq C \left(\frac{k^{1/9}}{d} \right)^{N+1} . \tag{6.88}$$

Hence (6.81) is satisfied. □

Therefore, provided appropriate choices of the parameters of the composite Filon-Clenshaw-Curtis method are made, the error of the fully-discrete Galerkin method remains the same as the error of the semi-discrete Galerkin method. In other words, the accuracy of the fully-discrete Galerkin method does not deteriorate due to inexact approximation of the entries of the stiffness matrix.

6.5 Numerical examples

In this section, we discuss the numerical results obtained for the fully-discrete Galerkin method.

Example 1.

Consider a problem of scattering by an ellipse illustrated in Figure 6.1 by a plane wave $u^I(\boldsymbol{x})$ travelling in the direction $\boldsymbol{a} = (1,0)^T$. In the numerical experiments below, we have taken

$$\Lambda_1 = [t_1 - c_1 k^{-1/3}, t_1 + c_2 k^{-2/9}],\ \Lambda_2 = [t_1 + c_2 k^{-2/9}, t_2 - c_2 k^{-2/9}],$$
$$\Lambda_3 = [t_2 - c_2 k^{-2/9}, t_2 + c_1 k^{-1/3}],\ \Lambda_4 = [t_2 + c_1 k^{-1/3}, 2\pi] \cup [0, t_1 - c_1 k^{-1/3}],$$

with the constants c_1 and c_2 chosen as

$$c_1 = 0.506L, \qquad c_2 = 0.366L, \tag{6.89}$$

where L denotes the circumference of an ellipse in Figure 6.1.

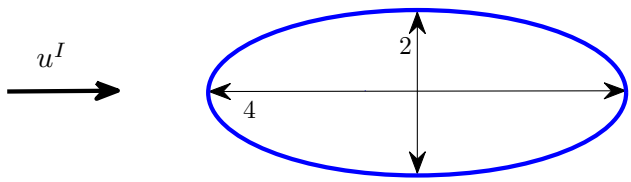


Figure 6.1: *In the Example 1, we consider scattering by an ellipse with major axis greater than the minor axis. The incident wavefield, u^I, is propagating in the direction of the vector $(1,0)^T$.*

The vector $\boldsymbol{d}$ in $\mathbb{V}_k^{\boldsymbol{d}}$ is equal to $\boldsymbol{d} = (d,d,d)^T$. The parameters q, N and M of the composite FCC quadrature used for the assembly of the stiffness matrix, are fixed,

$$q = 8, \quad N = 8, \quad M = 20, \tag{6.90}$$

and parameters k and d vary. The parameters q, N and M in (6.90) satisfy the conditions (6.86).

In Table 6.1 we display the relative errors of the fully-discrete Galerkin approximation for the cases of moderate k, defined as follows

$$e_d = \frac{\|v - \tilde{\tilde{v}}\|_{L^2([0,2\pi])}}{\|v\|_{L^2([0,2\pi])}}, \tag{6.91}$$

where the "exact" solution v is obtained via the Nystrom method [34, Section 3.5, pp66-77]. The relative errors were computed using Simpson's rule with a sufficient number of quadrature points.

Since the integration parameters q, N, M are fixed, the cpu time[4] for the computation of the fully-discrete solution remains the same for fixed d and varying k. We observe from the table that as k increases, the errors also increase but with much slower rate as the increase in k. This is consistent with Theorem 6.17. Furthermore, for sufficiently large d, the errors decrease as k grows which is a better result than the theory predicts.

Standard					Star-combined				
d	$k = 200$	$k = 400$	$k = 600$	cpu time	d	$k = 200$	$k = 400$	$k = 600$	cpu time
4	1.47(-2)	3.09(-2)	3.49(-2)	84	4	1.62(-2)	3.54(-2)	4.09(-2)	118
8	2.18(-3)	4.16(-3)	4.99(-3)	101	8	2.91(-3)	5.83(-3)	6.24(-3)	136
12	1.53(-3)	6.83(-4)	7.87(-4)	135	12	3.04(-3)	9.85(-4)	9.97(-4)	180
16	2.25(-3)	6.59(-4)	3.73(-4)	184	16	2.72(-3)	8.36(-4)	6.97(-4)	238

Table 6.1: *The errors e_d defined in (6.91) of the fully-discrete Galerkin method for solving the standard and star-combined integral equation for low/moderate k.*

The errors in Table 6.1 are presented for the cases of standard and star-combined formulations. In the star-combined case, the cpu time is slightly higher than the cpu time for the case of standard formulation. This is due mainly to the additional computation of the highly-oscillatory one-dimensional integrals, (6.44c) and (6.44d). However, as in case of the standard formulation, the cpu time remains to be fixed for growing k and only grows with increasing d. Moreover, the growth rate of the cpu time with increasing d in the case of star-combined formulation is the same as in the case of standard combined formulation, see Table 6.2.

Standard			Star-combined		
d	cpu time	ratio	d	cpu time	ratio
4	84		4	118	
8	101	1.20	8	136	1.15
12	135	1.33	12	180	1.32
16	184	1.36	16	238	1.32

Table 6.2: *The rate of growth in cpu time with growing d.*

In Table 6.3, we display the ratios of errors e_d/e_{d+4}. We conjecture that the errors in the Table 6.1 decrease exponentially with growing d. In other words, we expect

$$e_d \sim e^{-cd},$$

[4]The CPU time was monitored using *tic* and *toc* commands in MATLAB run on Intel Premium Dual CPU E2160@ 1.80 GHz 1.79 GHz, 1.98 GB RAM.

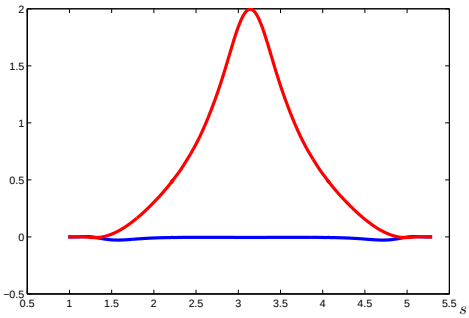


Figure 6.2: *The plot of the computed approximation of the function $V(\cdot, 600)$. The real part of the solution is plotted in blue and the imaginary part is plotted in red.*

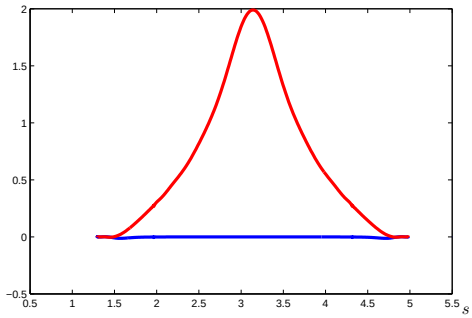


Figure 6.3: *The plot of the computed approximation of the function $V(\cdot, 6000)$. The real part of the solution is plotted in blue and the imaginary part is plotted in red.*

with some constant $c > 0$ independent of d. Then,

$$\frac{e_d}{e_{d+4}} = \frac{e^{-cd}}{e^{-c(d+4)}} = e^{c4} = const.$$

We observe from the first two rows of Table 6.3 that for fixed k, the errors indeed decay exponentially.

Standard			Star-combined		
ratio	$k = 400$	$k = 600$	ratio	$k = 400$	$k = 600$
e_4/e_8	7.43	6.99	e_4/e_8	6.07	6.55
e_8/e_{12}	6.09	6.43	e_8/e_{12}	5.92	6.26
e_{12}/e_{16}	1.03	2.11	e_{12}/e_{16}	1.18	1.43

Table 6.3: *The ratios of the errors of the fully-discrete Galerkin method for solving the standard and star-combined integral equations.*

The numerical approximation of $V(\cdot, k)$ for $k = 600$ is plotted in Figure 6.2.

Example 2.

Let us now consider the fully-discrete approximation for the cases of large values of k with Λ_j, $j = 1, 2, 3, 4$ and the parameters M, q and N chosen as in Example 1. In Table 6.4, we display the errors

$$e_d = \frac{\|v - \tilde{\tilde{v}}\|_{L^2([0,2\pi])}}{\|v\|_{L^2([0,2\pi])}}, \tag{6.92}$$

where the "exact" solution, v, is the solution of the fully-discrete Galerkin method with $d = 20$.

The cpu time remains the same, since the parameters for the numerical approximation of the entries of the stiffness matrix are fixed at the same values (6.90) as in the previous

example. Again, a k increases, the errors also increase but with much slower rate as the increase in k. We also observe, as in Example 1, for sufficiently large d, the errors decrease as k grows.

Standard				Star-combined			
d	$k = 4000$	$k = 8000$	cpu time	d	$k = 4000$	$k = 8000$	cpu time
4	8.931(-3)	3.780(-2)	84	4	3.796(-2)	7.862(-2)	118
8	1.783(-3)	6.011(-3)	101	8	5.368(-3)	8.320(-3)	136
12	6.625(-4)	6.281(-4)	135	12	8.051(-4)	7.287(-4)	180
16	7.088(-4)	1.927(-5)	184	16	9.603(-4)	2.685(-5)	238

Table 6.4: *The errors e_d defined in (6.92) of the fully-discrete Galerkin method for solving the standard and star-combined integral equations for large k.*

The numerical approximation of $V(\cdot, k)$ for $k = 6000$ is plotted in Figure 6.2.

Example 3

In this example, we consider the scattering by an ellipse illustrated in Figure 6.4.

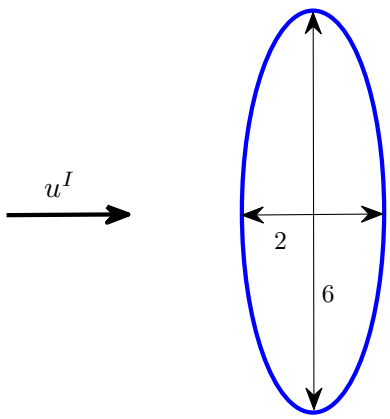


Figure 6.4: *The function u^I denotes the incident wavefield propagating in the direction of the vector $(1, 0)^T$ and scattering off the ellipse with the major axis smaller that the minor axis.*

Again, the parameters q, N and M for integration on each pair of the domains of the boundary are fixed as in (6.90), and parameters k and d vary. The constants c_1 and c_2 in (6.15) are chosen as follows:

$$\begin{aligned} c_1 &= 0.085L, \\ c_2 &= 0.057L, \end{aligned}$$

where L denotes the circumference of an ellipse in Figure 6.4. Note that the constants c_1 and c_2 are not the same as in the previous example (6.89) where we have considered the ellipse with the major axis bigger that the minor axis. Were the constants in this

example to be chosen as in (6.89), the numerical results would have shown much poorer convergence rates than in Example 1.

In Table 6.5, we display the errors

$$e_d = \frac{\|v - \tilde{\tilde{v}}\|_{L^2([0,2\pi])}}{\|v\|_{L^2([0,2\pi])}}$$

for fixed $d = 12$ and $k = 400$ and varying c_1 and c_2 for two types of ellipses. The first ellipse has curvature $\kappa(t) = 9$ at the transition points $t = t_1$ and $t = t_2$, see Figure 6.4; while the second ellipse has curvature $\kappa(t) = 1/3$ at the transition points, see Figure 6.1.

From Table 6.5, we observe that the accuracy of the fully-discrete Galerkin method is sensitive to the choice of the constants c_1 and c_2. In particular, in order to maintain the accuracy in both cases, the constants c_1 and c_2 must depend on the curvature at the transition points.

$a = 1,\ b = 3$ $\kappa(t_1) = \kappa(t_2) = 9$		$a = 3,\ b = 1$ $\kappa(t_1) = \kappa(t_2) = 1/3$	
c_1 and c_2	$k = 400$	c_1 and c_2	$k = 400$
$c_1 = 0.026L$ and $c_2 = 0.019L$	$6.23(-3)$	$c_1 = 0.851L$ and $c_2 = 0.637L$	$1.68(-3)$
$c_1 = 0.052L$ and $c_2 = 0.038L$	$2.29(-3)$	$c_1 = 1.019L$ and $c_2 = 0.731L$	$7.21(-4)$
$c_1 = 0.105L$ and $c_2 = 0.076L$	$9.79(-4)$	$c_1 = 1.019L$ and $c_2 = 0.731L$	$1.03(-3)$

Table 6.5: *The errors in the approximation of $V(\cdot, k)$ for two different type of ellipses: on the left, the results are displayed for ellipse with curvature $\kappa(t) = 9$ at the transition points (right columns) and the results on the right correspond to the ellipse with curvature $\kappa(t) = 1/3$.*

This can be explained by looking at the asymptotic expansion of $V(\cdot, k)$ in the transition zones given in (3.5). The expansion is given in terms of the derivatives of the Fock's integral $\Psi^{(l)}\left(k^{1/3}Z(s)\right)$, where $Z(s)$ is defined in (3.135),

$$Z(s) = 2^{-1/3}\int_{\boldsymbol{x}_0}^{\boldsymbol{x}} \kappa(s)^{2/3} ds,$$

where $\kappa(s)$ denotes the curvature of the boundary Γ at $\boldsymbol{\gamma}(s)$. Using the properties of the Fock's integral at large values of the argument, the k - dependent estimates of the derivatives of V were be derived, see Theorem 3.6. However, the expansion (3.5) suggests that the derivatives of V also depend on the curvature of the boundary in the transition zones as follows, see Theorem 3.6,

$$|D_s^n V(s,k)| \leq C_n \begin{cases} 1, & n = 0, 1 \\ k^{-1}(k^{1/3} + \kappa(s)|w(s)|)^{-n-2}, & n \geq 2, \end{cases}, \tag{6.93}$$

where $w(s) = (s - t_1)(t_2 - s)$.

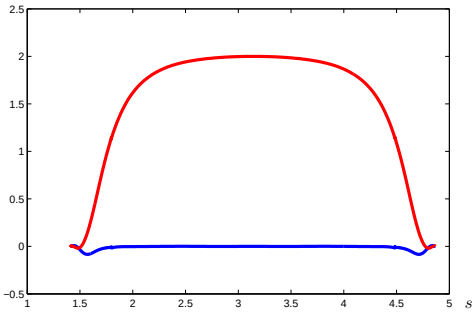


Figure 6.5: *The plot of the computed approximation of the function $V(\cdot, 600)$. The real part of the solution is plotted in blue and the imaginary part is plotted in red.*

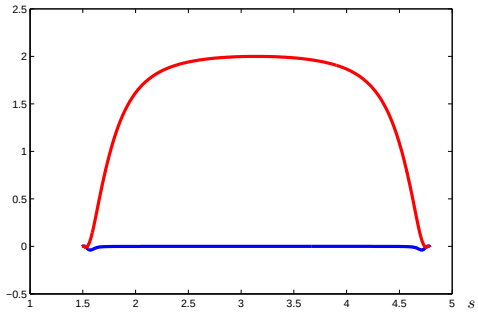


Figure 6.6: *The plot of the computed approximation of the function $V(\cdot, 6000)$. The real part of the solution is plotted in blue and the imaginary part is plotted in red.*

The curvature at the transition point t_1 on the ellipse in Figure 6.4 is substantially smaller than the curvature at t_1 for the ellipse in Figure 6.4. Therefore, it is reasonable to choose transition intervals Λ_1 and Λ_3 in this example to be smaller than those in Example 1.

