Numerical methods for the wave equation with uncertainty

BENJAMIN DAVID WEBER
Numerical methods for the wave equation with uncertainty

BENJAMIN DAVID WEBER
Abstract

The problem of uncertainty quantification for the wave equation with random scatterers is explained and motivated. The background of various methods for solving hyperbolic partial differential equations is discussed, and the details of the numerical solution method used are explained. A brief discussion of uncertainty quantification (UQ) is included with a description of three different sampling methods and a precise description of the UQ problem investigated. Numerical results for different families of random scatterers are presented, including regularity tests and convergence analysis of the different sampling methods.
Sammanfattning

# Contents

1 Introduction
   1.1 Overview of the problem ............................................. 1
   1.2 Motivation ............................................................. 1
   1.3 Related literature .................................................... 1
   1.4 Overview of the methodology ....................................... 2
   1.5 Overview of the results ............................................. 3
   1.6 Layout of paper ...................................................... 3

2 Solving the wave equation .............................................. 4
   2.1 Governing equation .................................................... 4
      2.1.1 The wave equation for scattering in $2D$ ........................ 4
      2.1.2 The wave equation with a PML boundary ....................... 7
   2.2 Numerical methods ................................................... 8
      2.2.1 Review of numerical solution methods .......................... 9
   2.3 Finite difference discretisation .................................... 12
      2.3.1 Notation .......................................................... 12
      2.3.2 4th order finite difference discretisation in space ............ 13
      2.3.3 Implementation of complex Dirichlet boundaries ............... 14
      2.3.4 Time discretisation ............................................. 16
      2.3.5 Implementation of PML ......................................... 18
   2.4 Validation of methods .............................................. 20
      2.4.1 Validation of basic solver ..................................... 20
      2.4.2 Validation of PML implementation ............................ 24

3 Uncertainty Quantification (UQ) ...................................... 27
   3.1 Overview of UQ ...................................................... 27
   3.2 Review of UQ methods ............................................... 28
      3.2.1 Overview of methods ........................................... 28
   3.3 Description of the problem in UQ ................................. 31
Chapter 1

Introduction

1.1 Overview of the problem

This paper focuses on the problem of scattering of plane-wave Gaussian pulses in two spatial dimensions for the wave equation from an object of uncertain shape. The uncertainty of said shape is represented by one or more random variables capable of taking a relatively large range of values, resulting in large families of shapes in which large variations in shape and size are possible.

1.2 Motivation

The motivation for approaching this problem is the broad range of subjects in which wave scattering occurs. In these fields, it is often the pattern of scattered energy that is measured. However, in order for these measurements to be of any use in determining the shape of the scatterer, knowledge of how uncertainty in the shape of the scatterer translates to changes in the scattered energy is required. This paper explores how relatively large uncertainty in the shapes of families of scatterers can affect the scattering pattern of the energy and its probability distribution.

1.3 Related literature

Much of existing literature in the field of uncertainty quantification for the wave equation falls into one or more of the following categories:

- Theoretical results pertaining to simple domains, such as in one dimension
• Theoretical analysis based upon small perturbations from a classical system
• Numerical analysis of systems in which the coefficients of the equation are uncertain
• Numerical analysis of random scatterers for single frequencies using the Helmholtz equation

A lot of existing literature in the field of uncertainty quantification for the wave equation has focussed more on theoretical results rather than computational results. For instance, Gao (2017) focuses on a relatively simple domain in order to produce analytical results about the system and its distribution. Other literature has approached the problem of a random wave-speed, rather than a random scatterer, such as Motamed, Nobile, Tempone (2013). In the field of random scatterers, a common approach is to work computationally with a single frequency, as in the paper by Tsuji, Xiu and Ying (2011). There work was also limited to relatively simple shapes with small perturbations. This leaves an area of research that this paper will begin to investigate; namely what happens when larger perturbations are allowed and more complex shapes are considered for a more generalised incoming wave?

1.4 Overview of the methodology

The methodology followed by this paper is as follows:

1. Describe the uncertainty in shape of a type of scatterer with a level-set function defined by a finite number of independent uniformly distributed random variables \( \{ X_1, \ldots, X_n \} \),

2. Use one of the sampling methods outlined in section 3 to select sample points,

3. Evaluate the numerical solution to the classical wave system described by each sample point up to a time \( T \) by which the scattering profile should have fully formed, using the numerical method outlined in section 2,

4. Approximate the scattered energy in each direction using the calculated numerical solution (see section 3.3 for details),
5. Calculate the expected value and variance of the scattering in each direction, and analyse convergence for a small number of selected angles.

All numerical experiments performed used MATLAB R2017a. The code was run on a laptop with the Windows 10 operating system and a 2.40GHz Intel Core i3-7100U CPU.

Note: The notation d.p. is used throughout this paper for 'decimal places' to denote the level of accuracy used for writing numbers in decimal form.

1.5 Overview of the results

It was found that sampling with around 2000 points was typically sufficient to see convergence with Monte Carlo sampling to an accuracy comparable to the accuracy of the numerical solution to each system, however the accuracy and convergence of other sampling methods was more dependent on the type of shape used and the number of random parameters defining it. Shapes with an increasing number of random parameters but retaining certain characteristics were found to be able to converge to a fixed probability distribution for scattering.

1.6 Layout of paper

This paper is divided into four main sections. In section 2, the governing differential equations for the system are described, and the numerical methods used to solve them are described in detail, including a discussion of stability and accuracy. In section 3, the problem as it relates to uncertainty quantification is defined in more precise detail, with a discussion of how the uncertainty was introduced to the system and the precise quantities of interest being analysed. In section 4 the different cases of the problem are defined, and the numerical results are presented with conclusions. In section 5 conclusions of the investigation are presented and summarised, together with suggestions for further research.
Chapter 2

Solving the wave equation

In this chapter the governing equations for the problem, including for an absorbing boundary condition in the form of a perfectly matched layer (PML) boundary, will be outlined, and a detailed description of the finite difference discretisation used to calculate solutions given, with a discussion of stability and experimental validation of the techniques.

2.1 Governing equation

2.1.1 The wave equation for scattering in $2D$

The equation at the core of this problem is the unforced wave equation with a constant wave-speed $c$ in $\mathbb{R}^2$ on the exterior of a bounded region $\Omega$ with Dirichlet boundary $\partial\Omega$ up to a time $T$:

$$\frac{\partial^2 u}{\partial t^2} - c^2 \Delta u = 0, \quad (x, y, t) \in \mathbb{R}^2 \setminus \Omega \times [0, T),$$

$$u(x, y, 0) = u_0(x, y), \quad (x, y) \in \mathbb{R}^2 \setminus \Omega,$$

$$\frac{\partial u}{\partial t} \bigg|_{(x,y,0)} = v_0(x, y), \quad (x, y) \in \mathbb{R}^2 \setminus \Omega,$$

$$u(x, y, t) = g(x, y, t), \quad (x, y, t) \in \partial\Omega \times [0, T),$$

where $(x, y)$ are cartesian coordinates, $t$ is the time, $u_0$ and $v_0$ describe the initial values of the solution function $u$ and its temporal derivative, and $g$ is the value of the solution function on the boundary $\partial\Omega$. As an example of a
system as described above, below a diagram of the domain is included when the object \( \Omega \) is a unit circle:

![Diagram of the domain for the wave equation on the exterior of an object. Here the object \( \Omega \) is shown in red and the exterior region where the solution \( u \) exists is shown in green.](image)

Since the focus of this paper is to analyse the scattering of a plane wave from the object represented by region \( \Omega \), it is necessary to formulate a related problem for the scattering field based on the solution \( u_{\text{wave}} \) in the absence of the object. In this case, scattering will refer to all changes in the computed solution due to the presence of the object \( \Omega \). As a consequence, this will include both reflections from the object and 'shadows' due to the object preventing the wave from entering a certain region. To do this, first consider the full solution \( u = u_{\text{scatter}} + u_{\text{wave}} \) on \( \mathbb{R}^2 \setminus \Omega \times [0, T) \), assume no scattering has occurred at a time \( t = 0 \) and apply a homogeneous Dirichlet boundary:

\[
\frac{\partial^2}{\partial t^2} (u_{\text{scatter}} + u_{\text{wave}}) = c^2 \Delta (u_{\text{scatter}} + u_{\text{wave}}), \quad (x, y, t) \in \mathbb{R}^2 \setminus \Omega \times [0, T),
\]

\[
(u_{\text{scatter}}(x, y, 0) + u_{\text{wave}}(x, y, 0)) = u_{\text{wave}}(x, y, 0), \quad (x, y) \in \mathbb{R}^2 \setminus \Omega,
\]

\[
\left. \frac{\partial}{\partial t} (u_{\text{scatter}} + u_{\text{wave}}) \right|_{(x,y,0)} = \left. \frac{\partial u_{\text{wave}}}{\partial t} \right|_{(x,y,0)}, \quad (x, y) \in \mathbb{R}^2 \setminus \Omega,
\]

\[
u_{\text{wave}}(x, y, t) + u_{\text{scatter}}(x, y, t) = 0, \quad (x, y, t) \in \partial \Omega \times [0, T).
\]
For the wave equation in 2D with a constant wave-speed on $\mathbb{R}^2$, there exists solutions of the form $u_{\text{wave}}(x, y, t) = U(x - ct)$, representing plane waves moving in the direction of $x$ increasing, where $U : \mathbb{R} \rightarrow \mathbb{R}$ is a twice differentiable function describing the wave’s profile. Since the function $u_{\text{wave}}$ obeys the wave equation for all $\mathbb{R}^2$, by the linearity of the differential operators the following problem is obtained for $u_{\text{scatter}}$:

\[
\frac{\partial^2 u_{\text{scatter}}}{\partial t^2} - c^2 \Delta u_{\text{scatter}} = 0, \quad (x, y, t) \in \mathbb{R}^2 \setminus \Omega \times [0, T),
\]

\[u_{\text{scatter}}(x, y, 0) = 0, \quad (x, y) \in \mathbb{R}^2 \setminus \Omega,\]

\[\frac{\partial u_{\text{scatter}}}{\partial t} \bigg|_{(x, y, 0)} = 0, \quad (x, y) \in \mathbb{R}^2 \setminus \Omega,\]

\[u_{\text{scatter}}(x, y, t) = -U(x - ct), \quad (x, y, t) \in \partial \Omega \times [0, T).\]

As an example of what the above refers to in practice, below are three plots, showing the solution for a Gaussian pulse, with its peak starting at $x = -2$, on the geometry shown in figure 1 at time $t = 2$. The first image is the computed solution in the absence of the scatterer, the second is the full solution and the third is the scattering only:
Figure 2.2: Image plots showing the computed solution of a plane wave Gaussian pulse moving to the right, for the cases when a circular object is absent and present, as well as a plot of only the scattering when present.