In Table 6.6 we display the errors of the fully-discrete Galerkin method for the cases of moderate k, defined as before

$$e_d = \frac{\|v - \tilde{v}\|}{\|v\|}, \tag{6.94}$$

where the "exact" solution v is obtained via the Nystrom method [34]. The relative errors were computed using Simpson's rule with a sufficient number of quadrature points.

From Table 6.4, we observe similar convergence rates as in Example 1: as k increases, the errors also increase but with much slower rate as the increase in k. Also, for sufficiently large d, errors decrease as k grows.

Standard					Star-combined				
d	$k = 200$	$k = 400$	$k = 600$	cpu time	d	$k = 200$	$k = 400$	$k = 600$	cpu time
4	1.04(-2)	2.07(-2)	2.95(-2)	84	4	2.58(-2)	4.32(-2)	3.49(-2)	118
8	2.41(-3)	3.91(-3)	4.86(-3)	101	8	5.02(-3)	7.67(-3)	5.91(-3)	136
12	1.87(-3)	9.97(-4)	9.36(-4)	135	12	4.21(-3)	1.51(-3)	9.82(-4)	180
16	2.01(-3)	1.23(-3)	5.82(-4)	184	16	6.07(-3)	1.06(-3)	7.86(-4)	238

Table 6.6: *The errors e_d defined in (6.91) of the fully-discrete Galerkin method for solving the standard and star-combined integral equation for low/moderate k.*

In Table 6.7, we display the ratios of errors e_d/e_{d+4}. As we observe from the Table 6.3, the errors decay exponentially as d grows until the error in the shadow part of the boundary becomes dominant.

The numerical approximation of $V(\cdot, k)$ for $k = 600$ and $k = 6000$ are plotted in Figure 6.5 and Figure 6.6 respectively.

Standard			Star-combined		
ratio	$k = 400$	$k = 600$	ratio	$k = 400$	$k = 600$
e_4/e_8	5.29	6.07	e_4/e_8	5.63	5.91
e_8/e_{12}	3.92	5.19	e_8/e_{12}	5.08	6.02
e_{12}/e_{16}	0.81	1.69	e_{12}/e_{16}	1.42	1.21

Table 6.7: *The ratios of the errors of the fully-discrete Galerkin method for solving the standard and star-combined integral equations.*

Example 4

Let us now compare the condition numbers of the stiffness matrices corresponding to the standard-combined and the star-combined formulations. Consider the problem of scattering by an ellipse illustrated in Figure 6.1. Let Λ_j, $j = 1, 2, 3, 4$ be chosen as in Example 1.

In Table 6.8, we display the condition numbers, $\text{cond}_d(k)$ of the stiffness matrices, entries of which are approximated by the numerical integration method described in Chapter 4. The condition numbers are obtained using the command ***cond*** in MATLAB. Also in Table 6.8, we display the ratios

$$\text{ratio} = \frac{\text{cond}_{d+4}(k)}{\text{cond}_d(k)},$$

between the condition numbers for growing d. We observe from the table that the the rate of growth of the condition numbers with increasing d, corresponding to standard and star-combined formulations, are similar.

Standard					Star-combined				
d	$k = 600$	ratio	$k = 6000$	ratio	d	$k = 600$	ratio	$k = 6000$	ratio
4	38.5		27.9		4	33.2		22.7	
8	74.9	1.94	59.6	2.14	8	74.9	2.25	56.6	2.49
12	121.1	1.61	95.9	1.61	12	144.3	1.93	117.5	2.08
16	187.5	1.54	139.7	1.46	16	240.4	1.67	203.1	1.73

Table 6.8: *Condition numbers for star-combined integral equation.*

In Table 6.9, we display the ratios

$$r(k) = \frac{\text{cond}_d(k)}{\text{cond}_d(10k)},$$

between the condition numbers for $k = 600$ and the condition numbers for $k = 6000$. From Table 6.9, we observe that the condition numbers grow very slowly (if at all) with increasing k.

In Figures 6.7 to 6.10, we plot the absolute values of the entries of the stiffness matrices

Formulation	Ratios	$d = 8$	$d = 12$
Standard	r(600)	1.06	1.14
	r(1200)	0.96	1.00
	r(2400)	0.97	1.02
Star	r(600)	1.06	1.04
	r(1200)	1.11	1.09
	r(2400)	1.10	1.11

Table 6.9: *Ratios* $\mathrm{cond}_d(600)/\mathrm{cond}_d(6000)$*, between the condition numbers for* $k = 600$ *and* $k = 6000$.

$\boldsymbol{R}$ and $\boldsymbol{A}$ corresponding to the standard and the star-combined formulations defined in (6.25) and (6.34) respectively with $k = 600$ and $k = 6000$. We observe from the plots that with increasing k, the absolute values of the diagonal entries also increase, while the off-diagonal entries remain relatively unchanged. The stiffness matrices are of the dimension $3(d + 1)$ with $d = 16$.

The absolute values of the off-diagonal elements are bigger than the upper bound on ε in (6.82). Therefore, by Proposition 6.16, in order to preserve the error estimates of the semi-Galerkin method, the entries cannot be approximated by zero.

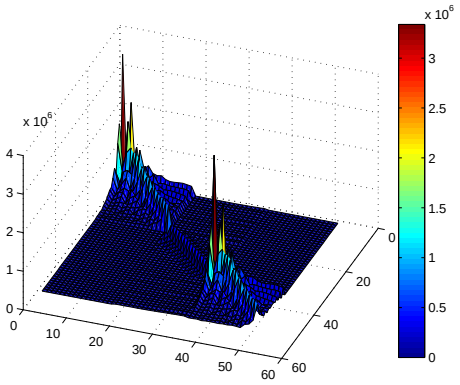


Figure 6.7: *The plot of absolute values of the entries of the stiffness matrix for the case when* $k = 600$ *corresponding to the standard combined formulation.*

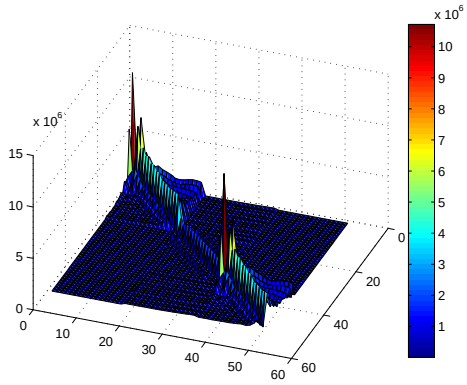


Figure 6.8: *The plot of absolute values of the entries of the stiffness matrix for the case when* $k = 600$ *corresponding to the star-combined formulation.*

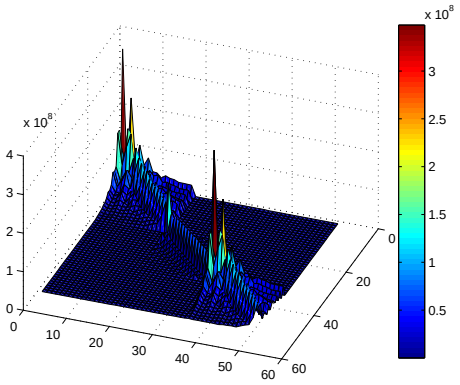


Figure 6.9: *The plot of absolute values of the entries of the stiffness matrix for the case when* $k = 6000$ *corresponding to the standard combined formulation.*

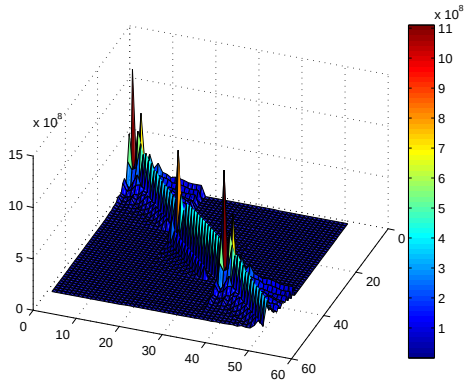


Figure 6.10: *The plot of absolute values of the entries of the stiffness matrix for the case when* $k = 6000$ *corresponding to the stan-combined formulation.*

Bibliography

[1] T. Abboud, J.-C. Nédélec, and B. Zhou. Méthodes des équations intégrales pour les hautes fréquences. *Comptes Rendus Mathématique. Académie des Sciences. Paris*, 164, 1994.

[2] T. Abboud, J.-C. Nédélec, and B. Zhou. Improvement of the integral equation method for high frequency problems. *Third International Conference on Mathematical Aspects of Wave Propagation Phenomena, SIAM*, pages 178–187, 1995.

[3] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, 1964.

[4] V. M. Babich. On shortwave asymptotics of Green's function for the Helmholtz eqution (in russian). *Matematicheskii sbornik*, 65 (107), number 4:576–630, 1964.

[5] V. M. Babich and V. S. Buldyrev. *Sort-wavelength diffraction theory. Asymptotic methods.* Springer-Verlag. Springer series on wave phenomena, 1990.

[6] V. M. Babich, D. B. Dement'ev, B. A. Samokish, and V. P. Smyshlyaev. On evaluation of the diffraction coefficients for arbitrary "nonsingular" directions of a smooth convex cone. *SIAM Journal on Applied Mathematics*, 60(2):536–573, 1999.

[7] I. Babuška, F. Ihlenburg, T. Strouboulis, and S. K. Gangaraj. A posteriori error estimation for finite element solutions of Helmholtz equation. Part I: the quality of local indicators and estimators. *International Journal for Numerical Methods in Engineering*, 40, Issue 18:3443–3462, 1997.

[8] I. Babuška, F. Ihlenburg, T. Strouboulis, and S. K. Gangaraj. A posteriori error estimation for finite element solutions of Helmholtz equation. Part II: estimation of the pollution error. *International Journal for Numerical Methods in Engineering*, 40:3883–3900, 1997.

[9] I. Babuška and F. Ihlenburg. Dispersion analysis and error estimation of Galerkin finite element methods for the Helmholtz equation. *International Journal for Numerical Methods in Engineering*, 38, Issue 22:3745–3774, 1995.

[10] I. Babuška and J. M. Melenk. The partition of unity method. *International Journal for Numerical Methods in Engineering*, 40, Issue 4:727–758, 1997.

[11] E. T. Bell. Partition polynomials. *Annals of Mathematics*, 29 (1/4):3846, 1928.

[12] T. Betcke, S. N. Chandler-Wilde, I. G. Graham, S. Langdon, and M. Lindner. Condition number estimates for combined potential integral operators in acoustics and their boundary element discretisation. *Numerical methods for partial differential equations*, 27:31–69, 2011.

[13] T. Betcke and E. A. Spence. Numerical estimation of coercivity constants for boundary integral operators in acoustic scattering. *SIAM Journal on Numerical Analysis*, 49 (4):1572–1601, 2011.

[14] N. Bleistein. *Mathematical Methods for wave phenomena.* Academic Press, 1984.

[15] B. D. Bonner, I. G. Graham, and V. P. Smyshlyaev. The computation of conical diffraction coefficients in high-frequency acoustic wave scattering. *SIAM Journal on Numerical Analysis*, 43:1202–1230, March 2005.

[16] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi. *Electromagnetic and acoustic scattering by simple shapes.* New York, Hemisphere Publishing Corp., 1987.

[17] H. Brakhage and P. Werner. Uber das dirichletsche aussenraumproblem fur die Helmholtzsche schwingungsgleichung. *Archiv der Mathematik*, 16:325 – 329, 1965.

[18] O. Bruno, O. Reitich, and C. Geuzaine. On the o(1) solution of multiple-scattering problems. *IEEE Transactions on Magnetics*, 41:1488–1491, 2005.

[19] O. P. Bruno and C. A. Geuzaine. An o(1) integration scheme for threedimensional surface scattering problems. *Journal of Computational and Applied Mathematics*, 204:463–476, 2007.

[20] O. P. Bruno, C. A. Geuzaine, J. A. M. Jr., and F. Reitich. Prescribed error tolerances within fixed computational times for scattering problems of arbitrarily high frequency: the convex case. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 362:629–645, 2004.

[21] O. P. Bruno and L. A. Kunyansky. Surface scattering in three dimensions: An accelerated high-order solver. *Proceedings: Mathematical, Physical and Engineering Sciences*, 457(2016):pp. 2921–2934, 2001.

[22] A. Buffa and P. Monk. Error estimates for the Ultra Weak Variational Formulation of the Helmholtz equation. *ESIAM: Mathematical modelling and Numerical analysis*, Volume 42, Number 6:925–940, 2008.

[23] V. S. Buslaev. Short wave asymptotic behavior in the problem of diffraction by smooth convex contours. (in russian). *Trudy Matematicheskogo Instituta imeni Steklova*, 73:14–117, 1964.

[24] O. Cessenat and B. Despres. Application of an Ultra Weak Variational Formulation of elliptic PDEs to the two-dimensional Helmholtz problem. *SIAM Journal on Numerical Analysis*, 35(1):255–299, 1998.

[25] S. N. Chandler-Wilde and I. G. Graham. Boundary integral methods in high-frequency scattering. *"Highly Oscillatory Problems", editors: B. Engquist, T. Fokas, E. Hairer, A. Iserles*, pages 154–193, Cambridge University Press, 2009.

[26] S. N. Chandler-Wilde, I. G. Graham, S. Langdon, and M. Lindner. Condition number estimates for combined potential boundary integral operators in acoustic scattering. *Journal of Integral Equations and Applications*, 21, Number 2:229–279, 2009.

[27] S. N. Chandler-Wilde and S. Langdon. A Galerkin boundary element method for high frequency scattering by convex polygons. *SIAM Journal on Numerical Analysis*, 45(2):610–640, 2007.

[28] S. N. Chandler-Wilde, S. Langdon, and L. Ritter. A high-wavenumber boundary-element method for an acoustic scattering problem. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 362(1816):647–671, 2004.

[29] T. M. Cherry. Uniform asymptotic formulae for functions with transition points. *Transactions of the American Mathematical Society*, 68(2):224–257, 1950.

[30] P. G. Ciarlet. *The Finite Element Method For Elliptic Problems.* SIAM, Classics in Applied Mathematics, 2002.

[31] C. W. Clenshaw and A. R. Curtis. A method for numerical integration on an automatic computer. *Numerische Mathematik*, 2:197–205, 1960.

[32] J. A. Cochran. The zeros of hankel functions as functions of their order. *Numerische Mathematik*, 7, Issue 3:238–250, 1965.

[33] D. Colton and R. Kress. *Integral equation methods in scattering theory.* A Wiley-Interscience Publication, JonhWiley&Sons, 1983.

[34] D. Colton and R. Kress. *Inverse acoustic and electromagnetic scattering theory.* Springer, 1998.

[35] G. M. Constantine and T. H. Savits. A multivariate Faà di Bruno formula with applications. *Transactions of the American Mathematical Society*, 348(2):503–520, 1996.

[36] P. Debye. Näherungsformeln für die zylinderfunktionen für groe werte des arguments und unbeschränkt veränderliche werte des index. *Mathematische Annalen*, 67 (4):535 –558, 1909.

[37] C. F. F. di Bruno. Note sur une nouvelle formule du calcul differentiel. *The Quarterly Journal of Mathematics*, 1:359–360, 1857.

[38] V. Dominguez. Public domain code. http://www.unavarra.es/personal/victor_dominguez/clenshawcurtisrule, 2009.

[39] V. Dominguez, I. G. Graham, T. Kim, and V. P. Smyshlyaev. Recent progress on hybrid numerical asymptotic boundary integral methods for high-frequency scattering problems. *Proceedings of UKBIM7*, pages 15–23, 2009.

[40] V. Dominguez, I. G. Graham, and V. P. Smyshlyaev. A hybrid numerical-asymptotic boundary integral method for high-frequency acoustic scattering. *Numerische Mathematik*, 106(3):471–510, 2007.

[41] V. Dominguez, I. G. Graham, and V. P. Smyshlyaev. Stability and error estimates for Filon-Clenshaw-Curtis rules for highly-oscillatory integrals. *IMA Journal Numerical Analysis*, 31 (4):1253–1280, 2011.

[42] F. Ecevit. *Integral equation formulations of electromagnetic and acoustic scattering problems: convergence of multiple scattering interations and high-frequency asymptotic expansions*. PhD thesis, University of Minnesota, 2005.

[43] F. Ecevit and F. Reitich. Analysis of multiple scattering iterations for high-frequency scattering problems. part I: the two-dimensional case. *Numerische Mathematik*, 114:271–354, 2009.

[44] G. A. Evans and J. R. Webster. A comparison of some methods for the evaluation of highly oscillatory integrals. *Journal of Applied Mathematics and Computer Science*, 112(1-2):55–69, 1999.

[45] L. Evans. *Partial differential equations*. American mathematical Society, 1998.

[46] V. Filippov. Rigorous justification of the shortwave asymptotic theory of diffraction in the shadow zone. *Journal Sov. Math.*, 6:577–626, 1976.

[47] L. Filon. On a quadrature formula for trigonometric integrals. *Proceedings of the Royal Society of Edinburgh*, 49:38–47, 1928.

[48] E. A. Flinn. A modification of Filon's method of numerical integration. *Journal of the ACM (JACM)*, 7(2):181–184, 1960.

[49] P. Gamallo and R. J. Astley. A comparison of two Trefftz-type methods: the Ultra-weak Variational Formulation and the least-squares method, for solving shortwave 2-d helmholtz problems. *International Journal for Numerical Methods in Engineering*, 71, Issue 4:406 – 432, 2007.

[50] W. M. Gentleman. Implementing Clenshaw-Curtis quadrature, I methodology and experience. *Communications of the ACM*, 15(5):337–342, 1972.

[51] E. Giladi and J. Keller. An asymptotically derived boundary element method for the Helmholtz equations. *Proceedings of the 20th Annual Review of Progress in Applied Computational Electromagnetics*, 2004.

[52] C. J. Gittelson, R. Hiptmair, and I. Perugia. Plane wave discontinuos Galerkin methods. *ESIAM: Mathematical modelling and Numerical analysis*, 43:297–331, 2009.

[53] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products.* Academic Press, 1994.

[54] A. Gray. *Modern differential geometry of curves and surfaces with Mathematica.* CRC Press LLC, 1998.

[55] M. Hardy. Combinatorics of partial derivatives. *The Electronic Journal of Combinatorics*, 13:1–13, 2006.

[56] A. I. Hascelik. Suitable Gauss and Filon-type methods for oscillatory integrals with an algebraic singularity. *Applied Numerical Mathematics*, 59(1):101–118, 2009.

[57] R. Hiptmair, A. Moiola, and I. Perugia. Error analysis of Trefftz-discontinuous Galerkin methods for the time-harmonic Maxwell equations. *Preprint IMATI-CNR Pavia, 5PV11/3/0*, 2011.

[58] M. Honnor, J. Trevelyan, and D. Huybrechs. Numerical evaluation of the two-dimensional partition of unity boundary integrals for Helmholtz problems. *Journal of Computational and Applied Mathematics. Eighth International Conference on Mathematical and Numerical Aspects of Waves*, 234, Issue 6:1656–1662, 2007.

[59] D. Huybrechs and S. Olver. Highly oscillatory quadrature. *"Highly Oscillatory Problems", editors: B. Engquist, T. Fokas, E. Hairer, A. Iserles*, pages 51–71, Cambridge University Press, 2009.

[60] D. Huybrechs and S. Vandewalle. On the evaluation of highly oscillatory integrals by analytic continuation. *SIAM Journal on Numerical Analysis*, 44(3):1026–1048, 2006.

[61] D. Huybrechs and S. Vandewalle. A sparse discretization for integral equation formulations of high frequency scattering problems. *SIAM Journal on Scientific Computing*, 29(6):2305–2328, 2007.

[62] A. Iserles. On the numerical quadrature of highly-oscillating integrals I: Fourier transforms. *IMA Journal Numerical Analysis*, 2003.

[63] A. Iserles. On the numerical quadrature of highly-oscillating integrals II: Irregular oscillators. *IMA Journal Numerical Analysis*, 2005.

[64] A. Iserles and S. P. Norsett. Efficient quadrature of highly oscillatory integrals using derivatives. *Proceedings Royal Soc. 461: 1383-1399*, 2005.

[65] J. B. Keller. Geometrical theory of diffraction. *Journal of the Optical Society of America A*, 52:116–130, 1962.

[66] R. Kress. *Linear Integral equations.* Springer-Verlag, 1989.

[67] R. Kress. On the numerical solution of a hypersingular integral equation in scattering theory. *Journal of Computational and Applied Mathematics*, 61:345–360, August 1995.

[68] R. Kress and W. T. Spassov. On the condition number of boundary integral operators for the exterior dirichlet problem for the helmholtz equation. *Numer. Math.*, 42:77–95, 1983.

[69] R. Leis. Zur dirichletschen randwertaufgabe des auenraumes der schwingungsgleichung. *Mathematische Zeitschrift*, 1965:205– 211, 90.

[70] D. Levin. Procedures for computing one- and two-dimensional integrals of functions with rapid irregular oscillations. *Mathematics of Computation*, 38(158):pp. 531–538, 1982.

[71] D. Levin. Analysis of a collocation method for integrating rapidly oscillatory functions. *Journal of Computational and Applied Mathematics*, 78:131–138, February 1997.

[72] D. Ludwig. Uniform asymptotic expansion for wave propagation and diffraction problems. *SIAM Review*, 12:325–331, 1970.

[73] Y. L. Luke. On the computation of oscillatory integrals. *Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge University Press*, 50:269–277, 1954.

[74] J. M. Melenk. On the convergence of Filon quadrature. *Journal of Computational and Applied Mathematics*, 234(6):1692–1701, 2010.

[75] R. B. Melrose and M. E. Taylor. Near peak scattering and the corrected Kirchhoff approximation for a convex obstacle. *Advances in Mathematics*, 55(3):242 – 315, 1985.

[76] W. R. Mendes. *The Numerical Solution Of Wiener-Hopf Integral Equations.* PhD thesis, University of Bath, 1988.

[77] P. Monk and D.-Q. Wang. A least-squares method for the Helmholtz equation. *Computer Methods in Applied Mechanics and Engineering*, 175, Issues 1-2:121–136, 1999.

[78] C. S. Morawetz and D. Ludwig. An inequality for the reduced wave operator and the justification of geometrical optics. *Communications on pure and applied mathematics*, 21:187–203, 1968.

[79] J. Oliver. Relative error propagation in the recursive solution of linear recurrence relations. *Numerische Mathematik*, 9:323–340, 1967.

[80] F. Olver. Error bounds for asymptotic expansions in turning-point problems. *J. Soc. Indust. Appl. Math.*, 12, Number 1:200–214, 1964.

[81] F. Olver. *Asymptotics and special functions.* Computer Science and Applied Mathematics. Academic Press. New York., 1974.

[82] F. W. J. Olver. The asymptotic expansion of Bessel functions of large order. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 247(930):328–368, 1954.

[83] S. Olver. Moment-free numerical approximation of highly oscillatory integrals with stationary points. *Euro. Jnl of Applied Mathematics*, 18:435–447, 2007.

[84] S. Olver. Fast, numerically stable computation of oscillatory integrals with stationary points. *BIT Numerical Mathematics*, 50:149–171, 2010.

[85] P. Ortiz and E. Sanchez. An improved partition of unity finite element model for diffraction problems. *International Journal for Numerical Methods in Engineering, John Wiley & Sons, Ltd.*, 50, Issue 12:2727–2740, 2001.

[86] O. I. Panich. On the solubility of exterior boundary-value problems for the wave equation and for a system of Maxwell's equations [in russian]. *Uspekhi Mat. Nauk*, 20:1(121):221–226, 1965.

[87] E. Perrey-Debain, O. Laghrouche, P. Bettess, and J. Trevelyan. Plane-wave basis finite elements and boundary elements for three-dimensional wave scattering. *Philosophical Transactions of the Royal Society of London. Series A:Mathematical, Physical and Engineering Sciences*, 362(1816):561–577, 2004.

[88] R. Piessens and M. Branders. Numerical solution of integral equations of mathematical physics, using Chebyshev polynomials. *Journal of Computational Physics*, 21:178–196, 1976.

[89] R. Piessens and M. Branders. Computation of Fourier transform integrals using Chebyshev series expansions. *Computing*, 32:177–186, 1984.

[90] R. Piessens and M. Branders. A numerical method for the integration of oscillatory functions. *BIT Numerical Mathematics*, 11:317–327, 1984.

[91] Y. Saad. *Iterative Methods for Sparse Linear Systems.* SIAM, Philadelphia, 2003.

[92] J. Saranen and G. Vainikko. *Periodic Integral and Pseudodifferential Equations with Numerical Approximation.* Springer.

[93] C. Schwab. *p− and hp− Finite Element Methods. Theory and Applications in Solid and Fluid Machanics.* Numerical MAthematics and Scientific Computation. Clarendon Press. Oxford., 1998.

[94] I. H. Sloan and W. E. Smith. Product integration with the Clenshaw-Curtis points: Implementation and error estimates. *Numerische Mathematik*, 34:387–401, December, 1980.

[95] V. I. Smirnov. *A course of higher Mathematics, Vol. 3 Pt. 2.* Nauka, Moscow (Addison-Wesley, Reading, MA 1964), 1969.

[96] E. A. Spence, S. N. Chandler-Wilde, I. G. Graham, and V. P. Smyshlyaev. A new frequency-uniform coercive boundary integral equation for acoustic scattering. *Communications on Pure and Applied Mathematics*, 64 (10):1384–1415, 2011.

[97] M. Spivak. *Calculus.* World Student Series Edition, University textbook, 1967.

[98] L. L. Thompson and P. M. Pinsky. A Galerkin least-squares finite element method for the two-dimensional Helmholtz equation. *International Journal for Numerical Methods in Engineering*, 38, Issue 3:371–397, 1995.

[99] L. N. Trefethen. Is Gauss quadrature better than Clenshaw-Curtis? *SIAM Review*, 50:67–87, February 2008.

[100] G. N. Watson. The diffraction of electric waves by the earth. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 95(666):pp. 83–99, 1918.

[101] R. Wong. *Asymptotic Approximation of Integrals.* Classics in Applied mathematics, 1989.

[102] S. Xiang. Efficient Filon-type methods for $\int_a^b f(x)\,\mathrm{e}^{\mathrm{i}\omega g(x)}\,\mathrm{d}x$. *Numerische Mathematik*, 105(4):633–658, 2007.

Appendices

Appendix A

Eikonal equation

We begin this section by describing the mathematical formulation of ray methods. This leads to well known *eikonal* and *transport* equations, see Sections A.1, and A.3. Solving the eikonal equation gives us greater physical understanding behind different types of ray propagation (e.g. reflected and diffracted rays).

Recall the scattering problem under consideration: find $u(\boldsymbol{x}) = u^S(\boldsymbol{x}) + u^I(\boldsymbol{x})$ that satisfies

$$\Delta u + k^2 u = 0 \quad \text{in } \mathbb{R}^2 \setminus \overline{\Omega}, \tag{A.1}$$

$$u^S + u^I = 0 \quad \text{on } \Gamma, \tag{A.2}$$

$$\lim_{r\to\infty} \sqrt{r}\left(\frac{\partial u^S}{\partial r} - iku^S\right) = 0. \tag{A.3}$$

When the incident wave hits an illuminated part of the boundary of the scatterer, it gives rise to a reflected wave. When the incident wave hits the boundary of the scatterer tangentially, i.e. transition part of the boundary, creeping and then diffracted rays arise. In later sections, we will discuss in detail how these can be found in practise.

An approximation to the scattered wave $u(\boldsymbol{x})$, solving the Helmholtz equation (A.1), is sought in the form of the ray expansion as follows:

$$u(\boldsymbol{x}, k) = \exp\left(ik\tau(\boldsymbol{x})\right) \sum_{j=0}^{\infty} \frac{A_j(\boldsymbol{x})}{(ik)^j} \ . \tag{A.4}$$

The unknown functions in (A.4), are the "Eikonal" $\tau(\boldsymbol{x})$ and the "asymptotes" $A_0(\boldsymbol{x})$, $A_1(\boldsymbol{x})$,.... The equations for $\tau(\boldsymbol{x})$ and $A_j(\boldsymbol{x})$, are obtained by formally substituting the ray expansion (A.4) into Helmholtz equation (A.1) and equating terms of order k^2, k^1, k^0, k^{-1} and so on. Namely,

(i) the Eikonal equation (by equating the terms of order k^2):

$$|\nabla\tau(\boldsymbol{x})|^2 = 1, \tag{A.5}$$

for $\boldsymbol{x} \in \mathbb{R}^2 \setminus (\Omega \cup \Gamma)$, and

(ii) the Transport equation (by equating the terms of order k^{1-n}, $n \geq 0$):

$$\triangle\tau(\boldsymbol{x})A_n(\boldsymbol{x}) + 2\left(\nabla\tau(\boldsymbol{x}) \cdot \nabla A_n(\boldsymbol{x})\right) = -\triangle A_{n-1}(\boldsymbol{x}), \quad n = 0, 1, 2, ..., \tag{A.6}$$

where, by convention, $A_{-1} \equiv 0$.