From this point on in this paper, the function \( u(x, y, t) \) will refer to the solution \( u_{\text{scatter}} \) of the above system, and \( U \) will refer to the incoming wave.

### 2.1.2 The wave equation with a PML boundary

In the above section, the wave equation for the exterior scattering field from an object \( \Omega \) was formulated. However, the system was given for an infinite domain, which cannot be realistically modelled with a finite number of numerical computations. To avoid this, a perfectly matched layer (PML) surrounding the region of interest \( \Lambda \) can be added. This heavily damps all waves within a specified region \( \Sigma \) while producing virtually no reflections at the boundary \( \partial \Sigma \) between the PML region and the undamped region \( \Lambda \). Using the derivation and formulation of the problem in two spatial dimensions by Grote and Sim (2010) gives the modified system:
\[
\begin{align*}
\frac{\partial^2 u}{\partial t^2} + (\zeta_1 + \zeta_2) \frac{\partial u}{\partial t} + \zeta_1 \zeta_2 u &= c^2 \Delta u + \frac{\partial \phi_1}{\partial x} + \frac{\partial \phi_2}{\partial y}, \\
\frac{\partial \phi_1}{\partial t} &= -\zeta_1 \phi_1 + c^2 (\zeta_2 - \zeta_1) \frac{\partial u}{\partial x}, \\
\frac{\partial \phi_2}{\partial t} &= -\zeta_2 \phi_2 + c^2 (\zeta_1 - \zeta_2) \frac{\partial u}{\partial y}, \\
u(x, y, 0) &= \frac{\partial u}{\partial t} \bigg|_{(x,y,0)} = 0, \\
\phi_1(x, y, 0) &= \phi_2(x, y, 0) = 0,
\end{align*}
\] , \quad (x, y, t) \in \Lambda \cup \Sigma \setminus \Omega \times [0, T),

where $\phi_1, \phi_2$ are auxiliary variables, $\zeta_1(x), \zeta_2(y)$ are variable coefficients describing the damping profile and $\Gamma$ is the exterior boundary of the region $\Lambda \cup \Sigma$. Stability of this system is proven in the paper by Grote and Sim (2010).

In order to set the damping to zero within the region $\Lambda$ and to reduce reflections due to numerical error when discretising on a finite difference grid, the coefficient functions $\zeta_1, \zeta_2$ should be defined so as to be zero in the region $\Lambda$, and twice differentiable on the boundary $\partial \Sigma$.

From this point on, the domain in which solutions are computed will be denoted $D$, where:

\[ D := \Lambda \cup \Sigma \setminus \Omega. \]

### 2.2 Numerical methods

The system that needs to be solved for this paper is the wave equation in $2D$ with a complex boundary; that is the boundary may have features that make it difficult to approximate numerically, such as large surface curvature, disjoint segments or concave sections. Solutions to systems with such complex boundaries are required for this paper because the aim is to analyse how relatively large amounts of uncertainty about the shape of a scatterer affects the scattering profile.
2.2.1 Review of numerical solution methods

The simplest approach, and the one adapted to this system below, is the finite difference method. The concept here is to first approximate spatial derivatives in the problem directly by taking the approximate solution at different times on a rectangular grid, then rewriting the system as a system of ordinary differential equations at each grid-point. At each grid-point, the values at nearby grid-points are expressed as Taylor expansions of the solution function and its derivatives at the grid-point being considered. A linear combination of these expansions is then chosen to eliminate a certain number of low order terms except for the desired derivative.

To better illustrate this approach, an example of this method of approximating derivatives is presented in 1D. The function \( f \) is approximated at grid-points \( x_i = x_0 + i \cdot h \), where \( x_0 \) is some constant, \( h \) is the distance between grid-points and \( i \) is an index for the grid-points. Taylor expansions around the point \( x_i \) for \( u(x_{i-1}) \), \( u(x_{i+1}) \) give:

\[
\begin{align*}
  u(x_{i-1}) = u(x_i) - h \frac{du}{dx} \bigg|_{x=x_i} + \frac{1}{2} h^2 \frac{d^2 u}{dx^2} \bigg|_{x=x_i} - \frac{1}{3!} h^3 \frac{d^3 u}{dx^3} \bigg|_{x=x_i} + O(h^4), \\
  u(x_{i+1}) = u(x_i) + h \frac{du}{dx} \bigg|_{x=x_i} + \frac{1}{2} h^2 \frac{d^2 u}{dx^2} \bigg|_{x=x_i} + \frac{1}{3!} h^3 \frac{d^3 u}{dx^3} \bigg|_{x=x_i} + O(h^4).
\end{align*}
\]

This gives the following approximations for the first and second derivatives of \( u \) at \( x_i \):

\[
\begin{align*}
  \frac{du}{dx} \bigg|_{x=x_i} = \frac{u(x_{i+1}) - u(x_{i-1})}{2h} + O(h^2), \\
  \frac{d^2 u}{dx^2} \bigg|_{x=x_i} = \frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1})}{h^2} + O(h^2).
\end{align*}
\]

With this approximation of the spatial derivatives, there are two main ways of implementing boundary conditions. The first is the simplest, however it only works on simple domains, such as rectangles in cartesian coordinates or circles in polar coordinates. By choosing the grid correctly in this case, the outermost rectangle of grid-points can be placed directly on the boundary, and boundary conditions applied directly. Any derivatives required can then be constructed from existing grid-points, using the method outlined above. The second method is slightly more complicated, but more widely applicable. The
grid is chosen to be large enough to entirely cover the domain. Points in the grid lying outside the boundary, but either horizontally or vertically adjacent to points inside the boundary are then labelled as ghost points, and values of the numerical solution at these points are calculated based on the solution inside the boundary and the stated boundary conditions. These values are then used to calculate the required derivatives at the boundary.

Approximating the spatial derivatives as above allows the wave equation to be expressed as a linear system of second order ODEs. A common approach to this is to first define a second variable $v$ as:

$$v(x) = \frac{du}{dt}, \quad \frac{d^2u}{dt^2} = \frac{dv}{dt}.$$  

This allows the second order system to be rewritten as a first order system, for which there are a wide range of solution methods available. The method employed in this paper was different, and involved directly approximating the second time derivative. More details are given in section 2.3 below.

A less direct approach is the finite volume method, as described by Leveque (2002). This technique is usually applied to hyperbolic equations that can be re-written in the first order form:

$$\frac{\partial u}{\partial t} + \nabla \cdot [F(u, x, y, t)] = g(x, y, t).$$  

Typically, a finite volume method is applied when the equation has certain conservation properties that should also be present in the numerical solution.

To obtain a finite volume method, the domain is first divided into a set of smaller regions, or cells. A volume integral of the above form of the equation is then performed on each cell, to rewrite the equation in terms of the integral of $u$ for each cell. Using Gauss’ theorem, the spatial derivative term can be replaced with a surface integral of the flux function $F(u, x, y, t)$ over the surface of the subregion. This can then be further simplified by noting that each subregion shares each part of its surface either with a different subregion or the outer boundary. The contributions from the outer boundary can be computed directly using the specified boundary conditions, while the surface integral for an area between two subregions will be equal in magnitude but with an opposite sign for each of the regions. This can be used to construct a system of linear equations by approximating the value of this integral, which in turn can be solved for using an established time-stepping method. Often the integral of
the flux function $F$ is substituted for a slightly different numerical flux function, in order to gain certain desirable convergence properties.

Finally, another common approach, particularly for problems with a more complex domain shape, is the **finite element method**, as described by Eriksson, Estep, Hansbo, Johnson (1996). This method is based on first finding the weak form of the equation for its spatial derivatives, then discretising this in space by approximating functions as the sum of a finite number of test functions to obtain an ODE system. This method is often favourable when using a complex domain or when an adaptive method of spatial discretisation is required, since the algorithms for generating a finite element discretisation are well developed.

The weak form of the spatial derivatives is formed by first defining a space $V$ of functions in which the solution should lie, then multiplying the equation by an arbitrary function $v$ from some trial space $V^*$ of functions, often the space $V$ or a vector space with similar functions to those in $V$, and integrating over some subregion $E$ of the domain, with boundary $\partial E$:

$$
\int_E v \frac{\partial^2 u}{\partial t^2} dx = \int_{\partial E} v \nabla u \cdot dS - \int_E \nabla u \cdot \nabla v dx + \int_E v f dx.
$$

Next, the domain as a whole is divided into smaller regions called elements. These are often triangles in practice, but other methods may use other polygons. The function spaces $V$ and $V^*$ are then restricted to a test space of functions $V_n$ and trial space $V_n^*$ that are defined piece-wise on each element, and that can be described by a finite number of variables defined at the nodes of the grid of elements in a linear fashion. The solution $u$ is then approximated as an element of $V_n$. For hyperbolic problems such as the wave equation, a common choice for these restricted spaces is a space of functions that are continuous on individual elements, but can be discontinuous over the domain of the equation. This allows for discretisations in which the computed values at a given point are affected less by values at points further away in space, which is more natural for wave systems. The methods this produces are called discontinuous Galerkin, or $dG$, methods (Hesthaven, 2008). The above form of the equation is then evaluated by substituting $v$ for basis functions of $V_n^*$, which leads to a linear system of ODEs.
2.3 Finite difference discretisation

In this section, the precise details of the finite difference discretisation used for this problem will be given, together with validity tests to demonstrate their accuracy and stability.

2.3.1 Notation

In order to fully describe the finite difference discretisation for this problem, a brief discussion of the notation is required. Firstly, a square grid was used for \( u \), with the same step-size \( h \) used in both the \( x \) and \( y \) directions. The dimensions of the grid were \((x,y) \in [-x^*,x^*]^2\), with \( x^* \) chosen so that \( D \subset [-x^*,x^*]^2 \). The number of grid-points in either direction of the grid is denoted \( N \), where \((N-1)h = 2x^* \). A staggered grid was used for the variables \( \phi_1 \) and \( \phi_2 \) introduced for the PML boundary, which was displaced from the normal grid by a distance \( \frac{1}{2}h \) in both the \( x \) and \( y \) directions. A time-step of \( k \) was used for the time discretisation, with \( h \) and \( k \) linked by the relation:

\[
k = \frac{\lambda_{CFL} h}{c},
\]

where \( \lambda_{CFL} \) is the CFL number for the discretisation.

Points on the grids are denoted as \((x_i,y_j)\), and time-steps as \( t_m \), using the following definition:

\[
x_i = -x^* + ih, \quad y_j = -x^* + jh, \quad t_m = mk.
\]

For a function \( v \) defined in the domain \( D \) of the problem, the approximates value of \( v \) on the grids are notated as

\[
v_{i,j}^m \approx y(x_i, y_j, t_m).
\]

In cases where the grid-point to which the indices \( i, j \) refer lies outside of the domain \( D \), the value of \( v_{i,j}^m \) or \( v_{i+\frac{1}{2},j+\frac{1}{2}}^m \) should be taken to be a linearly extrapolated value of \( v \) at a corresponding ghost point, based on the data from a neighbouring point on the interior of the domain. For more details on how this was accomplished, see section 2.3.3.

For the purpose of applying linear operators to the approximate values at a given time step, it was necessary to construct a vector of values corresponding
to points inside the domain. For the approximate values at a time \( t = t_m \), this vector is denoted as \( \mathbf{v}^m \), where

\[
\mathbf{v}^m = \{ v_{i,j}^m : (x_i, y_j) \in D \}.
\]

Finally, to define the boundaries, the zero contour of a level-set function was used, which allowed uncertainty to be easily incorporated into the shape. This level-set function is denoted as \( \Phi(x, y) \), and its values on the two grids as \( \Phi_{i,j} \) and \( \Phi_{i+\frac{1}{2},j+\frac{1}{2}} \) respectively. Values of \( \Phi_{i,j} \geq 0 \) corresponded to grid-points inside the object \( \Omega \), whereas values \( \Phi_{i,j} < 0 \) represented a point in the domain \( D \).

To represent this boundary numerically, the distance to the boundary from a neighbouring interior grid-point was approximated as \( \alpha_{i,j,d}h \), where \( i, j \) represent the coordinates of the interior grid-point adjacent to the boundary, and \( d \in \{x-, x+, y-, y+\} \) represents the direction to the boundary from that grid-point. The values of \( \alpha \) were calculated by linear interpolation of the level-set function, with full details given below in section 2.3.3. A visualisation of two of these distances can be seen in figure 3, also in section 2.3.3.

### 2.3.2 4th order finite difference discretisation in space

The first step in discretising the equations was to perform a semi-discretisation by separating the time and space derivatives. This is accomplished by first leaving the time derivatives as continuous derivatives, but discretising the spatial domain, which transforms the wave equation from a PDE to an ODE for a vector of high dimension.

On the interior of \( \Lambda \setminus \Omega \), this only requires a discretisation of the Laplacian operator:

\[
\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.
\]

Since the derivatives in the two dimensions are independent of each other, the simplest approach to this is to apply the 2nd order accurate central difference formula in each direction, which yields the basic 5-point stencil for the Laplacian:

\[
\Delta u \approx \frac{u_{i+1,j}^m - 2u_{i,j}^m + u_{i-1,j}^m}{h^2} + \frac{u_{i,j+1}^m - 2u_{i,j}^m + u_{i,j-1}^m}{h^2},
\]

\[
= \frac{u_{i+1,j}^m + u_{i-1,j}^m + u_{i,j+1}^m + u_{i,j-1}^m - 4u_{i,j}^m}{h^2} = (D_{x,+}D_{x,-} + D_{y,+}D_{y,-}) u^m.
\]
This approximation has an associated error of $O(h^2)$, however it is possible to improve this accuracy to $O(h^4)$ by adding correction terms representing $-\frac{1}{12} h^2 \left( \frac{\partial^4}{\partial x^4} + \frac{\partial^4}{\partial y^4} \right) u$ (Kreiss, Petersson, Yström 2002). Since these correction terms use fourth order derivatives they require a larger number of grid-points to approximate accurately, which is difficult at the boundary $\partial \Omega$. To account for this, an indicator matrix $G$ was defined on the grid, defined as follows:

$$(Gy)_{i,j}^m = \begin{cases} y_{i,j}^m & \text{Point } (i,j) \text{ is not directly adjacent to } \partial \Omega, \\ 0 & \text{Point } (i,j) \text{ is directly adjacent to } \partial \Omega. \end{cases}$$

In the above, directly adjacent should be taken to mean that one or more of the neighbouring grid-points to $(i,j)$ lies within the region $\Omega$. From this, the extension to fourth order accuracy on the interior of $\Lambda$ was made by:

$$\Delta u \approx \left( \left( I - \frac{1}{12} h^2 D_{x,+} + D_{x,-}G \right) D_{x,+} + D_{x,-} \right) u^m + \left( I - \frac{1}{12} h^2 D_{y,+} + D_{y,-}G \right) D_{y,+} + D_{y,-} u^m. \tag{2.1}$$

In the above, $I$ refers to the identity matrix. The result of the above spatial discretisation is that the stencil is second order at the boundary, fourth order away from the boundary, and a combination of the two discretisations near to the boundary.

### 2.3.3 Implementation of complex Dirichlet boundaries

In order to define the complex boundary $\partial \Omega$ the approach presented by Kreiss, Petersson, Yström (2002) is followed. In addition, the precise shape of the boundary $\partial \Omega$ was defined using a level-set function $\Phi(x,y)$. This allowed points inside of $\Omega$ to be detected at the start, and removed from the discretisation. On the finite difference grid $\partial \Omega$ was approximated by calculating values of $\alpha_{i,j,x-}, \alpha_{i,j,x+}, \alpha_{i,j,y-}, \alpha_{i,j,y+}$, where appropriate for each point directly adjacent to $\partial \Omega$:

$$\begin{align*}
\alpha_{i,j,x-} \Phi_{i-1,j} + (1 - \alpha_{i,j,x-}) \Phi_{i,j} &= 0, \\
\alpha_{i,j,x+} \Phi_{i+1,j} + (1 - \alpha_{i,j,x+}) \Phi_{i,j} &= 0, \\
\alpha_{i,j,y-} \Phi_{i,j-1} + (1 - \alpha_{i,j,y-}) \Phi_{i,j} &= 0, \\
\alpha_{i,j,y+} \Phi_{i,j+1} + (1 - \alpha_{i,j,y+}) \Phi_{i,j} &= 0.
\end{align*}$$

In the above equations, the subscripts $x\pm, y\pm$ refer to the positive and negative directions along the $x$- and $y$-axes respectively. The above calculations were
only performed for grid-points inside the computed domain and adjacent to the boundary, and only needed to be performed once since the boundary $\partial \Omega$ was taken to be constant with time. To visualise how the values of $\alpha$ actually relate to the grid and the boundary, the following diagram is included:

![Diagram of the finite difference grid close to $\partial \Omega$, with grid points inside $D$ shown in green, grid points inside $\Omega$ in blue, and the boundary $\partial \Omega$ in red. The distance $\alpha_{i,j,x} h$ is shown for one point with a purple line, and $\alpha_{i,j,y} h$ with a yellow line.](image)

The values of $\alpha$ computed above can now be used to linearly extrapolate values of $u$ for ghost points inside of $\Omega$, in order to implement the Dirichlet boundary condition on $\partial \Omega$. This gives the following approximations of values at the ghost points, which have second order accuracy in $h$:

$$u_{i-1,j}^m = -\frac{1 - \alpha_{i,j,x}}{\alpha_{i,j,x}} u_{i,j}^m - \frac{1}{\alpha_{i,j,x}} U (x_i - \alpha_{i,j,x} h - ct_m), \quad (2.2)$$

$$u_{i+1,j}^m = -\frac{1 - \alpha_{i,j,x}}{\alpha_{i,j,x}} u_{i,j}^m - \frac{1}{\alpha_{i,j,x}} U (x_i - \alpha_{i,j,x} h - ct_m), \quad (2.3)$$

$$u_{i,j-1}^m = -\frac{1 - \alpha_{i,j,y}}{\alpha_{i,j,y}} u_{i,j}^m - \frac{1}{\alpha_{i,j,y}} U (x_i - ct_m), \quad (2.4)$$

$$u_{i,j+1}^m = -\frac{1 - \alpha_{i,j,y}}{\alpha_{i,j,y}} u_{i,j}^m - \frac{1}{\alpha_{i,j,y}} U (x_i - ct_m). \quad (2.5)$$
In the above, \( U \) is the wave function defined in section 2.1.1. The ghost values from equation (2) – (5) can then be substituted into equation (1) as required, giving the semi-discretised wave equation:

\[
\frac{\partial^2 u}{\partial t^2} = c^2 \left( (A_1 + A_2) u + g(t) \right),
\]

\[
\left. \frac{\partial u}{\partial t} \right|_{t=0} = 0.
\]

In the above equation, the diagonal matrix \( A_1 \) represents the contribution of \( u_{m}^{i,j} \) to the extrapolated values at the ghost points, \( A_2 \) represents the discretisation of the derivative in \( A \setminus \Omega \), and \( g(t) \) represents the contribution of the non-homogeneous Dirichlet boundary.