A.1 The Eikonal equation

The Eikonal equation (A.5), is a non-linear first-order partial differential equation that can be solved using method of charachteristics, see for example [45] or [14]. Solving the Eikonal equation (A.5) is equivalent to finding $\tau(x_1, x_2)$ that satisfies the following equation

$$F(x_1, x_2, p, q, \tau) = p^2 + q^2 - 1 = 0, \quad \text{where} \quad p = \tau_{x_1}, \quad q = \tau_{x_2}, \tag{A.7}$$

where $(\cdot)_{x_1}$ and $(\cdot)_{x_1}$ denote partial derivatives with respect to x_1 and x_2 respectively. The method of characteristics reduces the problem of solving non-liner first order PDE (A.7) to a system of quasi-linear ODEs along a characteristic curve. Suppose that (x_1, x_2, p, q, τ), solve a family of a general first-order ODE system

$$\frac{dx_1}{d\sigma} = 2p, \quad \frac{dx_2}{d\sigma} = 2q, \quad \frac{d\tau}{d\sigma} = 2, \quad \frac{dp}{d\sigma} = 0, \quad \frac{dq}{d\sigma} = 0,$$

where σ is a parameter that relates to the distance along the characteristic curve, and (x_1, x_2, p, q, τ) smoothly depends on the additional parameter s, and the mapping $(s, \sigma) \to (x_1, x_2)$ is invertible in some domain $D \in \mathbb{R}^2$, then

$$\tau(x_1, x_2) = \tau\left(s\left(x_1, x_2\right), \sigma\left(x_1, x_2\right)\right)$$

solves the equation (A.7). The charachteristic curve describes a ray. The solutions to this system of equations are:

$$x_1(s, \sigma) = x_{01}(s) + 2p_0(s)\sigma, \qquad x_2(s, \sigma) = x_{02}(s) + 2q_0(s)\sigma, \tag{A.8}$$

$$p(s) = p_0(s, \sigma), \qquad q(s, \sigma) = q_0(s), \tag{A.9}$$

$$\tau(s, \sigma) = \tau_0(s) + 2\sigma, \qquad 1 = p^2(s) + q^2(s). \tag{A.10}$$

Here $x_{01}(s)$, $x_{02}(s)$, $q_0(s)$, $p_0(s)$, $\tau_0(s)$ are defined from the appropriate initial conditions. Note that by eliminating σ, we obtain the following solution for the Eikonal

$$\tau(\boldsymbol{x}) = \tau(\boldsymbol{x_0}) + |\boldsymbol{x} - \boldsymbol{x_0}|, \tag{A.11}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}_0$ correspond to the same charachteristic curve, i.e. the same value of s in (2.11) - (2.13). This equation has the following physical meaning: the Eikonal is determined as the value of τ at the boundary, at the point $\boldsymbol{x_0}$, plus the length of the scattered ray $\boldsymbol{x}_0\boldsymbol{x}$. It is not obvious from the equation (A.11) alone, however, what the scattered rays are. In the following two sections, we determine the scattered rays by studying the system of equations (A.10) in more detail.

A.2 Reflected rays

In the illuminated region, the incident wave gives rise to a reflected wave. We denote τ_I and τ_R as the incident and reflected components of the eikonal, that are required to satisfy the following condition, consistently with (A.2):

$$\tau_0^I(s) = \tau^I(x_{01}(s), x_{02}(s)) = \tau_0^R(s). \tag{A.12}$$

This is the initial condition, with respect to the parameter σ, for the ODE system (A.11). Differentiating (A.12) with respect to s, and using (A.7) we deduce that

$$p_0^R x_{01}^{'}(s) + q_0^R x_{02}^{'}(s)) = (\tau_0^I)^{'}(s), \qquad p_0^I x_{01}^{'}(s) + q_0^I x_{02}^{'}(s) = (\tau_0^I)^{'}(s). \tag{A.13}$$

Note that the left hand sides of (A.13) represent the dot-products of the unit vector $\left(p_0^R, q_0^R\right)$ and $\left(p_0^I, q_0^I\right)$, respectively, with the tangent $(x_{01}^{'}(s), x_{02}^{'}(s)))$ to the boundary. Since, by (2.11)-(2.13), $\left(p_0^I, q_0^I\right)$ and $\left(p_0^R, q_0^R\right)$ are unit vectors in the direction of the incident and reflected rays respectively, the incident ray and the reflected ray must make equal angles with the tangent to the boundary at each point. Denote

$$p_0^I(s) = \cos\alpha_I(s), \qquad q_0^I(s) = \sin\alpha_I(s),$$
$$x_{01}^{'}(s)/\sqrt{x_{01}^{'}(s) + x_{02}^{'}(s)} = \cos\theta(s), \qquad x_{02}^{'}(s)/\sqrt{x_{01}^{'}(s) + x_{02}^{'}(s)} = \sin\theta(s),$$

Then,

$$p_0^R(s) = \cos\alpha_R(s), \quad q_0^R(s) = \sin\alpha_R(s), \tag{A.14}$$

where $\alpha_R(s) = 2\theta(s) - \alpha_I(s)$, see Figure A.1. This is a well known Snell's law that provides simple geometric description of the wave motion: the angle of reflection equals to the angle of incidence.

A.3 Amplitude of the wave fronts

Transport equations have the following solutions, see e.g.[14] or [5]:

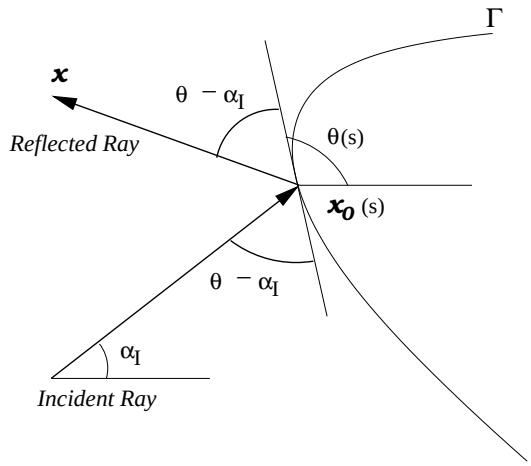


Figure A.1: *Directions of the incident and reflected waves.*

$$A_j(\sigma) = \frac{1}{\left(\sqrt{J(\sigma)J(\sigma_1)} - J(\sigma)\right)} \int_{\sigma_1}^{\sigma} \sqrt{J(\sigma')}\nabla^2 A_{j-1}(\sigma')d\sigma',$$

where,

$$J(\sigma) = \left| \boldsymbol{n}(s) \cdot \frac{d\boldsymbol{x}(s,\sigma)}{ds} \right|. \tag{A.15}$$

Since $u(\boldsymbol{x}) = u^I(\boldsymbol{x}) + u^R(\boldsymbol{x})$ (and both, incident and reflected waves satisfy (A.4)) and $\tau^I(\boldsymbol{x}) = \tau^R(\boldsymbol{x})$ for $\boldsymbol{x} \in \Gamma$, we deduce the asymptotic expansion of the normal derivative of $u(\boldsymbol{x})$,

$$\begin{aligned} \frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{x}) &:= 2ik\left(\nabla\tau \cdot \boldsymbol{n}\right)(\boldsymbol{x}) \exp\left(ik\tau(\boldsymbol{x})\right) A_0(\boldsymbol{x}) \\ &+ \sum_{j=0}^{\infty} \frac{1}{(ik)^j} \left(A_{j+1}\nabla\tau + \nabla A_j\right)(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}) \exp\left(ik\tau(\boldsymbol{x})\right). \end{aligned}$$

However, in the transition regions, the Jacobian (A.15) vanishes. In other words, the ray coordinates (s, σ) can no longer be uniquely represented by the cartesian coordinates x_1 and x_2. Hence, the ray method is not valid in the neighbourhood of the transition domains.

In the case of plane wave incidence, we find $\tau(\boldsymbol{x}) = \boldsymbol{x} \cdot \boldsymbol{a}$ for $\boldsymbol{x} \in \Gamma$. Therefore, since $A_0(\boldsymbol{x}) = 1$, $\boldsymbol{x} \in \Gamma$, we find the main order term for the expansion of the normal derivative,

$$\frac{\partial u}{\partial \boldsymbol{n}}(\boldsymbol{x})_m = 2ik(\boldsymbol{a} \cdot \boldsymbol{n}) \exp\left(ik(\boldsymbol{x} \cdot \boldsymbol{a})\right), \tag{A.16}$$

where the subscript "'m"' denotes the "'main order term"'.

Appendix B

Method of characteristics

In order to find eikonal $\tau(\boldsymbol{x})$, the partial differential equation (A.5) is solved using method of characteristics. Method of characteristics solves first order non-linear partial differential equations of the type

$$F(x_1, x_2, p, q, \tau) = 0, \quad \text{where} \;\; p = \tau_{x_1}, \;\; q = \tau_{x_2}, \tag{B.1}$$

by reducing the problem into a system of quasi-linear partial differential equations:

$$\frac{dx_1}{F_p} = \frac{dx_2}{F_q} = \frac{d\tau}{pF_p + qF_q} = -\frac{dp}{F_{x_1} + pF_\tau} = -\frac{dq}{F_{x_2} + qF_\tau}. \tag{B.2}$$

If we introduce a new parameter σ that indicates the distance along the characteristic curve, then we can rewrite (B.2) as follows

$$\frac{dx_1}{d\sigma} = F_p, \quad \frac{dx_2}{d\sigma} = F_q, \quad \frac{d\tau}{d\sigma} = pF_p + qF_q, \tag{B.3}$$

$$\frac{dp}{d\sigma} = -F_{x_1} - pF_\tau, \quad \frac{dq}{d\sigma} = -F_{x_2} - qF_\tau. \tag{B.4}$$

To find a solution τ that satisfies (B.1) and passes through the initial curve

$$x_1 = x_{01}(s), \;\; x_2 = x_{02}(s), \quad \tau = \tau_0(s),$$

where s is the variable of parametrisation of the initial curve, we solve the system of equations (B.4), with $p = p_0(s)$ and $q = q_0(s)$ determined from the following equations:

$$F(x_{01}(s), x_{02}(s), \tau_0(s), p_0(s), q_0(s)) = 0 \tag{B.5a}$$

$$\tau_0^{'}(s) = p_0(s)x_{01}^{'}(s) + q_0(s)x_{02}^{'}(s). \tag{B.5b}$$

It is possible to solve the system (B.5b) in the neighbourhood of initial point s_0, provided that

$$J(s, \sigma) = \det \begin{bmatrix} (x_1)_s & (x_2)_s \\ (x_1)_\sigma & (x_2)_\sigma \end{bmatrix} \neq 0. \tag{B.6}$$

Then integrating (B.4), we obtain the solution:

$$x_1 = x_1(s, \sigma),\ x_2 = x_2(s, \sigma), \quad \tau = \tau(s, \sigma), \quad p = p(s, \sigma), \quad q = q(s, \sigma).$$

From these equations, we can obtain τ.

Appendix C

Locations of stationary points

In this section, we examine the locations of the stationary points of $\psi_{[s]}$, i.e. those points $(s,t) \in ([0,2\pi] \times [0,2\pi])$ satisfying

$$\psi'_{[s]}(t) := \boldsymbol{\gamma}'(t)\left(\boldsymbol{a} - \frac{\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|}\right) = 0. \tag{C.1}$$

In particular, we aim to prove Theorem 4.16 in Chapter 4. The proof of Theorem 4.16 requires two intermediate results that we prove in Lemma C.2 and Lemma C.5.

We begin this section with the definition of convex curves [54]. Any straight line L divides $\mathbb{R}^2$ into two half-planes, H_1 and H_2, such that:

$$H_1 \cup H_2 = \mathbb{R}^2 \quad \text{and} \quad H_1 \cap H_2 = L,$$

we say that a curve C lies on one side of the straight line L if either $C \subset H_1$ or $C \subset H_2$.

Definition C.1. *A plane closed curve is convex if it lies on one side of each of its tangent lines.*

This definition can also be understood in terms of the behaviour of the pair of vectors $(\boldsymbol{T}, \boldsymbol{N})$ - tangent and normal to Γ. Let $\{\boldsymbol{\gamma} : [0,2\pi] \to \mathbb{R}^2\}$ be the arc-length parametrization of the smooth, convex, closed curve $\Gamma \subset \mathbb{R}^2$. Then, the vector

$$\boldsymbol{T}(s) := \boldsymbol{\gamma}'(s),$$

is the unit tangent vector of Γ at a point $\boldsymbol{\gamma}(s)$ on the boundary. There are two unit vectors perpendicular to $\boldsymbol{T}$, we choose to define $\boldsymbol{N}$, the normal vector of Γ at $\boldsymbol{\gamma}(s)$, as a unit vector obtained by rotating $\boldsymbol{T}$ clockwise by $\pi/2$.

Since Γ always lies on one side of its tangent lines, we deduce that unit vector $\boldsymbol{T}$ rotates in only one direction (clockwise or anti-clockwise) as it moves along the curve in the direction of increasing s. We choose the direction of increasing s to be anti-clockwise (such closed curves are called positively-oriented). Then the tangent $\boldsymbol{T}$ and the outward normal, $\boldsymbol{N}$, also rotate anti-clockwise.

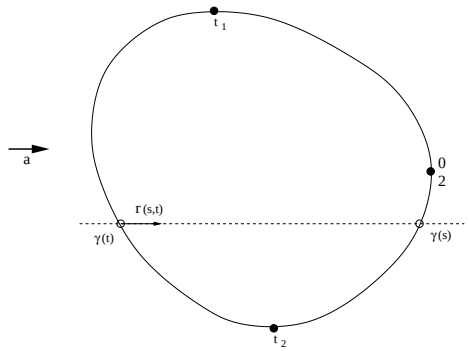


Figure C.1: *Stationary s and t are opposite.*

The function $\psi'_{[s]}(t)$ can be written as follows, in terms of three unit vectors:

$$\psi'_{[s]}(t) := \boldsymbol{T}(t) \cdot (\boldsymbol{a} - \boldsymbol{\rho}(s,t)) \,, \tag{C.2}$$

where $\boldsymbol{a}$ is fixed vector denoting the direction of incident wave, and vector $\boldsymbol{\rho}(s,t)$ is defined as follows

$$\boldsymbol{\rho}(s,t) := \frac{\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)}{|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)|}, \quad \text{hence} \lim_{t \to s^{\pm}} \boldsymbol{\rho}(s,t) = \mp \boldsymbol{T}(s).$$

It is possible to determine from the convexity of the boundary, the behaviour of the angles between the tangent vector $\boldsymbol{T}(t)$ and $\boldsymbol{a}$ and similarly between $\boldsymbol{T}(t)$ and $\boldsymbol{r}(s,t)$.