One important issue to note is that the non-zero elements of \( A_1 \) can become very large compared to \( A_2 \) in the case when one or more of the values for \( \alpha \) is close to 0, or when the boundary \( \partial \Omega \) passes very close to a grid point, since they are of the form:

\[
\left( A_1 \right)_{(i,j),(i,j)} = -\frac{1}{h^2} \sum_{d=x+,y+,y-} m_d \frac{1}{\alpha_{i,j,d}} \\
\]

where \( m_d \) is equal to 1 for directions \( d \) in which \( \alpha_{i,j,d} \) was calculated, and 0 otherwise. This means that the ratio between eigenvalues of \( A_1 + A_2 \) can become arbitrarily large, and affect the stability properties of the equation. As a result the boundary terms will need special treatment when discretising the time derivative.

### 2.3.4 Time discretisation

In this case, the approach to time discretisation used by Kreiss, Petersson, Yström (2002) was adopted, in which the second order time derivative was discretised directly with a central difference approximation:

\[
\left. \frac{\partial^2 u}{\partial t^2} \right|_{t=t_m} \approx \frac{u^{m+1} - 2u^m + u^{m-1}}{k^2}.
\]

As mentioned in the previous section, the matrix \( A_1 \) can present stability problems, depending on the exact shape of \( \partial \Omega \). If an explicit time-stepping method is applied to the semi-discretised equation using \((A_1 + A_2) u^m\) to approximate
$(A_1 + A_2)\ u$, the following restriction (Kreiss, Petersson, Yström, 2002) is placed on the time-step size:

$$\left( \max_l |\lambda_l| + \min_l |\lambda_l| \right) k^2 \leq 4,$$

where $\lambda_l$ are the eigenvalues of the matrix $A_1 + A_2$. Due to the elements of $A_1$ potentially being arbitrarily large, this places an extremely tight restriction on the time-step. To resolve this, a semi-implicit approach to the time discretisation is taken:

$$u^{m+1} - 2u^m + u^{m-1} = c^2 \left( A_1 \left( \frac{u^{m+1} + u^{m-1}}{2} \right) + A_2 u^m + g^m \right),$$

$$\left( I - \frac{1}{2} c^2 k^2 A_1 \right) (u^{m+1} + u^{m-1}) = (2I + c^2 k^2 A_2) u^m + c^2 k^2 g.$$

This gives the time-stepping formula presented by Kreiss, Petersson, Yström (2002) by which the next step $u^{m+1}$ can be calculated from the previous two steps. Since matrix $A_1$ is a diagonal matrix, the cost of inversion in the above time-stepping formula is minimal. Stability for this method now follows from the results presented in the same paper. $A_1$ is an unbounded negative semi-definite matrix ($v^T A_1 v \leq 0$ for all $v \in \mathbb{R}^N$ for some large $N$) and $A_2$ is negative definite. This means that the stability analysis applied to such systems will be applicable here, and the time-stepping method will be stable for CFL numbers less than $\frac{1}{\sqrt{2}}$.

In order to begin the calculations, it is necessary to not only have the initial data given by $u_{i,j}^0 = u_0(x_i, y_j) = 0$, but also an estimate of the previous time-step $u^{-1}$. Due to the similar issues with the values of $\alpha$, this time with the vector $g(0) = \mathcal{O} \left( \frac{\max \alpha_{i,j,d}}{\alpha_{i,j,d}} \right)$, it is difficult to calculate this to second order accuracy using a Taylor expansion:

$$u(x_i, y_j, t - 1) = u(x_i, y_j, 0) - k \frac{\partial u}{\partial t}\bigg|_{x_i, y_j, 0} + \frac{1}{2} k^2 \frac{\partial^2 u}{\partial t^2}\bigg|_{x_i, y_j, 0} + \mathcal{O} (k^3),$$

$$\therefore \ u^{-1} \approx \frac{1}{2} k^2 g(0).$$

In the above, the explicit evaluation of $g(0)$ can produce extremely large error oscillations that do not decay with time, even though this is only performed for a single time-step. The simplest way to avoid this is to choose the wave profile $U$ such that at time $t = 0$ the value of $U$ at the boundary $\partial \Omega$ is either 0 or close to 0, so that the approximation $u_{i,j}^{-1} = 0$ can be made with reasonable accuracy.
2.3.5 Implementation of PML

In modifying the wave equation to include a PML region, two new variables \( \phi_1, \phi_2 \) were introduced. In each of the three governing differential equations, \( u \) had different levels of spatial derivatives to \( \phi_1 \) and \( \phi_2 \); when \( u \) had terms with either second order or no spatial derivatives, \( \phi_1, \phi_2 \) had only terms with first order spatial derivatives, and vice versa. As a result, it was practical to define \( \phi_1, \phi_2 \) on a spatially staggered grid.

Due to the position of the staggered grid in relation to the normal grid, first derivatives on both grids were calculated as the average of two central difference approximations:

\[
\frac{\partial \phi_1}{\partial x} \bigg|_{i,j} \approx \frac{1}{2} \left( \frac{\phi_{1,i+\frac{1}{2},j} - \phi_{1,i-\frac{1}{2},j}}{h} + \frac{\phi_{1,i+\frac{1}{2},j+\frac{1}{2}} - \phi_{1,i-\frac{1}{2},j+\frac{1}{2}}}{h} \right) = A_{\phi_1,u} \phi_1,
\]

\[
\frac{\partial \phi_2}{\partial y} \bigg|_{i,j} \approx \frac{1}{2} \left( \frac{\phi_{2,i,j+\frac{1}{2}} - \phi_{2,i,j-\frac{1}{2}}}{h} + \frac{\phi_{2,i+\frac{1}{2},j} - \phi_{2,i-\frac{1}{2},j}}{h} \right) = A_{\phi_2,u} \phi_2,
\]

\[
\frac{\partial u}{\partial x} \bigg|_{i+\frac{1}{2},j+\frac{1}{2}} \approx \frac{1}{2} \left( \frac{u_{i+1,j} - u_{i,j}}{h} + \frac{u_{i+1,j+1} - u_{i,j+1}}{h} \right) = A_{u,\phi_1} u,
\]

\[
\frac{\partial u}{\partial y} \bigg|_{i+\frac{1}{2},j+\frac{1}{2}} \approx \frac{1}{2} \left( \frac{u_{i,j+1} - u_{i,j}}{h} + \frac{u_{i+1,j+1} - u_{i+1,j}}{h} \right) = A_{u,\phi_2} u.
\]

These approximations gave second order accuracy, however in this case it was sufficient since the aim of implementing PML is not to produce a high accuracy solution in the PML region, but rather to heavily damp any non-zero solutions. Therefore it was sufficient that the method had second order accuracy for \( \phi_1, \phi_2 \) and that the boundary of the PML region \( \partial \Sigma \) did not produce reflections.

The next consideration for the modified equations was that of the boundaries, both for the variables \( \phi_1, \phi_2 \) on \( \partial \Omega \) and for \( u, \phi_1, \phi_2 \) on \( \Gamma \). For the value of \( u \) at \( \Gamma \), it was simple to incorporate \( \Gamma \) into the calculations of \( \alpha \) as a homogeneous Dirichlet boundary, by modifying the level-set function to incorporate both \( \partial \Omega \) and \( \Gamma \). For \( \phi_1, \phi_2 \) this problem was trivially solved in both cases. For \( \partial \Omega \), the coefficients \( \zeta_1, \zeta_2 \) were 0 for the entirety of \( \Lambda \), which meant that \( \phi_1, \phi_2 \) were constant and 0 in this region, so there would be no boundary contributions from either at \( \partial \Omega \) and it could simply be omitted from calculations. Similarly for the region of \( \Sigma \) close to \( \Gamma \), provided that the PML calculations in \( \Sigma \) damped the solution, the values of \( u \) and \( \nabla u \) were sufficiently small to
neglect entirely, meaning $\phi_1, \phi_2$ could also be assumed to be 0 and allowing $\Gamma$ to also be omitted from calculations for $\phi_1, \phi_2$.

To write the above as a semi-discretised equation, first introduce four diagonal matrices $Z_1, Z_2, Z_1', Z_2'$ defined by:

$$(Z_1 u^m)_{i,j} = \zeta_1(x_i) u^m_{i,j},$$

$$(Z_2 u^m)_{i,j} = \zeta_2(y_j) u^m_{i,j},$$

$$(Z'_1 \phi_1^m)_{i+\frac{1}{2},j+\frac{1}{2}} = \zeta_1(x_{i+\frac{1}{2}}) \phi_1^m_{i+\frac{1}{2},j+\frac{1}{2}},$$

$$(Z'_2 \phi_2^m)_{i+\frac{1}{2},j+\frac{1}{2}} = \zeta_2(y_{j+\frac{1}{2}}) \phi_2^m_{i+\frac{1}{2},j+\frac{1}{2}}.$$
\[
\frac{1}{2} c^2 A_1 \left( u^{m+1} + u^{m-1} \right) + c^2 A_2 u^m + c^2 g^m + A_{\phi_1,u} \phi_1^m + A_{\phi_2,u} \phi_2^m,
\]
\[
\phi_1^{m+1} - \phi_1^m = -\frac{1}{2} Z_1' \left( \phi_1^{m+1} + \phi_1^m \right) + \frac{1}{2} c^2 \left( Z_2' - Z_1' \right) A_{u,\phi_1} \left( u^{m+1} + u^m \right),
\]
\[
\phi_2^{m+1} - \phi_2^m = -\frac{1}{2} Z_2' \left( \phi_2^{m+1} + \phi_2^m \right) + \frac{1}{2} c^2 \left( Z_1' - Z_2' \right) A_{u,\phi_2} \left( u^{m+1} + u^m \right),
\]
\[u^{-1} = u^0 = 0,\]
\[\phi_1^0 = \phi_2^0 = 0.\]

Time-stepping is then performed by solving for \(u^{m+1}\), followed by \(\phi_1^{m+1}\) and \(\phi_2^{m+1}\):

\[
\left( I + \frac{1}{2} k Z_1 + Z_2 \right) - \frac{1}{2} c^2 k^2 A_1 \right) u^{m+1} = \left( 2I - k^2 Z_1 Z_2 + c^2 k^2 A_2 \right) u^m
\]
\[
+ \left( -I + \frac{1}{2} k \left( Z_1 + Z_2 \right) + \frac{1}{2} c^2 k^2 A_1 \right) u^{m-1} + k^2 A_{\phi_1,u} \phi_1^m + k^2 A_{\phi_2,u} \phi_2^m + c^2 k^2 g^m,
\]

\[
\left( I + \frac{1}{2} k Z_1' \right) \phi_1^{m+1} = \frac{1}{2} c^2 k \left( Z_2' - Z_1' \right) A_{u,\phi_1} \left( u^{m+1} + u^m \right) + \left( I - \frac{1}{2} k Z_1' \right) \phi_1^m,
\]

\[
\left( I + \frac{1}{2} k Z_2' \right) \phi_2^{m+1} = \frac{1}{2} c^2 k \left( Z_1' - Z_2' \right) A_{u,\phi_2} \left( u^{m+1} + u^m \right) + \left( I - \frac{1}{2} k Z_2' \right) \phi_1^m.
\]

A discussion of the theoretical stability of this discretisation is beyond the scope of this thesis. The stability is tested practically below in section 2.4.2.

## 2.4 Validation of methods

In order to provide validation that the above numerical methods are stable and will actually provide results with the desired accuracy, numerical experiments will now be presented to demonstrate their effectiveness.

### 2.4.1 Validation of basic solver

To provide validation of the basic solver, that is the numerical method applied to the wave equation without a PML boundary, the wave equation on a bounded interior domain (the boundary \(\partial \Omega\) encircles the solution region \(D\)) was used. In addition, a forcing term \(f\) was applied to the system \((u_{tt} - c^2 \Delta u = f)\) to
allow the system to be defined with a known exact solution (Kreiss, Petersson, Ystrom, 2002). This allowed the error of the numerical solution to be calculated exactly for convergence analysis, rather than being approximated using a test run with a small $h$ as a model solution for comparison. For the purposes of these tests, the error was defined as:

$$e(h, k, m) = \max_{i,j} |u_{exact}(x_i, y_j, t_m) - u_{num,h,k}(x_i, y_j, t_m)|,$$

where $u_{exact}$ is the exact solution to the system, and $u_{num,h,k}$ is the numerical solution obtained using a spatial step-size of $h$ and a time-step of $k$.

The test case chosen was that of sine waves moving in the direction of $x$ increasing inside an ellipse with axes lying parallel to the $x$- and $y$-axes:

$$D = \left\{(x, y) : x^2 + \left(\frac{y}{0.75}\right)^2 < 1\right\},$$

$$u_{exact}(x, y, t) = \sin(4\pi(x - t)) \sin(4\pi y).$$

By setting $c = 1$, this gave the following initial, boundary and forcing terms for the system:

$$u_0(x, y) = \sin(4\pi x) \sin(4\pi y),$$
$$v_0(x, y) = -4\pi \cos(4\pi x) \sin(4\pi y),$$
$$f(x, y, t) = 16\pi^2 \sin(4\pi(x - t)) \sin(4\pi y),$$
$$g(x, y, t) = \sin(4\pi(x - t)) \sin(4\pi y),$$

where $u_0, v_0, g$ are the functions defined in section 2.1.1, and $f$ is the forcing term. To show what this solution actually looks like, below is a plot of the numerical solution at time $t \approx 5$, using $h = 0.0111(4d.p.)$ (note: in this paper $d.p.$ stands for decimal places) and a CFL number of 0.5:
Figure 2.4: Numerical solution of the above system using the discretisation specified above.

The stability of the method can be demonstrated by running a numerical system over a long period of time, and showing that the error does not grow with time. To do this, a simulation was run with $h = 0.0111(4\text{d.p.})$ and $\lambda_{CFL} = 0.5$ up until a time $T = 80$, and the error as defined above was recorded at each time-step. Plotting this error against time produces the following graph:
Figure 2.5: Plot of the error of the solution over time until $t = 80$, using the above discretisation.

As can be seen above, the error does not continue to grow with time, even for long time periods. Given that the time periods required for the investigation are much shorter than this (typically up until $t = 5$ or less), this method is suitably stable.

To check that the accuracy is of the correct order, the above system was solved numerically for different values of the step-size $h$. To avoid the issues of instabilities with the initial conditions, the time-stepping method was initialised with the exact values of $u_{i,j}^0$ and $u_{i,j}^1$ and the simulation was started at $m = 1$.

To test convergence with $h$, a range of $h$-values given by $h \in \left\{ \frac{1}{20+5n} : n = 1, 2, \ldots, 46 \right\}$ were used with $CFL = 0.5$. When compared to the exact solution, this resulted in the following graph:
In the above figure, the desired second order convergence (from the time-stepping due to a constant CFL number) can be clearly seen.

2.4.2 Validation of PML implementation

To test the validity of applying the PML boundary with the above method, first a test case was run (scattering of a Gaussian plane-wave from the unit circle) up to a time \( T = 10 \), or around twice as long as the numerical experiments performed in section 4 below. The system used had the same dimensions for \( \Lambda \) and \( \Sigma \) as the numerical experiments as described in section 4. The largest absolute value of the solution at each time-step was then recorded, which produced the following graph:
As can be seen above, the values remain bounded for the entire simulation, with the solution effectively dropping to zero after a certain time period as expected. This indicates that the method is stable enough for use in the experiments in section 4.

To ensure that implementing the PML boundary does not affect the accuracy of the solution inside the domain of interest \( \Lambda \), another test was performed without a PML boundary using a larger domain \( \Lambda ' = [-3.8, 3.8]^2 \). The solution inside the region \([-2.2, 2.2]^2\) was then compared for a simulation with a PML boundary and the solution without to calculate errors. Doing so produced the following graph:
Figure 2.8: Semi-logarithmic graph showing the error of the PML solution inside $\Lambda$ compared to a test solution on a larger domain. Note that after a certain time period the test case becomes unreliable due to the wave being reflected back into the domain.

The time period $t < 1$ is not shown above since this was before scattering occurred, so no errors were produced. The main features in the above occur at the times $t \approx 2$ and $t \approx 5$. The simulation modelled a Gaussian plane-wave with its peak starting at $x = -2$ and incident on a unit circle scatterer with wave-speed $c = 1$. This meant that it first reaches the PML boundary at $t \approx 2.2$ after reflecting from the scatterer, indicating the change here is due to the effect of the PML. From the above, it is clear that the PML method produces comparable accuracy to simply using a larger domain, since the error remains small, despite showing a small change due to the reflection.

The next major change occurs at around $t \approx 5.4$. For the simulation on the larger domain, this is when theory predicts the outgoing wave re-enters the test region after reflecting from the outer boundary. However, the stability plot shows that the solution with PML is effectively 0 everywhere for this time period, indicating that this does not conceal inaccuracies in the PML solution.
Chapter 3

Uncertainty Quantification (UQ)

This section will provide a brief overview of what is meant by uncertainty quantification, the various methods for UQ that will be used, and a description of the precise problem in UQ that this paper will investigate.

3.1 Overview of UQ

The aim of uncertainty quantification is to analyse problems in which one or more of the input parameters, denoted $X$, defining the problem is uncertain, either due to a lack of knowledge about the parameters, or due to some inherent randomness in the system itself. The objective is then to determine the statistics of some observable quantity of interest (QoI) $Y[X]$ of the system, based on the statistics of the uncertain input parameters. Often it is of interest to calculate the expected value $\bar{Y} = \mathbb{E}[Y[X]]$ and the variance $\sigma_Y^2 = \mathbb{E}\left[(Y[X] - \bar{Y})^2\right]$ as numerical indicators of the distribution.