The variable t in the illuminated part of the boundary

In the interval (t_1, t_2) the following holds

$$\begin{aligned} \boldsymbol{N}(t_1) \cdot \boldsymbol{a} &= 0, \\ \boldsymbol{N}(s) \cdot \boldsymbol{a} &< 0, \qquad s \in (t_1, t_2), \\ \boldsymbol{N}(t_2) \cdot \boldsymbol{a} &= 0. \end{aligned}$$

Lemma C.2. *Given $t \in (t_1, t_2)$, there exist unique $s \in [0, t_1) \cup (t_2, 2\pi]$, such that*

$$\psi'_{[s]}(t) = 0.$$

In fact, it is necessary and sufficient for s to satisfy the following equation:

$$\boldsymbol{\rho}(s,t) = \boldsymbol{a}. \tag{C.3}$$

One of the consequences of the lemma, is that, given $t \in \Lambda_2$, there are no stationary points $s \in [a, d]$, provided that the transition intervals Λ_1 and Λ_3 extend further into illuminated zone than into the shadow.

To prove Lemma C.2, we study the angles between vectors $\boldsymbol{a}$, $\boldsymbol{\rho}(s,t)$ and $\boldsymbol{T(t)}$, behaviour

of which can be sufficiently fully determined from the fact that the boundary Γ is convex. For this, we require a few intermediate results that we present in Proposition C.3 and Proposition C.4.

Proposition C.3. *Define $\alpha(s,t)$ as the angle between $\boldsymbol{T}(t)$ and $\boldsymbol{\rho}(s,t)$, measured anti-clockwise from $\boldsymbol{T}(t)$, see Figure C.2, then:*

$$\alpha(s,t) \in (0,\pi), \quad \textit{for all } s,t \in [0,2\pi], \quad s \neq t.$$

Proof. At the point $\boldsymbol{\gamma}(t)$, $\boldsymbol{\rho}(s,t)$ is the unit vector directed from $\boldsymbol{\gamma}(t)$ to $\boldsymbol{\gamma}(s)$. But $\boldsymbol{\gamma}(s)$ is the point on the curve Γ which lies on one side of the tangent line $L : c\boldsymbol{T}(t) + \boldsymbol{\gamma}(t)$, see Figure C.2. Since $\boldsymbol{T}$ rotates anti-clockwise, the angle α belongs to the interval $[0,\pi]$. Moreover, $\alpha = 0$ or $\alpha = \pi$ only when $t \to s^{\pm}$, i.e. when $\boldsymbol{\rho}(s,t)$ is parallel to $\boldsymbol{T}(t)$. Hence $\alpha(s,t) \in (0,\pi)$. □

We can deduce from the proposition that the following holds:

$$\boldsymbol{N}(t) \cdot \boldsymbol{\rho}(s,t) < 0, \quad \forall s,t \in [0,2\pi], \quad s \neq t,$$

and similarly,

$$\boldsymbol{N}(s) \cdot \boldsymbol{\rho}(s,t) > 0, \quad \forall s,t \in [0,2\pi], \quad s \neq t.$$

To see this, recall the definition of convexity. The tangent line passing through the point $\boldsymbol{\gamma}(t)$, divides $\mathbb{R}^2$ into two half-planes one of which contains the curve Γ and towards which vector $\boldsymbol{\rho}(s,t)$ is directed (because it is always directed into interior of Γ). Vector $\boldsymbol{N}(t)$ however, always directed into exterior of Γ and hence the result.

Proposition C.4. *Define $\beta(t)$ as the angle between $\boldsymbol{T}(t)$ and $\boldsymbol{a}$, measured anti-clockwise from $\boldsymbol{T}(t)$, see Figure C.3, then*

$$\beta(t) \in (0,\pi), \quad \textit{for all } t \in (t_1,t_2).$$

Proof. Consider the function $g(t) := \boldsymbol{T}(t) \cdot \boldsymbol{a}$, then

$$g(t) := |\boldsymbol{T}(t)||\boldsymbol{a}| \cos(\beta(t)) = \cos(\beta(t)). \tag{C.4}$$

However, $g'(t) = \boldsymbol{T}'(t) \cdot \boldsymbol{a} = \boldsymbol{\gamma}''(t) \cdot \boldsymbol{a} > 0$ for all $t \in (t_1,t_2)$ (since $\boldsymbol{\gamma}''(t)$ is parallel to $\boldsymbol{N}(t)$ but faces inwards). Thus, $g(t)$ is a monotonically increasing function for $t \in (t_1,t_2)$, or equivalently $\cos(\beta(t))$ is monotonically increasing, and:

$$g(t_1) = -1 \quad \Rightarrow \quad \beta(t_1) = \pi,$$

$$g(t_2) = 1 \quad \Rightarrow \quad \beta(t_2) = 0,$$

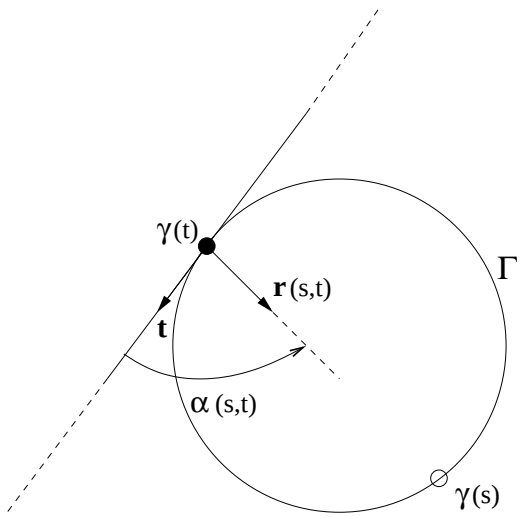


Figure C.2: $\alpha(s,t)$ *as the angle between* $\boldsymbol{T}(t)$ *and* $\boldsymbol{\rho}(s,t)$*, measured anti-clockwise from* $\boldsymbol{T}(t)$*.*

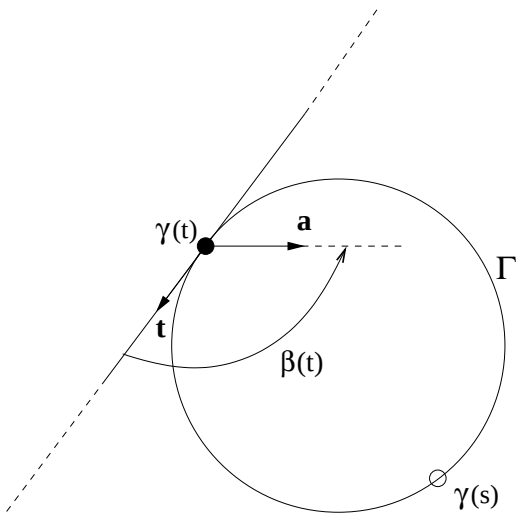


Figure C.3: $\beta(t)$ *as the angle between* $\boldsymbol{T}(t)$ *and* $\boldsymbol{a}$*, measured anti-clockwise from* $\boldsymbol{T}(t)$*.*

now from (C.4), the result follows. □

proof of Lemma C.2.

Proof. With $\alpha(s,t)$ and $\beta(t)$ as before, we can write the vector $\boldsymbol{a}$ and $\boldsymbol{\rho}(s,t)$ in terms of the pair of orthogonal vectors $\boldsymbol{T} := \boldsymbol{T}(t)$ and $\boldsymbol{N} := \boldsymbol{N}(t)$ as follows:

$$\begin{aligned} \boldsymbol{\rho}(s,t) &= \boldsymbol{T}\cos(\alpha(s,t)) - \boldsymbol{N}\sin(\alpha(s,t)), \\ \boldsymbol{a} &= \boldsymbol{T}\cos(\beta(t)) - \boldsymbol{N}\sin(\beta(t)). \end{aligned}$$

Then, using elementary trigonometric identities,

$$\begin{aligned} \psi'_{[s]}(t) &= \cos(\beta(t)) - \cos(\alpha(s,t)) \\ &= -2\sin\left(\frac{\beta(t)+\alpha(s,t)}{2}\right)\sin\left(\frac{\beta(t)-\alpha(s,t)}{2}\right). \end{aligned}$$

Thus $\psi'_{[s]}(t) = 0$ when

1. $\frac{\beta(t)+\alpha(s,t)}{2} = 0$, or
2. $\frac{\beta(t)+\alpha(s,t)}{2} = \pi$, or
3. $\frac{\beta(t)-\alpha(s,t)}{2} = 0$, or
4. $\frac{|\beta(t)-\alpha(s,t)|}{2} = \pi$.

We eliminate cases 1,2 and 4 since $\alpha(s,t),\ \beta(t) \in (0,\pi)$, see Proposition C.3 and Propo-

sition C.4. Case 3 implies

$$\boldsymbol{a} = \boldsymbol{\rho}(s,t),$$

which is only possible when s does not belong to the interval (t_1, t_2). The later can be deduced as follows: for all $(s,t) \in [0, 2\pi],\ s \neq t$,

$$\boldsymbol{N}(s) \cdot \boldsymbol{\rho}(s,t) > 0.$$

However, since $\boldsymbol{a} = \boldsymbol{\rho}(s,t)$, it follows $\boldsymbol{N}(s) \cdot \boldsymbol{a} > 0$. Therefore, $s \notin (t_1, t_2)$. This completes the proof. □

t in the transition parts of the boundary

Consider the case when $t \in [0, t_1] \cup [t_2, 2\pi]$ and $t > s$ (the above-the-diagonal domains).

Recall the transition domains Λ_1 and Λ_3,

$$\begin{aligned} \Lambda_1 \quad &= \quad [a,b], \qquad \text{with} \qquad a = t_1 - \delta, \qquad b = t_1 + \varepsilon, \\ &\text{also} \quad a' = t_1 + \delta. \end{aligned}$$

and

$$\begin{aligned} \Lambda_3 \quad &= \quad [c,d], \qquad \text{with} \qquad c = t_2 - \varepsilon, \qquad d = t_2 + \delta, \\ &\quad\ \ d' = t_2 - \delta. \end{aligned}$$

Lemma C.5. *Given $t \in [0, t_1]$, there exists unique $s \in [0, 2\pi]$ such that*

$$\psi'_{[s]}(t) = 0. \tag{C.5}$$

Moreover, for $t \in [0, t^*]$, where $t^* \in (0, t_1)$ can be found by solving for $t \in (0, t_1)$,

$$\psi'_{[0]}(t) = 0,$$

the unique point s that satisfies (C.5) is in $s \in [s^*, 2\pi]$, where s^* is found by solving $\boldsymbol{r}(s^*, 0) = \boldsymbol{a}$ (for the case of a circle $s^* = \pi$). While for $t \in [t^*, t_1]$, equation (C.5) holds for $s \in [0, t_1]$ with $s < t$.

Proof. Similarly to the proof of Proposition C.4, we can show that when $t \in (0, t_1)$,

$$\beta(t) \in (\pi, 3\pi/2). \tag{C.6}$$

To see this, recall that $g'(t) = \boldsymbol{\gamma}''(t) \cdot \boldsymbol{a} < 0$ when $t \in [0, t_1] \cup [t_2, 2\pi]$. Hence $g(t)$ is decreasing function for $t \in (0, t_1)$, or equivalently $\cos(\beta(t))$ is decreasing. But $g(0) = 0$

and $g(t_1) = -1$ and $\beta(t) > \pi$, hence $\beta(t) \in (\pi, 3\pi/2)$.

On the other hand, from Proposition C.3 we deduce that for all s and t, $s \neq t$,

$$\alpha(s,t) \in (0,\pi). \tag{C.7}$$

The function $\psi'_{[s]}(t)$ is again zero when either of the equations

$$\begin{array}{lll} 1. & \frac{\beta(t)+\alpha(s,t)}{2} = 0, & \text{or} \\ 2. & \frac{\beta(t)+\alpha(s,t)}{2} = \pi, & \text{or} \\ 3. & \frac{\beta(t)-\alpha(s,t)}{2} = 0, & \text{or} \\ 4. & \frac{|\beta(t)-\alpha(s,t)|}{2} = \pi, & \end{array}$$

hold. Using (C.6) and (C.7), we eliminate cases 1,3 and 4.

Therefore, for $t \in (0, t_1)$, the pair (s,t) satisfies $\psi'_{[s]}(t) = 0$ if and only if the equation

$$\beta(t) = 2\pi - \alpha(s,t), \tag{C.8}$$

holds. Suppose for $t \in (0, t_1)$, there exist s_1 and s_2 that satisfy (C.8),

$$\begin{aligned} \beta(t) &= 2\pi - \alpha(s_1,t), \\ \beta(t) &= 2\pi - \alpha(s_2,t). \end{aligned}$$

Then $\alpha(s_1,t) = \alpha(s_2,t)$, and therefore $\boldsymbol{r}(s_1,t) = \boldsymbol{r}(s_2,t)$, thus $s_1 \equiv s_2$.

Now, recall that $\alpha(s,t)$ is the angle between the tangent vector $\boldsymbol{T}(t)$ and the vector $\boldsymbol{\rho}(s,t)$, i.e. given t and $\alpha(s,t)$ from (C.8), the corresponding stationary s can be uniquely determined.