As an example of where UQ may be useful, consider weather forecasting. In principle, it is possible to make accurate predictions for the future of the weather based on the weather at the moment. However, perfect measurements can never be made. It is only possible to measure variables such as wind speed, humidity or temperature at a relatively small number of points, leaving the distribution of these values uncertain elsewhere. In addition, there are uncertainties present in the measurements themselves. It is important to know how these uncertainties will affect the probability distribution of the amount of rainfall in a given area, or the cloud coverage.
3.2 Review of UQ methods

3.2.1 Overview of methods

There are three main methods of uncertainty quantification that this paper will use; Monte Carlo (MC), quasi-Monte Carlo (QMC) and sparse grids. The first of these, Monte Carlo, is the most widely applicable. The approach is extremely simple; generate sample points $X_i$ randomly according to the required probability distribution, then evaluate the classical system obtained by replacing the random parameters with the fixed values given by the sample point. For this paper, the MATLAB function rand() was used to generate an array of random numbers of the required size. By analysing the results across all sample points, an experimental approximation for the distribution of $Y[X]$ is obtained. For example, the expected value and variance of $Y$ are obtained using the sample formulae, where $N_{\text{points}}$ is the number of sample points used:

$$\bar{Y} \approx \frac{1}{N_{\text{points}}} \sum_{i=1}^{N_{\text{points}}} Y[X_i],$$

$$\sigma_Y^2 \approx \frac{1}{N_{\text{points}} - 1} \sum_{i=1}^{N_{\text{points}}} (Y[X_i] - \bar{Y})^2.$$ 

The only restriction on when Monte Carlo integration will converge is that the function being sampled is Lebesgue integrable. Provided this condition is met it will converge for a high enough value of $N_{\text{points}}$ with an error of $O\left(\frac{1}{N_{\text{points}}}\right)$ (Caflisch, 1998), regardless of the dimensionality of the problem. However, for certain problems the coefficient of $N_{\text{points}}^{-\frac{1}{2}}$ may be large, leading to significant errors unless $N_{\text{points}}$ is very large. It is for this reason that other methods of UQ are also considered.

Quasi-Monte Carlo methods are similar to Monte Carlo methods, except the sampling is not random. Instead, it is based on a sequence of quasi-random numbers on the interval $[0,1)$ designed to be more evenly distributed than purely random sampling of the same interval, which are then transformed to represent the required probability distribution. For this paper, Halton sequences, as described by Caflisch (1998) were used. After this, the mathematics are the same as for MC, with the sample formulae again being used to calculate expected values and variances. To show the difference between MC and QMC, below are two scatter plots, in which points were sampled by both MC and QMC with a Halton sequence:
Figure 3.1: Scatter plots showing Monte Carlo sampling (left) and quasi-Monte Carlo sampling (right) over $[0, 1)^2$ with a uniform distribution

Quasi-Monte Carlo is constrained by the fact that its sample points are not truly random. As a result the method requires a continuous function to produce good convergence results. Given that this condition is satisfied, quasi-Monte Carlo integration typically converges with an error of $O\left(N_{\text{points}}^{-1} \log^{k} N_{\text{points}}\right)$, for some constant $k$ that can depend on the dimensionality of the problem (Caflisch, 1998). However, for high dimensionality the value of $k$ or the constant of proportionality may become large, and the method will require large values of $N_{\text{points}}$ for an accurate result, similar to standard Monte Carlo.

The third technique is sparse grids. This technique is completely different to MC and QMC, in that it is not based on sampling a certain number of points. Instead, it is based on representing smooth functions on a sparse grid of $O\left(n^{d-1} 2^n\right)$ points, where $n$ is the discretisation level of the grid and $d$ is the dimension of the problem. The size of this grid scales with $h_n = 2^{-n}$ (Garcke, 2012). Each of the points in this grid is then assigned a weight $w_i$ for numerical integration, giving the following formulae for expected value and variance, where $X_i$ are the sample points of a multi-dimensional uniformly distributed random variable $X$, $N_{\text{points}}$ is the number of points in the sparse grid and $V$ is the volume of the sample space:

$$\bar{Y} \approx \frac{1}{V} \sum_{i=1}^{N_{\text{points}}} w_i Y[X_i],$$

$$\sigma_Y^2 \approx \frac{1}{V} \sum_{i=1}^{N_{\text{points}}} w_i (Y[X_i] - \bar{Y})^2.$$
The sparse grid method is heavily constrained by smoothness requirements. In particular, the mixed second derivatives of the function being integrated must exist and be bounded in order to obtain the convergence rate stated above (Garcke, 2012). In practice, this rules out both problems with discontinuities and problems where a small change in the input parameters can create a large change in the solution.

Sparse grids are preferable to other numerical integration methods, such as the trapezoidal rule, when the number of dimensions is high. The number of grid-points required for a certain accuracy is notably lower for sparse grids than the trapezoidal rule, due to the number of grid points for the trapezoidal rule scaling with $O(n^d)$ for an error of $O(n^{-2})$.

The convergence of the sparse grid method for a sufficiently smooth function is $O(n^{d-1/2-2n})$ (Garcke, 2012). However, this is subject to certain requirements on the system. As an example of what different levels of sparse grids look like, below are plots showing three different levels for the 2D sample space $[0, 1]^2$:
3.3 Description of the problem in UQ

3.3.1 Random parameters for the problem

For the problem addressed in this paper, the uncertainty comes in the form of the shape of the scatterer. To implement this for a vector of uniformly distributed independent random variables $X$, the boundary of the scatterer was represented as the zero contour of a level-set function $\Phi(x, y, X)$. The level-sets used were not necessarily continuous in $(x, y)$, but were chosen so that the zero contour would vary continuously with the parameters $X$. Three families of level-set functions were considered, which are described in more detail in section 4.

3.3.2 Quantities of interest (QoI)

For this problem, the quantities of interest were the energy intensities of the scattered wave in various directions. This was recorded as the maximum en-
nergy density $E$ in time recorded at each measurement point:

$$E(x, y) := \sup_{t > 0} [u(x, y, t)^2] .$$

To approximate this with the given time and space discretisations, the following approximation was made:

$$E(x, y) \approx \max_{m=0,1,\ldots,\lfloor T \rfloor} [u_m^* (x, y)^2] ,$$

where $u_m^*(x, y)$ is the approximation of $u(x, y, t_m)$ obtained by bi-linear interpolation at $(x, y)$ from the four grid-points closest to it:

$$u_m^*(x_{i+\beta_1}, y_{j+\beta_2}) = (1 - \beta_1) \left((1 - \beta_2) u_{i,j}^m + \beta_2 u_{i,j+1}^m\right) + \beta_1 \left((1 - \beta_2) u_{i+1,j}^m + \beta_2 u_{i+1,j+1}^m\right) .$$

The maximum value was chosen instead of an integral over time due to the Gaussian profile over time of the Dirichlet boundary condition on $\partial \Omega$ at any given point. This meant that the profile over time of the solution at a point away from the boundary could be assumed to be approximately Gaussian with the same width, meaning the integral over time of the scattered energy would be proportional to the maximum value, so it did not matter which was chosen. This also meant that the recorded value could be assumed to occur within a certain time period $0 < t < T$ after the start of the simulation, with $T$ given by the wave speed and the size of the measurement ring defined below.

To measure these quantities, a ring of points of radius $R_E$ was defined around the origin, so that the object $\Omega$ would lie entirely within the ring. The points were evenly distributed around the ring, and spaced so that the distance between them was approximately the same as the distance between points on the finite difference grid. The value of $u_m^*$ was then computed at each point and squared. The maximum value over the duration of the simulation was then recorded for each measurement point. These values will be denoted as $\bar{E}(\theta)$, where $\theta$ denotes the angular coordinate of the measurement point:

$$\bar{E}(\theta) := \max_{m=0,1,\ldots,\lfloor \frac{T}{\Delta t} \rfloor} [u_m^*(R_E \cos(\theta), R_E \sin(\theta))^2] .$$

As well as the overall distribution of the energy scattering profile, the distribution of scattered energy at certain angles around the object were of interest. As well as providing quantitative rather than qualitative information,
they offered the possibility to numerically analyse the convergence of the UQ methods employed for each test case. Three angles were chosen for this task: \(\theta = 0\) (directly behind the object from the perspective of the incoming wave), \(\theta = \pi\) (directly in front of the object) and \(\theta = \frac{1}{2} \pi\) (off to the side of the object). Only one side of the object was considered for more precise analysis, since in each test case the reflection of the object in the \(x\)-axis was equally probable as the object itself, so the distributions in the maximum energy are theoretically symmetrical.

### 3.4 Practical considerations

#### 3.4.1 Computational costs

There were several contributions to the computational cost that affected the choice of the maximum value of \(N_{\text{points}}\) for each case. The most direct one was that every sample point taken meant another classical system that had to be numerically computed, meaning the computation time scaled with \(O(N_{\text{points}})\). This was easily manageable with MC and QMC, however due to the sparse grid method using \(O(n^{d-1}2^n)\) points at level \(n\), the number of levels that could reasonably be computed quickly became limited in higher dimensions.

Another contribution to the cost was the number of grid-points. While it was possible in some test cases to avoid objects of extremely small size, or with details on a very small scale, it was not possible in every case to prevent small objects. As an example, see the bubble array described below in section 4. This meant that in some cases, it was necessary to take larger values of \(N\) to ensure accuracy. Due to the requirement to take a sufficiently small value of \(h\) and the computational cost of large \(N\), this limited the size of the domain \(\Lambda \cup \Sigma\) that could be used, which in turn limited the possible values of the measurement ring radius \(R_E\). It was also impractical with the resources available to compute any systems for sufficiently large values of \(N\), due to the memory requirements of storing all the matrices and vectors required for the numerical method.

Another contribution to the computational cost was the shape of the scatterer. The solution to the equation was only calculated in each case for grid-points outside the scatterer, which meant for larger scatterers more points would lie inside, and thus the matrices and vectors used by the solver would be of a smaller size. However due to the fixed minimum and maximum sizes of the
objects, typically the effect of the smaller matrices would be negligible compared to the larger number of grid-points that always lie outside of $\Omega$.