Since $\beta(t) \in (\pi, 3\pi/2)$ is monotonically decreasing function of t, from equation (C.8) we deduce that $\alpha(s,t)$ is monotone increasing function of t and $\alpha(s,t) \in (\pi/2, \pi)$. In other words, as t increases from 0 to t_1, the angle $\alpha(s,t)$ is monotonically increasing, assigning correspondingly growing s. Moreover,

$$\text{when} \quad t = 0, \quad \beta(t) = 3\pi/2, \quad \rightarrow \quad \alpha(s,t) = \pi/2 \quad \rightarrow \quad s = s^*, \tag{C.9}$$

$$\text{when} \quad t = t_1, \quad \beta(t) = \pi, \quad \rightarrow \quad \alpha(s,t) = \pi \quad \rightarrow \quad s = t_1. \tag{C.10}$$

This concludes the proof. □

The result of Lemma C.5 can be interpreted as follows: as t monotonically increases from 0 to t_1, the corresponding unique stationary point s also moves along the boundary monotonically increasing in value starting from s^* (until it reaches 2π, where it switches

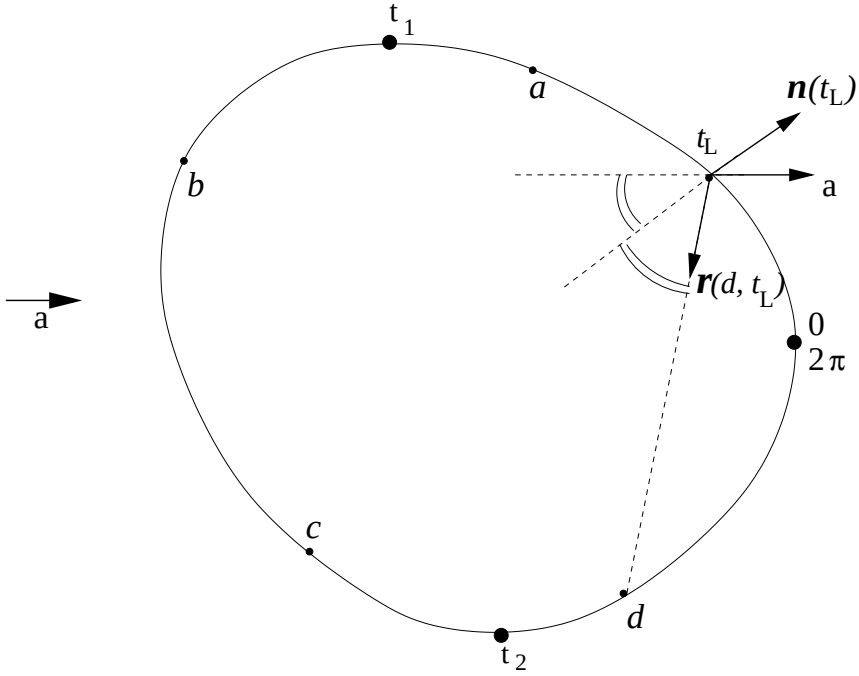


Figure C.4: *The point t_L that represent the limit of how far the transition zone Λ_1 can extend into the shadow is illustrated. There are similar restrictions for the domain Λ_3 that can be derived similarly to Λ_1.*

back to 0) to t_1.

There exists a point $t_L \in (0, t_1)$, such that

$$\psi'_{[d]}(t_L) = 0,$$

where $s = d$ is the end point of the Λ_3: $\Lambda_3 = [c, d]$. From Lemma C.5, we deduce that for $t > t_L$, and $t < t_1$, there are no points $s \in \Lambda_1 \cup \Lambda_2 \cup \Lambda_3$, with $s < t$, such that $\psi'_{[s]}(t) = 0$ holds.

Lemma C.6. *Given $t \in [t_2, 2\pi]$, there exist unique $s \in [0, 2\pi]$, such that (C.5) holds. Moreover, for $t \in [t^*, 2\pi]$, where $t^* \in (t_2, 2\pi)$ can be found by solving for $t \in (t_2, 2\pi)$,*

$$\psi'_{[0]}(t) = 0,$$

the unique point s that satisfies (C.5) is in $s \in [0, s^*]$, where s^* is defined as before: $\boldsymbol{r}(s^*, 0) = \boldsymbol{a}$. While for $t \in [\pi, t^*]$, equation (C.5) holds for $s \in [t_2, 2\pi]$ with $t < s$.

Proof. Follows analogously to the proof of Lemma C.5 □

Corollary C.7. *For $t \in [a, t_1)$ and $s \in [a, d]$, $s < t$, the condition*

$$\psi'_{[s]}(t) \neq 0, \tag{C.11}$$

holds, provided that

$$a > t_L,$$

where t_L can be determined from the equation:

$$\boldsymbol{a} \cdot \boldsymbol{n}(t_L) = -\boldsymbol{r}(d, t_L) \cdot \boldsymbol{n}(t_L),$$

or equivalently $\Psi_t(d, t_L) = 0$.

Remark C.8. *Let us consider the lower-triangular domain of $(\Lambda_1 \times \Lambda_1)$ with $\Lambda_1 := [a, b] = [a, t_1] \cup [t_1, b]$.*

Corollary C.7 implies the condition (C.11) is satisfied for $t \in [a, t_1]$ $s \in \Lambda_1$ with $s < t$. On the other hand, Lemma C.2 implies that given $t \in [t_1, b]$, the equation (C.5) holds for $s < t_1$. Therefore, the condition (C.11) is satisfied for $t \in [t_1, b]$ and $s \in \Lambda_1$ with $s < t$.

Therefore, Hypothesis A is satisfied for $(s, t) \in \Lambda_1 \times \Lambda_1$ with $s > t$.

Appendix D

Chebyshev-weighted L_2-norms.

Definition D.1. *We call the functional* $\|\cdot\|_{L_2[-1,1],\omega} : V \to \mathbb{R}$, *defined as*

$$\|f\|_{L_2[-1,1],\omega} := \left(\int_{-1}^{1} \frac{(f(x))^2}{\sqrt{1-x^2}} dx\right)^{1/2}$$

the Chebyshev-weighted L_2*-norm.*

Note that since

$$\int_{-1}^{1} T_n(x)T_m(x)\frac{dx}{\sqrt{1-x^2}} = \begin{cases} 0 & : n \neq m \\ \pi & : n = m = 0 \\ \pi/2 & : n = m \neq 0 \end{cases}$$

for $f(x) = \sum_{j=0}^{p} \alpha_j T_j(x)$, $x \in [-1,1]$:

$$\|f\|^2_{L_2[-1,1],\omega} = \int_{-1}^{1} \frac{(f(x))^2}{\sqrt{1-x^2}} dx = \sum_{j=0}^{p}\sum_{m=0}^{p} \alpha_j\alpha_m \int_{-1}^{1} \frac{T_j(x)T_m(x)}{\sqrt{1-x^2}} dx = \frac{\pi}{2}\left(\sum_{j=0}^{p} |\alpha_j|^2\right).$$

Similarly, for $g(x) = \sum_{j=0}^{p} \alpha_j S_j(x) = \sum_{j=0}^{p} \alpha_j T_j\left(\frac{2(x-a)}{(b-a)} - 1\right)$, $x \in [a,b]$

$$\int_{-1}^{1} \frac{T_n(t)T_m(t)}{\sqrt{1-t^2}} dt = \frac{(b-a)}{2}\int_{a}^{b} \frac{S_n(x)S_m(x)}{\sqrt{\frac{2(x-a)}{b-a}}\sqrt{\frac{2(b-x)}{b-a}}} dx = \tag{D.1}$$

$$\int_{a}^{b} \frac{S_n(x)S_m(x)}{\sqrt{(a-x)(x-b)}} dx = \begin{cases} 0 & : n \neq m \\ \pi & : n = m = 0 \\ \pi/2 & : n = m \neq 0 \end{cases} \tag{D.2}$$

hence

$$\|g\|^2_{L_2[a,b],\omega} = \int_{a}^{b} \frac{(g(x))^2}{\sqrt{(x-a)(b-x)}} dx = \sum_{j=0}^{p}\sum_{m=0}^{p} \alpha_j\alpha_m \int_{a}^{b} \frac{S_j(x)S_m(x)}{\sqrt{(x-a)(b-x)}} dx = \frac{\pi}{2}\left(\sum_{j=0}^{p} |\alpha_j|^2\right).$$

Therefore $\|v\|_{L_2[-1,1],\omega} = \|v\|_{L_2[a,b],\omega}$. The following proposition shows how L_∞-norm is bounded by the Chebyshev-weighted L_2-norm.

Proposition 1.

$$||v(x)||^2_{L_\infty([-1,1])} \leq \frac{2(p+1)}{\pi}\int_{-1}^{1}\frac{(v(x))^2}{\sqrt{1-x^2}}dx := \frac{2(p+1)}{\pi}||v(x)||^2_{L_2,\omega} \tag{D.3}$$

Proof. Since $v(x) = \sum_{j=0}^{p} V_j T_j(x)$,

$$v^2(x) = \left(\sum_{j=0}^{p} V_j T_j(x)\right)^2 \leq (p+1)\left(\sum_{j=0}^{p} V_j^2\right).$$

Hence

$$v^2(x) \leq \frac{2(p+1)}{\pi}\int_{-1}^{1}\frac{v^2(x)}{\sqrt{1-x^2}}dx$$

and $(D.3)$ follows. □

Lemma D.2 (Inverse estimate for weighted norm)**.** *For $v(x) = \sum_{j=0}^{p} V_j T_j(x)$, the Chebyshev-weighted L_2-norm is bounded by L_2-norm as follows:*

$$||v(x)||_{L_2([-1,1]),\omega} \leq \left(\frac{32(p+1)}{\pi}\right)^{1/2} ||v(x)||_{L_2([-1,1])}$$

Proof. Set $\varepsilon_{\mu p} = 1 - \cos(1/\mu p)$ for $\mu \geq 1$. Since

$$\int_{-1}^{1}\frac{v^2(x)}{\sqrt{1-x^2}}dx = \int_{-1+\varepsilon_{\mu p}}^{-1}\frac{v^2(x)}{\sqrt{1-x^2}}dx + \int_{-1+\varepsilon_{\mu p}}^{1-\varepsilon_{\mu p}}\frac{v^2(x)}{\sqrt{1-x^2}}dx + \int_{1-\varepsilon_{\mu p}}^{1}\frac{v^2(x)}{\sqrt{1-x^2}}dx \tag{D.4}$$

and

1. $$\max_{x\in[-1,-1+\varepsilon_{\mu p}]} |v(x)| \leq ||v(x)||_{L_\infty(I)}$$

2. $$\max_{x\in[-1+\varepsilon_{\mu p},1-\varepsilon_{\mu p}]} \frac{1}{\sqrt{1-x^2}} \leq \frac{1}{\sqrt{1-(1-\varepsilon_{\mu p})^2}} = \frac{1}{\sin(1/\mu p)} \leq 2\mu p$$

3. $$\int_{1-\varepsilon_{\mu p}}^{1}\frac{dx}{\sqrt{1-x^2}} = \pi/2 - \arcsin(1-\varepsilon_{\mu p}) = \frac{1}{\mu p}$$

we deduce

$$\int_{-1}^{1}\frac{v^2(x)}{\sqrt{1-x^2}}dx \leq 2\mu p\int_{-1+\varepsilon_{\mu p}}^{1-\varepsilon_{\mu p}} v^2(x)dx + 2||v(x)||^2_{L_\infty(I)}\int_{1-\varepsilon_{\mu p}}^{1}\frac{dx}{\sqrt{1-x^2}} \tag{D.5}$$

$$\leq 2\mu p\int_{-1+\varepsilon_{\mu p}}^{1-\varepsilon_{\mu p}} v^2(x)dx + \frac{2}{\mu p}||v(x)||^2_{L_\infty(I)} \tag{D.6}$$

Using Proposition 1 we can bound the latter integral as follows:

$$\int_{-1}^{1} \frac{v^2(x)}{\sqrt{1-x^2}} dx \leq 2\mu p \int_{-1}^{1} v^2(x) dx + \frac{4(p+1)}{\pi \mu p} \int_{-1}^{1} \frac{v^2(x)}{\sqrt{1-x^2}} dx.$$

Therefore for $\mu > \mu^*$, $\mu^* = \frac{8(p+1)}{\pi p}$:

$$\int_{-1}^{1} \frac{v^2(x)}{\sqrt{1-x^2}} \leq \left(\frac{2\mu p}{1 - \frac{4(p+1)}{\pi \mu p}} \right) \int_{-1}^{1} v^2(x) dx.$$

Note that

$$\min_{\mu \in (\mu^*, \infty)} \left(\frac{2\mu p}{1 - \frac{4(p+1)}{\pi \mu p}} \right) = \frac{32(p+1)}{\pi}$$

Finally,

$$\int_{-1}^{1} \frac{v^2(x)}{\sqrt{1-x^2}} dx \leq \frac{32(p+1)}{\pi} \int_{-1}^{1} v^2(x) dx.$$

□

Appendix E

Behaviour of the integrand in the transformed integrals.

We have shown in Chapter 4 that the double integral (4.11) can be written in the form,

$$I_k^{[\tau_0,\tau_{\max}]}[F] := \sum_{j=0}^{N-1} \int_{\tau_j}^{\tau_{j+1}} F_j(\tau)\exp(ik\tau)\mathrm{d}\tau, \tag{E.1}$$

with an integrand function F_j of the form

$$F_j(\tau) = \int_{r_{1,j}(\tau)}^{r_{2,j}(\tau)} \frac{M(s,\psi_{[s]}^{-1}(\tau))}{\left|\psi'_{[s]}(\psi_{[s]}^{-1}(\tau))\right|}\mathrm{d}s,$$

where $r_{1,j}(\tau)$ and $r_{2,j}(\tau)$ are the upper and lower boundaries defined in e.g. Table 4.2. The integral of the form (E.1) can be efficiently approximated using Filon-type quadrature. In order to determine the accuracy of such approximation, the regularity of the functions F_j must be known.

Let us introduce, for convenience, a function G,

$$G(s,t) := \frac{M(s,t)}{\psi'_{[s]}(t)} = M_1(s,t)\partial_{\boldsymbol{n}(s))}\Phi_k(s,t) + M_2(s,t)\Phi_k(s,t), = G_1(s,t) + G_2(s,t), \tag{E.2}$$

where functions $M_1(s,t)$ and $M_2(s,t)$ are smooth. Note that since Hypothesis A is satisfied, the function $\psi'_{[s]}(t)$ does not vanish and and the function $1/\psi'_{[s]}(t)$ is smooth. In the notation introduced in (E.2), the function $\psi'_{[s]}(t)$ has been absorbed in $M_1(s,t)$ and $M_2(s,t)$.

In this section, we return to the example from Section 4.3 where the original domain of integration is $(\Lambda_2 \times \Lambda_2)^+$, see (4.96),

$$\begin{aligned} J_k &:= \int_b^c \int_s^c M(s,t)\exp\left(ik\Psi(s,t)\right) dtds \\ &= \int_0^{\tau_1} F_1(\tau)\exp(ik\tau)d\tau + \int_{\tau_1}^{\tau_{\max}} F_2(\tau)\exp(ik\tau)d\tau, \end{aligned}$$

with $F_1 : [0, \tau_1] \to \mathbb{R}$ and $F_2 : [\tau_1, \tau_{\max}] \to \mathbb{R}$ defined as,

$$F_1(\tau) = \int_b^{r(\tau)} G\left(s, \psi_{[s]}^{-1}(\tau)\right) ds, \qquad F_2(\tau) = \int_{r_1(\tau)}^{r_2(\tau)} G\left(s, \psi_{[s]}^{-1}(\tau)\right) ds,$$

where,

$$r(\tau) = \left(\psi^{[c]}\right)^{-1}(\tau),$$

and

$$r_1(\tau) = \left(\psi^{[c]}\right)^{-1}(\tau) \in [b, \xi_c], \quad \text{and} \quad r_2(\tau) = \left(\psi^{[c]}\right)^{-1}(\tau) \in [\xi_c, c],$$

where ξ_c is defined in (4.95).

We will prove in this section that F_1 has a log-singularity at $\tau = 0$ and F_2 has a square-root singularity at $\tau = \tau_{\max}$. We do this in Section E.1 and Section E.1 respectively. We will also prove in Section E.2 that away from these singularities, functions F_1 and F_2 are smooth.

In the notation of (E.2), we can write

$$F_1(\tau) = \int_b^{r(\tau)} G_1(s, \psi_{[s]}^{-1}(\tau)) ds + \int_b^{r(\tau)} G_2(s, \psi_{[s]}^{-1}(\tau)) ds = A(\tau) + B(\tau). \tag{E.3}$$

E.1 Log-singularity

Lemma E.1 (Logarithmic singularity, single-layer case)**.** *The function $B(\tau)$ in (E.3), defined as*

$$B(\tau) = \int_b^{r(\tau)} G_2(s, \psi_{[s]}^{-1}(\tau)) ds = \int_b^{r(\tau)} M_2(s, \psi_{[s]}^{-1}(\tau)) \Phi(s, \psi_{[s]}^{-1}(\tau)) ds \tag{E.4}$$

where $\Phi_k(s,t)$ is the fundamental solution of the Helmholtz equation,

$$\Phi_k(s,t) = \frac{i}{4} H_0^{(1)}\left(k\left|\boldsymbol{\gamma}(s) - \boldsymbol{\gamma}(t)\right|\right), \tag{E.5}$$

can be written in the form

$$B(\tau) = \left(\mathcal{E}_1 \circ r_2\right)(\tau) + \left(\mathcal{E}_2 \circ r_2\right)(\tau) \log \tau, \tag{E.6}$$

where $\mathcal{E}_m(x) : [a,b] \to \mathbb{R}$, $m = 1, 2$, is defined as

$$\mathcal{E}_m(x) = \int_b^x E_m(s, \psi_{[s]}^{-1}(\phi(x)) \mathrm{d}s, \tag{E.7}$$

where $E_m(s,t)$, $m = 1, 2$, are smooth functions on $[\Lambda_l, \Lambda_j]$, $l, j = 1, 2, 3$.

Proof. We assume for the purpose of this example that the boundary Γ is analytic. We can write the function $\Phi_k(s,t)$ as follows:

$$\Phi_k(s,t) = \Phi_1(s,t) + \Phi_2(s,t)\log\left(4\sin^2\frac{s-t}{2}\right), \tag{E.8}$$

where $\Phi_1(s,t)$ and $\Phi_2(s,t)$ are analytic functions [34], since Γ is analytic. Now, for simplicity, we write the integral in (E.4) as

$$B(\tau) = \int_b^{r_2(\tau)} G_1(s,\psi_{[s]}^{-1}(\tau))ds, \tag{E.9}$$

where $G_1(s,t)$ can now be written using (E.8) as

$$G_1(s,t) = E_1(s,t) + E_2(s,t)\log\left(4\sin^2\frac{s-t}{2}\right), \tag{E.10}$$

where $E_1(s,t)$ and $E_2(s,t)$ are smooth functions. Then adding and subtracting $E_2(s,t)\log\tau^2$, we obtain

$$\begin{aligned} G_1(s,\psi_{[s]}^{-1}(\tau)) &= E_1(s,\psi_{[s]}^{-1}(\tau)) + E_2(s,\psi_{[s]}^{-1}(\tau))\log\left(\frac{4\sin^2\left(\frac{s-\psi_{[s]}^{-1}(\tau)}{2}\right)}{\tau^2}\right) \\ &+ E_2(s,\psi_{[s]}^{-1}(\tau))\log\tau^2. \end{aligned} \tag{E.11}$$

Then, we can prove that the second term in (E.11) is smooth by expanding the function $\psi_{[s]}^{-1}(\tau)$ in Taylor series around $\tau = 0$:

$$\begin{aligned} \psi_{[s]}^{-1}(\tau) &= \psi_{[s]}^{-1}(0) + \left(\psi_{[s]}^{-1}\right)'(0)\tau + \frac{1}{2}\left(\psi_{[s]}^{-1}\right)''(\xi)\tau\xi, \quad \text{where } \xi \in (0,\tau) \\ &= s + C_1\tau + C_2\tau\xi. \end{aligned}$$

Then,

$$\sin^2\left(s - \psi_{[s]}^{-1}(\tau)\right) = C\tau^2 + O(\tau^4).$$

Therefore, $4\sin^2\left(\frac{s-\psi_{[s]}^{-1}(\tau)}{2}\right)/\tau^2$ is a smooth function of $\tau \in [0,\tau_{\max}]$. Finally returning to the equation (E.11), we obtain

$$G_1(s,\psi_{[s]}^{-1}(\tau)) = \tilde{E}_1(s,\psi_{[s]}^{-1}(\tau)) + E_2(s,\psi_{[s]}^{-1}(\tau))\log\tau, \tag{E.12}$$

where $\tilde{E}_1(s,t)$ and $E_2(s,t)$ are smooth functions. Substituting this into (E.9), we deduce

$$B(\tau) = \int_b^{r_2(\tau)} \tilde{E}_1(s,\psi_{[s]}^{-1}(\tau))ds + \log\tau\int_b^{r_2(\tau)} E_2(s,\psi_{[s]}^{-1}(\tau))\mathrm{d}s. \tag{E.13}$$

This concludes the proof. □

Remark E.2. *Lemma E.1 proves that the function $B(\tau)$ defined in (E.3) has a log-singularity as $\tau \to \infty$. The function $A(\tau)$ also has a logarithmic singularity as $\tau \to 0$ since the kernel in the integrand function of A can be decomposed similarly to (E.10), [67]. Analogous result can be shown for the function A in a similar manner.*

Square-root singularity In this section we will consider the behaviour of the function F_2 at $\tau = \tau_{\max}$, where

$$\tau_{\max} = \psi^{[c]}(\xi_c)$$

with ξ_c defined in (4.95) as

$$\left(\phi^{[c]}\right)'(\xi_c) = 0, \qquad \left(\phi^{[c]}\right)''(\xi_c) \neq 0. \tag{E.14}$$

Proposition E.3. *The function $F_2(\tau)$ can be written as*

$$\begin{aligned} F_2(\tau) &= \int_{r_1(\tau)}^{\xi_c} G(s, \psi_{[s]}^{-1}(\tau))ds + \int_{\xi_c}^{r_2(\tau)} G(s, \psi_{[s]}^{-1}(\tau))ds, \\ &= \mathcal{G} \circ r_1(\tau) + \mathcal{G} \circ r_2(\tau), \end{aligned} \tag{E.15}$$

where $\mathcal{G}(x) : [\xi_c, b] \to \mathbb{R}$,

$$\mathcal{G}(x) = \int_{\xi_c}^{x} G(s, \psi_{[s]}^{-1}(\phi(x))ds. \tag{E.16}$$

Proof. From Lemma 4.11 and (E.14), it follows that functions $r_1(\tau)$ and $r_2(\tau)$, have a square-root singularity at $\tau = \tau_{\max}$ and then using Theorem 4.10 we conclude that F_2 has a square-root singularity at $\tau = \tau_{\max}$. □

In the next section, we prove that the function G in (E.16) and the functions $\mathcal{E}_m$, $m = 1, 2$ defined in (E.7) are smooth.

E.2 Smoothness of the integrand function away from singularities

The main result in this section is Theorem E.10, where we prove that the function $\mathcal{G}$ defined in (E.16) and the function $\mathcal{E}$ defined in (E.7) are smooth. We will prove Theorem E.7 using two lemmas that we present first. In Lemma E.5, we prove a simpler result for an operator $\mathcal{F}(x) := \int_{s_0}^{x} F(s, x)\mathrm{d}s$ with well behaved integrand function F. In Lemma E.6, we consider more complicated integrand: a composite of two well behaved functions. Finally, in the Theorem E.7, we specify the conditions on the integrand in order to ensure

the smoothness of the operator. Now what remains is to show that the integrands in $\mathcal{E}$ and $\mathcal{G}$ satisfy these conditions. We do this in Lemma E.9.

We introduce the following notation: let I_x, I_s and I_t be three intervals in $\mathbb{R}$ such that $I_x \subset I_s$. Let $F : I_s \times I_x \to \mathbb{R}$ be given. We denote the n-th derivative of F with respect to its j-th argument as $\partial_j^n F$. Similarly, we denote the n-th derivative of F with respect to s variable as $\partial_s^n F$ and t variable as $\partial_t^n F$.

Definition E.4. *We say that $F(s,x)$ satisfies the* ***condition A*** *if there exist constants A_n and $A'_{n,m}$, such that*

$$\sup_{\substack{s\in I_s\\ x\in I_x}} \left|\left(\partial_2^n F\right)(s,x)\right| \leq A_n, \tag{A1}$$

$$\sup_{x\in I_x} \left|\frac{d^m}{dx^m}\left[\left(\partial_2^n F\right)(x,x)\right]\right| \leq A'_{n,m}. \tag{A2}$$

Lemma E.5. *Suppose that $I_x \subset I_s$, and condition A holds. For $s_0 \in I_s$, $x \in I_x \subset I_s$ define*

$$\mathcal{F}(x) = \int_{s_0}^{x} F(s,x)ds.$$

Then $\mathcal{F}$ is infinitely continuously differentiable and for $n \geq 1$,

$$\left(\frac{d^n\mathcal{F}}{dx^n}\right)(x) = \sum_{j=0}^{n-1} \frac{d^{n-j-1}}{dx^{n-j-1}}\left(\left(\partial_2^j F\right)(x,x)\right) + \int_{s_0}^{x} \left(\partial_2^n F\right)(s,x)ds. \tag{E.18}$$

Proof. We begin by deriving the formula (E.18). The first derivative of $\mathcal{F}$ is

$$\begin{aligned}
\frac{d\mathcal{F}}{dx}(x) &= \lim_{\Delta x\to 0}\frac{\mathcal{F}(x+\Delta x)-\mathcal{F}(x)}{\Delta x} = \lim_{\Delta x\to 0}\frac{\int_{s_0}^{x+\Delta x} F(s,x+\Delta x)ds - \int_{s_0}^{x} F(s,x)ds}{\Delta x}\\
&= \lim_{\Delta x\to 0}\frac{\int_{x}^{x+\Delta x} F(s,x+\Delta x)ds}{\Delta x} - \lim_{\Delta x\to 0}\frac{\int_{s_0}^{x}\left(F(s,x+\Delta x)-F(s,x)\right)ds}{\Delta x}\\
&= F(x,x) + \int_{s_0}^{x}\left(\partial_2^1 F\right)(s,x)ds.
\end{aligned}$$

The formula (E.18) then follows by induction. Since $F(s,x)$ satisfies the condition A, $\frac{d^n\mathcal{F}}{dx^n}(x)$ is bounded for all $n \geq 1$, $x \in I_x$. Since all of the derivatives of $\mathcal{F}(x)$ are bounded, $\mathcal{F}(x)$ is infinitely continuously differentiable. □

Now consider the following example:

$$F(s,x) = G\left(s,\eta(s,x)\right), \tag{E.19}$$

where the functions $G : I_s \times I_t \to \mathbb{R}$ and $\eta : I_s \times I_x \to I_t$ satisfy the following conditions,

$$\sup_{\substack{s\in I_s\\ t\in I_t}} |(\partial_1^m \partial_2^n G)(s,t)| \leq B_{n,m}, \tag{B1}$$

and

$$\eta(x,x) = b = const \quad \text{for all} \quad x \in I_x, \tag{C1}$$

$$\sup_{\substack{s\in I_s\\ x\in I_x}} |(\partial_2^n \eta)(s,x)| \leq C_n, \tag{C2}$$

$$\sup_{x\in I_x} \left| \frac{d^m}{dx^m} \left[(\partial_2^n \eta)(x,x) \right] \right| \leq C'_{n,m}. \tag{C3}$$

Lemma E.6. *If $G(s,t)$ satisfies (B1) and $\eta(s,x)$ satisfies (C1), (C2) and (C3), then $F(s,x)$, defined in (E.19), satisfies condition A.*

Proof. Using Faa di Bruno's formula, we deduce that

$$\begin{aligned}(\partial_2^n F)(s,x) &= \left(\frac{\partial^n}{\partial x^n}\right)[G(s,\eta(s,x))] \\ &= \sum_{j=1}^{n} \left(\partial_2^j G\right)(s,\eta(s,x)) \mathcal{B}_{n,j}\left((\partial_2^1\eta)(s,x),\ldots,(\partial_2^{n-j+1}\eta)(s,x)\right),\end{aligned}$$

where $\mathcal{B}_{n,j}$ is the Bell Polynomial [11], whose coefficients are finite. From (B1) and (C2), we deduce that (A1) holds.

On the other hand, using (C1), we deduce that

$$(\partial_2^n F)(x,x) = \sum_{j=1}^{n} \left(\partial_2^j G\right)(x,b) P_{n,j}\left((\partial_2^1\eta)(x,x),\ldots,(\partial_2^{n-j+1}\eta)(x,x)\right). \tag{E.22}$$

Now, the function $(\partial_2^n F)(x,x)$ is a smooth function with respect to the variable x since (B1) and (C3) hold. Thus (A2) is satisfied and condition A holds. □

Theorem E.7. *Consider the function $\mathcal{G}(x)$, defined as*

$$\mathcal{G}(x) = \int_{s_0}^{x} G(s,\eta(s,x))\, ds, \tag{E.23}$$

where $G(s,t)$ and $\eta(s,x)$ satisfy (B1) and (C1), (C2), (C3). Then

$$\mathcal{G} \in \mathcal{C}^\infty(I_x). \tag{E.24}$$

Proof. From Lemma E.6, we deduce that $F(s,x) = G(s,\eta(s,x))$, satisfies the condition A. Therefore, by Lemma E.5, $\mathcal{G}$ is infinitely continuously differentiable. □

Definition E.8. *We define the function $\psi_{[s]} : I_t \to [\tau_0, \tau_{\max}]$ as follows: for a fixed $s \in I_s$,*

$$\psi_{[s]}(t) := \Psi(s,t). \tag{E.25}$$

In the following lemma, we prove that the function $\eta(s,x) = \psi_{[s]}(\phi(x))$ satisfies conditions (C1), (C2) and (C3).