### 3.4.2 Tests of regularity for sparse grid method

Before beginning tests on the convergence for the various methods, the regularity of the problem was considered; that is how smoothly the values $\bar{E}(\theta)$ vary with the parameters. Rather than testing every parameter individually, a couple of parameters expected to contribute significantly were chosen. To account for the other parameters, a random line segment through the sample space was taken, by selecting two random sample points $X_0$ and $X_1$, then linearly interpolating between them with some parameter $\mu \in [0, 1]$:

$$X_\mu = (1 - \mu)X_0 + \mu X_1.$$  

To ensure the line is long enough to be representative of the space, the line was rejected and a new one selected if it was shorter than 0.5 when the random parameters were scaled to the interval $[0, 1]$. Sample points were then taken for evenly spaced values of $\mu$, and the values of the quantities of interest calculated, as well as approximations of their first and second derivatives with respect to $\mu$ using the finite difference formulae described in section 2.2.1.

Calculating the first and second derivatives in this manner had the drawback that every point had an error associated with it from the finite difference discretisation, which could not be assumed to be continuous with $\mu$. As a result, the number of sample points for $\mu$ that could be taken was limited, since the first derivative had an associated error of $O(h^2 \Delta \mu^{-1} + \Delta \mu^2)$, and the second derivative an error of $O(h^2 \Delta \mu^{-2} + \Delta \mu^4)$, where $\Delta \mu$ is the step-size in $\mu$. As a result, 40 points were chosen as an intermediate value to minimise both sources of error.
Chapter 4

Numerical results

This section will provide more detailed descriptions of the scattering shapes considered for UQ, reasons for considering each shape and a discussion of the numerical results gathered.

4.1 Details of the system and discretisation

In section 2, the governing equations being studied were stated in general terms, however for this section it is possible to specify the system in more detail, by giving the precise functions and definitions used.

For the purposes of this problem, the region of interest was chosen as $\Lambda = [-2.2, 2.2]^2$ and the region $\Omega$ was defined by a level-set function $\Phi$, chosen so that $\Omega$ would lie entirely within a circular region of radius 1.75 centred on the origin. This allowed the measurement points to be placed in a ring around the origin of radius 2, and the PML region to be defined as $\Sigma = [-2.5, 2.5]^2 \setminus \Lambda$. A diagram of the different regions is included below, with the object $\Omega$ coloured red, the PML region blue and the region with the usual wave equation ($\Lambda \setminus \Omega$) in green:
Figure 4.1: Diagram of the computational domain for the studied problem, with the object coloured red, the region with the simple wave equation coloured green and the PML region coloured blue. The object shown here is the unit circle.

The PML coefficient functions $\zeta_1, \zeta_2$ were defined with the same profile as in Grote, Sim (2010):

$$
\zeta_1 (x) = \begin{cases} 
80 \left( \frac{|x| - 2.2}{2.5 - 2.2} - \frac{1}{2\pi} \sin \left( \frac{2\pi (|x| - 2.2)}{2.5 - 2.2} \right) \right) & |x| > 2.2 \\
0 & |x| \leq 2.2 
\end{cases}, \\
\zeta_2 (y) = \begin{cases} 
80 \left( \frac{|y| - 2.2}{2.5 - 2.2} - \frac{1}{2\pi} \sin \left( \frac{2\pi (|y| - 2.2)}{2.5 - 2.2} \right) \right) & |y| > 2.2 \\
0 & |y| \leq 2.2 
\end{cases}.
$$

While many choices were possible for the unscattered wave profile $U$, in this case a Gaussian pulse was chosen:

$$
U(z) = e^{-\omega^2(z+2)^2}.
$$

The initial displacement for the pulse was chosen to ensure that the pulse does not significantly overlap the boundary $\partial \Omega$ at time $t = 0$, so that the largest value that the function $u_{\text{wave}}$ could take on the boundary $\partial \Omega$ would be $e^{-\pi^2} \approx 5.1723 \times 10^{-5}$ (4d.p.).
In order to construct a finite difference grid for the above problem, the size of the grid was chosen as $x^* = 2.6$, which ensured the exterior points of the grid lay outside $\Lambda \cup \Sigma$ for any value of $N$ sufficiently large enough to give a reasonable level of accuracy.

4.2 Case 1: Perturbed bubble

The perturbed bubble is the unit circle with a large surface perturbation generated by six random parameters. The purpose of testing this shape was to have a relatively simple shape generated by multiple parameters over a relatively large range, in order to test that the UQ methods chosen would converge for the problem given the required regularity conditions. A CFL number of 0.5 was used for this case with a $300 \times 300$ finite difference grid.

4.2.1 Precise description of shape

To define this shape, 2D polar coordinates were used to define the outline as a function $r = f(\theta)$, where $r$ is the radial coordinate and $\theta$ is the angular coordinate, and then the level-set function for numerically representing the boundary was defined as $\Phi(r, \theta) = f(\theta) - r$. In defining the function $f(\theta)$, six uniformly distributed random parameters $\{R_1, R_2, R_3, R_4, R_5, R_6\}$, each on the interval $[0.8, 1.2]$, were chosen. The function was then defined as:

$$f(\theta) = f_0 + f_1 \cos(\theta) + f_2 \sin(\theta) + f_3 \cos(2\theta) + f_4 \sin(2\theta) + f_5 \cos(3\theta),$$

where the coefficients $\{f_0, \ldots, f_5\}$ were chosen such that $f(\Theta_j) = R_j$, where $\Theta_j = -\pi + \frac{1}{3}(j - 1)\pi$. This function can be shown to always lie within the interval $[0.4, 1.6]$, and will therefore provide a suitable shape for working with this problem. For more details, see the appendix. Two examples of this shape with their corresponding numerical solutions at $t \approx 2$ are shown below.

Note: for the first solution, the scattered wave has started to enter the PML region, which is why the scattering appears weaker in front of the object.
Figure 4.2: Outlines of two examples of the perturbed bubble shape, with the scattered wave at \( t \approx 2 \)

### 4.2.2 Tests of regularity for sparse grid method

To test the regularity of this shape, 40 test points were used with the method for testing all parameters given in section 3.4.2. This produced the following graphs for the measured energy approximations \( \bar{E}(0), \bar{E}(\frac{1}{2}\pi), \bar{E}(\pi) \) and their derivatives:
From the figure 4.3, it can be seen that the measured energy intensities at the three given angles are approximately continuous, which indicates that both MC and QMC should converge. There are oscillations in the derivatives, which indicates that this problem in UQ may not be sufficiently regular for sampling methods such as sparse grids, either due to the system itself being irregular, or due to the implementation of the boundaries, such as when a grid-point transitions from being in $\Omega$ to being in $D$.

### 4.2.3 Convergence study for the UQ methods

To analyse the convergence for this case, the solution obtained with quasi-Monte Carlo for 1600 sample points was assumed to be approximately correct, since the regularity checks indicated that the quantities of interest were approximately continuous in the random parameters. Doing this produced the following convergence plots for the mean and variance of $\bar{E}(\theta)$ at $\theta = 0, \frac{1}{2}\pi, \pi$:
Figure 4.4: Plots of the convergence with the number of sample points $N_{\text{points}}$ for the expected value of $\bar{E}(\theta)$ at three different points on the measurement ring.
Figure 4.5: Plots of the convergence with the number of sample points $N_{\text{points}}$ for the variance of $E(\theta)$ at three different points on the measurement ring.

In figures 4.4 (mean values of $\bar{E}$) and 4.5 (variance of $\bar{E}$), the general trend of the error with MC sampling can be seen to follow the expected maximum error convergence of $O\left(\frac{1}{N_{\text{points}}}\right)$, which implies that the choice of QMC as the model solution was suitable for the system. QMC also had a consistently lower error than MC after some value of $N_{\text{points}}$ for all three angles, as expected from the theoretical maximum error with sufficient regularity.

In each case the sparse grid method did not show significant convergence, as predicted by the regularity checks. This indicates that this problem in UQ may lack sufficient regularity for sparse grids to be effective.

4.2.4 Scattering profile

To gain an idea of what the distribution of the scattered energies around this shape looks like, polar plots of the mean energy $\bar{E}$ and $\bar{E} \pm \sigma_E$, where $\sigma_E$ is the variance, are presented below, when calculated by either MC with 1600 sample points, QMC with 1600 sample points or the sparse grid method at the
4\textsuperscript{th} level:

Figure 4.6: Polar plot of the mean scattered energy in each direction, and the range of values covered by one standard deviation, as calculated by MC with 1600 points.

Figure 4.7: Polar plot of the mean scattered energy in each direction, and the range of values covered by one standard deviation, as calculated by QMC with 1600 points.
Figure 4.8: Polar plot of the mean scattered energy in each direction, and the range of values covered by one standard deviation, as calculated by the sparse grid method at the fourth level.

From the convergence analysis, it is known that all three methods obtain an error of less than $10^{-2}$ for both the mean and variance. Here it can be visibly seen in figures 4.6 - 4.8 that the sparse grid method converges more slowly for the variance than either MC or QMC.

In terms of the shape of the profile, the large recorded values behind the object are due to the 'shadow' of the object being recorded as part of the scattering solution. The variance here is also effectively 0, which is expected from the object always containing a finite region around the origin with a radius $\rho \approx 0.8$. Given this is the shadow of the object, it is likely that the range of angles for which this is observed would decrease with a larger measurement ring, since the measurement points for $|\theta| < \arctan\left(\frac{\rho}{R_E}\right) \approx 0.3805(4d.p.)$ will always lie behind the object.