Lemma E.9. *For $\phi : I_x \to [\tau_0, \tau_{\max}]$ such that $\phi(x) := \psi_{[s]}(b)$ where $\psi_{[s]} : I_t \to [\tau_0, \tau_{\max}]$, $s \in I_s$ is defined in (E.25), define the function $\eta : I_s \times I_x \to I_t$ as follows:*

$$\eta(s,x) = \psi_{[s]}^{-1}\left(\phi(x)\right). \tag{E.26}$$

Then, $\eta(s,x)$ satisfies (C1), (C2) and (C3), provided $\psi_{[s]}^{'}(t) \neq 0$.

Proof. **Condition (C1).** We begin with the identity

$$\psi_{[s]}^{-1}\left(\psi_{[s]}(t)\right) = t, \iff \psi_{[s]}^{-1}\left(\psi_{[s]}(b)\right) = b, \iff \psi_{[s]}^{-1}\left(\phi(s)\right) = b.$$

The final equation is obtained by the definition of $\psi_{[s]}(t)$ and $\phi(s)$. Substituting $s = x$, we obtain the result.

Condition (C2). Assume that $\psi_{[s]}^{'}(t) \neq 0$, $s \in I_s$, $t \in I_t$. We know that the phase-function $\Psi(s,t)$, defined in (4.12), satisfies (B1), provided the boundary Γ is smooth. Therefore $\psi_{[s]}(t)$ is continuously differentiable and has non-zero derivative. By the inverse function theorem, the inverse of $\psi_{[s]}(t)$ is also continuously differentiable. Then for fixed s, $\eta(s,x)$ is a composition of two continuously differentiable functions and therefore must satisfy $(C2)$.

Condition (C3). Consider the function $t = \psi_{[s]}^{-1}(\tau)$. The first derivative is continuous:

$$(\psi_{[s]}^{-1})'(\tau) = \frac{1}{(\partial_2^1\Psi)(s, \psi_{[s]}^{-1}(\tau))}.$$

If we differentiate this function again, we obtain:

$$\left(\psi_{[s]}^{-1})''(\tau)\right) = \partial_2^1\left(\frac{1}{(\partial_2^1\Psi)(s, \psi_{[s]}^{-1}(\tau))}\right) = -\frac{\left[(\partial_2^2\Psi)(s, \psi_{[s]}^{-1}(\tau))\right]}{\left[\psi_{[s]}^{'}(\psi_{[s]}^{-1}(\tau))\right]^2},$$

which is bounded. Similarly, for $j \geq 1$, $(\psi_{[s]}^{-1})(j)(\tau)$ is equal to: using Faa di Bruno's

formula,

$$\left(\psi_{[s]}^{-1}\right)^{(n+1)}(\tau) = \sum_{j=1}^{n} \partial_2^j \left(\frac{1}{(\partial_2^1\Psi)(s,\psi_{[s]}^{-1}(\tau))}\right) \mathcal{B}_{n,j}\left((\psi_{[s]}^{-1})'(\tau),\ldots,(\psi_{[s]}^{-1})^{(n-j+1)}(\tau)\right). \tag{E.27}$$

Thus, if the first derivative of $\psi_{[s]}^{-1}(\tau)$ is bounded, the all derivatives are also bounded. This can be shown by induction using (E.27).

Also using Faa di Bruno's formula, we deduce

$$\left(\partial_2^n\eta\right)(s,x) = \sum_{j=1}^{n} \left(\psi_{[s]}^{-1}\right)^{(j)}(\phi(x))\mathcal{B}_{n,j}\left(\phi'(x),\ldots,\phi^{n-j+1}(x)\right). \tag{E.28}$$

Since $\phi(x)$ is a smooth and bounded function, we conclude that the condition (C3) holds. □

Theorem E.10. *Functions $\mathcal{G}(x)$ defined in (E.16) and $\mathcal{E}(x)$ defined in (E.7) are smooth.*

Proof. The condition (B1) is satisfied by the definitions of $G(s,t)$ and $E(s,t)$ in (E.16) and (E.7). Therefore, the result follows from Theorem E.7 and Lemma E.9. □

Appendix F

Proof of Lemma 4.8

The proof of Lemma 4.8 requires an intermediate result which we prove in Lemma F.1.

Lemma F.1. *Let*

$$g(\boldsymbol{x}) = |\boldsymbol{x}|, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

For each $\boldsymbol{p} \in \mathbb{N}^d$ *there exist functions* a_j *and* $b_j \in C^\infty\left(\mathbb{R}^d\right)$ *such that*

$$\left(D_{\boldsymbol{x}}^{\boldsymbol{p}} g\right)(\boldsymbol{x}) = \sum_{j=0}^{|\boldsymbol{p}|} a_j(\boldsymbol{x}) b_j\left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right)|\boldsymbol{x}|^{1-j},$$

where $D_{\boldsymbol{x}}^{\boldsymbol{p}}$ *denotes any partial derivative of order* $|\boldsymbol{p}|$ *with respect to* $\boldsymbol{x}$*.*

Hence if $\boldsymbol{x}$ *ranges over a bounded domain in* $\mathbb{R}^d$*,*

$$\left|\left(D_{\boldsymbol{x}}^{\boldsymbol{p}} g\right)(\boldsymbol{x})\right| \leq C_{\boldsymbol{p}}\left(|\boldsymbol{x}|^{1-|\boldsymbol{p}|}\right), \tag{F.1}$$

with $C_{\boldsymbol{p}}$ *independent of* $\boldsymbol{x}$*.*

Proof. The proof follows by induction on $n := |\boldsymbol{p}|$. For $n = 0$ (F.1) holds because

$$\left(D_{\boldsymbol{x}}^{\boldsymbol{p}} g\right)(\boldsymbol{x}) = g(x) = |\boldsymbol{x}| = |\boldsymbol{x}|^{1-0}.$$

Now suppose that (F.1) is true for $|\boldsymbol{p}| = n$. Then if $\boldsymbol{p} = n + 1$, we have

$$D^{\boldsymbol{p}} = \frac{\partial}{\partial x_i} D^{\boldsymbol{p}'},$$

for some $|\boldsymbol{p}'| = n$ and some $i = 1, 2, 3$. Then

$$\left(D_{\boldsymbol{x}}^{\boldsymbol{p}} g\right)(\boldsymbol{x}) = \frac{\partial}{\partial x_i}\left\{\sum_{j=0}^{n} a_j(\boldsymbol{x}) b_j\left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right)|\boldsymbol{x}|^{1-j}\right\}.$$

Then the result follows since

$$\begin{aligned}
\frac{\partial}{\partial x_i}\left\{b_j\left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right)\right\} &= \sum_{k=1}^{d}\frac{\partial b_j}{\partial x_k}\left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right)\frac{\partial}{\partial x_i}\left(\frac{x_k}{|\boldsymbol{x}|}\right) \\
&= \sum_{k=1}^{d}\frac{\partial b_j}{\partial x_k}\left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right)\left(\delta_{ik}-\frac{x_i x_k}{|\boldsymbol{x}|^3}\right) \\
&= \frac{\partial b_j}{\partial x_i}\left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right)-\sum_{k=1}^{d}\left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right)\frac{\partial b_j}{\partial x_k}\left(\frac{x_i}{|\boldsymbol{x}|}\right)\left(\frac{x_k}{|\boldsymbol{x}|}\right)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial x_i}\left\{|x|^{1-j}\right\} &= x_i|\boldsymbol{x}|^{-j-1}=\frac{x_i}{|\boldsymbol{x}|}|\boldsymbol{x}|^{-j} \\
&= \frac{x_i}{|\boldsymbol{x}|}|\boldsymbol{x}|^{1-(j+1)}.
\end{aligned}$$

□

Proof of Lemma 4.8

Proof. Since

$$r(s,t)=g(\boldsymbol{\gamma}(s)-\boldsymbol{\gamma}(t)),$$

we have by Lemma F.1,

$$\left(D^{\boldsymbol{p}}_{(s,t)}r\right)(\boldsymbol{x})=\left(\frac{\partial}{\partial s}\right)^{p_1}\left(\frac{\partial}{\partial t}\right)^{p_2}\left\{g(\boldsymbol{\gamma}(s)-\boldsymbol{\gamma}(t))\right\}, \tag{F.2}$$

where

$$D^{\boldsymbol{p}}_{(s,t)}r(s,t):=\frac{\partial^{|\boldsymbol{p}|}}{\partial s^{p_1}\partial t^{p_2}}\ (r(s,t)).$$

The right hand side of (F.2) is a linear combination of

$$\left(D^{\boldsymbol{q}}_{\boldsymbol{x}}g\right)(\boldsymbol{\gamma}(s)-\boldsymbol{\gamma}(t)),$$

for $|\boldsymbol{q}|\leq|\boldsymbol{p}|$ with smooth coefficients. Hence the result follows from Lemma F.1. □

Appendix G

Proof of Proposition 3.22

Proof. We begin by writing out the equation (3.117) in terms of the function W which represents the Airy function Ai or Bi,

$$\left(\Delta + k^2\right)\left(W\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right)\mathcal{A}(\boldsymbol{x},\gamma,k) + ik^{-1/3}W'\left(-k^{2/3}\mu(\boldsymbol{x},\gamma)\right)\mathcal{B}(\boldsymbol{x},\gamma,k)\right)e^{ik\xi(\boldsymbol{x},\gamma)} = 0.$$

First, recall that $\Delta(fg) = \nabla\cdot\nabla(fg) = g\Delta f + 2\nabla f\cdot\nabla g + f\Delta g$, therefore

$$\left(\Delta + k^2\right)e^{ik\xi}D = e^{ik\xi}\left(\left(\Delta + k^2\right)D + 2ik\nabla D\cdot\nabla\xi - k^2|\nabla\xi|^2 D + ik(\Delta\xi)D\right) = 0. \quad \text{(G.1)}$$

Further, remembering that the Airy function W satisfies the Airy equation (3.23), we deduce

$$\begin{aligned}
\nabla D &= \sum_{j=0}^{N} k^{-j}\left(W\left(-k^{2/3}\mu\right)\nabla A_j - k^{2/3}W'\left(-k^{2/3}\mu\right)\nabla\mu A_j\right)\\
&+ ik^{-1/3}\sum_{j=0}^{N} k^{-j}\left(W'\left(-k^{2/3}\mu\right)\nabla B_j - k^{2/3}W''\left(-k^{2/3}\mu\right)\nabla\mu B_j\right)\\
&= \sum_{j=0}^{N} k^{-j}\left(W\left(-k^{2/3}\mu\right)\nabla A_j - k^{2/3}W'\left(-k^{2/3}\mu\right)\nabla\mu A_j\right)\\
&+ ik^{-1/3}\sum_{j=0}^{N} k^{-j}\left(W'\left(-k^{2/3}\mu\right)\nabla B_j + k^{4/3}\mu W\left(-k^{2/3}\mu\right)\nabla\mu B_j\right)
\end{aligned}$$

and

$$\begin{aligned}
\Delta D &= \sum_{j=0}^{N} k^{-j}\Big[W\left(-k^{2/3}\mu\right)\Delta A_j - 2k^{2/3}W'\left(-k^{2/3}\mu\right)(\nabla A_j\cdot\nabla\mu)\\
&\qquad - k^{2/3}W'\left(-k^{2/3}\mu\right)\Delta\mu A_j - k^2\mu W\left(-k^{2/3}\mu\right)A_j|\nabla\mu|^2\Big]\\
&+ ik^{-1/3}\sum_{j=0}^{N} k^{-j}\Big[W'\left(-k^{2/3}\mu\right)\Delta B_j + 2k^{4/3}W\left(-k^{2/3}\mu\right)\mu(\nabla B_j\cdot\nabla\mu)\\
&\qquad + k^{4/3}W\left(-k^{2/3}\mu\right)|\nabla\mu|^2 B_j + k^{4/3}\mu W\left(-k^{2/3}\mu\right)\Delta\mu B_j\\
&\qquad -k^2\mu W'\left(-k^{2/3}\mu\right)|\nabla\mu|^2 B_j\Big].
\end{aligned}$$

Substituting ∇D and ΔD into equation (G.1),we obtain

$$
\begin{aligned}
e^{-ik\xi}\left(\left(\Delta+k^2\right)\left[e^{ik\xi}D\right]\right) = \sum_{j=0}^{N} k^{-j}\Big[&W\left(-k^{2/3}\mu\right)\Delta A_j - 2k^{2/3}W'\left(-k^{2/3}\mu\right)(\nabla A_j\cdot\nabla\mu) \\
&- k^{2/3}W'\left(-k^{2/3}\mu\right)\Delta\mu A_j - k^2\mu W\left(-k^{2/3}\mu\right)A_j|\nabla\mu|^2\Big] \\
&+ ik^{-1/3}\sum_{j=0}^{N} k^{-j}\Big[W'\left(-k^{2/3}\mu\right)\Delta B_j + 2k^{4/3}W\left(-k^{2/3}\mu\right)\mu(\nabla B_j\cdot\nabla\mu) \\
&+ k^{4/3}W\left(-k^{2/3}\mu\right)|\nabla\mu|^2 B_j + k^{4/3}\mu W\left(-k^{2/3}\mu\right)\Delta\mu B_j - k^2\mu W'\left(-k^{2/3}\mu\right)|\nabla\mu|^2 B_j\Big] \\
&\sum_{j=0}^{N} k^{-j}\left(k^2\left(1-|\nabla\xi|^2\right)+ik\Delta\xi\right)\Big[W\left(-k^{2/3}\mu\right)A_j + ik^{-1/3}W'\left(-k^{2/3}\mu\right)B_j\Big] \\
&+ 2ik\sum_{j=0}^{N} k^{-j}\left(W\left(-k^{2/3}\mu\right)\nabla A_j\cdot\nabla\xi - k^{2/3}W'\left(-k^{2/3}\mu\right)\nabla\mu\cdot\nabla\xi A_j\right) \\
&+ (2ik)ik^{-1/3}\sum_{j=0}^{N} k^{-j}\left(W'\left(-k^{2/3}\mu\right)\nabla B_j\cdot\nabla\xi - k^{4/3}\mu W\left(-k^{2/3}\mu\right)\nabla\mu\cdot\nabla\xi B_j\right),
\end{aligned}
$$

and comparing coefficients in k^2 and $k^{2-1/3}$, we obtain:

$$
\begin{cases}
\left(|\nabla\xi|^2+\mu|\nabla\mu|^2-1\right)A_0 - 2(\nabla\mu\cdot\nabla\xi)\mu B_0 = 0, \\
\left(|\nabla\xi|^2+\mu|\nabla\mu|^2-1\right)B_0 + 2(\nabla\mu\cdot\nabla\xi)A_0 = 0.
\end{cases}
$$

In order to make sure that $A_0(\boldsymbol{x},\gamma)\neq 0$ and $B_0(\boldsymbol{x},\gamma)\neq 0$, we require $\mu(\boldsymbol{x},\gamma)$ and $\xi(\boldsymbol{x},\gamma)$ to satisfy (3.118). □