In front of the object the scattered energy is lower, however the variance is higher, which indicates the shape of the object has a significant effect on the directions in which energy is reflected. To give a more complete idea of the profile, the following density plot is provided, where the value at each point corresponds to the probability density of $\bar{E}(\theta)$ taking that value, for a fixed value of $\theta$: 
CHAPTER 4. NUMERICAL RESULTS

Figure 4.9: Probability density plot of the recorded energy intensity $\bar{E}(\theta)$, where $\theta$ (given in radians) is fixed in each case. QMC sampling was used to generate the plot.

From figure 4.9 it can be seen that the energy is more variable for the front two thirds of the object. The range of angles for which the value of $E$ is effectively constant across the random variables appears to correspond very closely to the range of measurement points which lie directly behind the object, which indicates this is would change with a larger value of $R_E$.

4.3 Case 2: Artillery shell

This test case gives an approximate outline of an artillery shell in 2D. The purpose is to provide a test case in which the level-set is discontinuous and the resulting boundary has sharp corners, however the shape itself varies continuously with the random parameters that define it.

4.3.1 Precise description of shape

The level-set for this case was generated by three random parameters, representing the radius ($R$), length ($L$) and a parameter describing the eccentricity of the curved edges $(y_c)$. These parameters were uniformly distributed over the following intervals:
The level-set function was then given in cartesian coordinates by:

\[
\Phi(x, y) = \begin{cases} 
1 - a \left(x - \frac{1}{2}L\right)^2 - b \left(|y| - y_c\right)^2 & x \leq \frac{1}{2}L \\
-1 & x > \frac{1}{2}L 
\end{cases}
\]

\[
a = \frac{R^2 - 2Ry_c}{L^2 \left(R - y_c\right)^2}
\]

\[
b = \frac{1}{\left(R - y_c\right)^2}
\]

The result is a numerical boundary composed of two elliptical arcs and one parabolic arc approximating a vertical line segment, since the boundary is calculated by linear interpolation on the finite difference grid. The elliptical arcs are chosen to be reflections of each other in the x-axis, to pass through the points \((-\frac{1}{2}L, 0)\) and \((\frac{1}{2}L, \pm R)\) and to be parallel to the x-axis at \((\frac{1}{2}L, \pm R)\). Examples of this shape with the corresponding solution at \(t \approx 2\) are presented below:

Figure 4.10: Plots of the artillery shell shape with different input parameters, and the corresponding solution at \(t \approx 2\)
4.3.2 Regularity tests

To test the regularity of the solution, the same approach of selecting a random line in the sample space was taken, which yielded the following three graphs:

![Plots of $\bar{E}(\theta)$ and its derivatives along a randomly selected line in sample space](image)

Figure 4.11: Plots of $\bar{E}(\theta)$ and its derivatives along a randomly selected line in sample space

From the figure 4.11, the value of $\bar{E}(\pi)$ appears to behave regularly for most values of $\mu$, with only small fluctuations in derivatives occurring for $\mu \leq 0.1$. In addition, the oscillations in the derivatives of $\bar{E} \left( \frac{1}{2} \pi \right)$ are relatively small, which could indicate that the sparse grid method will perform reasonably well for this shape. In addition, the improved regularity seen above, when compared to the perturbed bubble, indicates that the issues with regularity seen before are due to the dimensionality of the system itself, rather than being a property of the spatial discretisation.

4.3.3 Convergence tests

To test the convergence of the sampling methods for this shape, 2000 sample points were taken for MC and QMC sampling, and sparse grids were used to
the fifth level of accuracy. Once again, the solution with the largest number of sample points from QMC was used as a model solution, due to the continuity of the tested quantities. This resulted in the following convergence plots:

Figure 4.12: Convergence plots for the expected value of $\bar{E}(\theta)$ at $\theta = 0, \frac{1}{2}\pi, \pi$
Here the same trend is seen with MC sampling, which indicates that the choice of the QMC solution with 2000 points as a model solution was reasonable. In addition, for most of the plots in figures 4.12 and 4.13, QMC can also be seen to have a noticeably faster convergence rate than MC for the number of sample points taken, with convergence beginning to approach $O\left(\frac{1}{N_{\text{points}}}\right)$.

The sparse grid method had a better performance for this test case, and often performed equally well or better than the QMC method for a comparable number of grid-points. This indicates that it is a good choice for calculating the mean and variance of this type of system for relatively simple shapes with a smaller number of dimensions.
### 4.3.4 Scattering profile

![Scattering profile for the shell test case using MC sampling with 2000 sampling points](image)

Figure 4.14: Scattering profile for the shell test case using MC sampling with 2000 sampling points.
Figure 4.15: Scattering profile for the shell test case using QMC sampling with 2000 sampling points

Figure 4.16: Scattering profile for the shell test case using sparse grid integration at the fifth level
In figures 4.14 - 4.16, the overall appearance of the plots is the same. This fits with the results seen in the convergence analysis, where all three sampling methods were seen to converge for both expected values and variance.

It can be seen that the variance and expected value directly in front of the shell are effectively 0. This fits with intuition about the system, since the shell has a pointed tip at the front, at which no point is parallel to the wave-front. As a result, it is expected that no energy is reflected directly back, but instead deflected away at an angle.

In this case the variance is high behind the object, where previously a relatively constant shadow was seen. This is likely due to the shell having a much broader range of widths (along the y-axis) than the perturbed bubble. This means the only a very slim range of measurement points are guaranteed to be directly behind the object. This also means that the effects of diffraction are more likely to noticeably the results here. To get a better idea of the distribution of $\bar{E}(\theta)$, a probability density plot is presented below:

![Probability density distribution for $\bar{E}(\theta)$ for the shell test case. The figure was generated using QMC sampling.](image)

Figure 4.17: Probability density distribution for $\bar{E}(\theta)$ for the shell test case. The figure was generated using QMC sampling.

In figure 4.17 it can be seen that behind the object there is a small range of values where the variance begins to reduce, however there is still noticeably
more variance than before. The range of angles to the side of the object where
the variance begins to increase due to the variable curvature of the sides can
also be more clearly seen here.

4.4 Case 3: Array of four bubbles

This test case gives an array of four non-overlapping circles placed almost
completely randomly within a circular region of radius 1.75 centred on the
origin. The purpose here is to test how the results of UQ will change when
there are multiple smaller objects that can potentially interact with each other’s
scattering profiles, and to what extent this will violate the regularity constraints
of the methods used.

4.4.1 Precise description of shape

In order to generate an array of $n_b$ bubbles, $3n_b + 1$ random parameters were
used. Of these, $n_b$ were used to define radial coordinates of the centres of
the bubbles ($\mathbf{R} = [R_1, \ldots, R_{n_b}]$, $R_j \sim U[0.25, 1.5]$), $n_b$ to define the lo-
cal angular coordinates ($\Xi, \Xi_j \sim U[\pi/4n_b, 7\pi/4n_b]$), $n_b$ to define individual radii
($S, S_j \sim U[0.5, 1]$) and one to define an off-set angle $\Theta_0 \sim U[0, 2\pi/n_b]:$

1. The circular region of radius 1.75 is divided into $n_b$ angular segments,
each with an angular range of $\frac{2\pi}{n_b}$

2. The angular segments are all rotated about the origin anti-clockwise by
an angle of $\Theta_0$

3. The three parameters $R_j, \Xi_j, S_j$ are used to define the bubble in segment
$j$, lying between angular coordinates $\Theta_0 + \frac{2\pi}{n_b} (j - 1)$ and $\Theta_0 + \frac{2\pi}{n_b} j$,
$j = 1, \ldots, n_b$

4. The centre of the bubble in segment $j$ is defined in 2D polar coordinates
as $(R_j, \theta_j)$:

$$\theta_j = \Theta_0 + \frac{2\pi}{n_b} (j - 1) + \Xi_j.$$

5. The minimum distance from the centre of each bubble to either the edge
of the circular region or the edges of segment $j$ is calculated:

$$\rho_{max,j} = \min \left\{ 1.75 - R_w, R_w \sin \Xi_j, R_w \sin \left( \frac{2\pi}{n_b} - \Xi_j \right) \right\}$$
6. The radius of the bubble in segment $j$ is calculated as

$$\rho_j = \rho_{\text{max},j} S_j$$

To illustrate what this shape looks like with $n_b = 4$, four examples are presented below:

![Approximate solutions at t=3.005](image)

Figure 4.18: Plots of the bubble array shape when 4 bubbles are generated, with the solution at $t \approx 3$

### 4.4.2 Regularity tests

The same regularity checks were performed for the array of four bubbles as for the single perturbed bubble and the shell, resulting in the following plots:
In this test case, the values of $\bar{E}(\theta)$ can be seen to oscillate slightly for $\theta = 0$ and $\pi$. This in turn can be seen as larger oscillations in the first and second derivatives. In addition, for $\theta = \frac{1}{2} \pi$ the value of $\bar{E}$ is approximately constant for most value of $\mu$, until rising quickly for $\mu > 0.9$. Although the value appears to be continuous itself, the first derivative appears to be discontinuous, indicating this system may not be regular enough for good convergence with the sparse grid method.

### 4.4.3 Convergence tests

Below are convergence plots for the expected value and variance of $\bar{E}(\theta)$ for $\theta = 0, \frac{1}{2} \pi, \pi$. QMC with 2000 points was used as the model solution:
Figure 4.20: Convergence plots for the expected value of $\hat{E}(\theta)$ at $\theta = 0, \frac{1}{2} \pi, \pi$
Once again, the convergence of the maximum error using MC sampling follows the general trend predicted by theory. While QMC performs well compared to MC in some cases, in others there is less difference in performance between the two. This is likely due to the system having a higher number of dimensions than the previous two cases (13 dimensions, as opposed to 6 for the perturbed bubble or 3 for the shell). This indicates that for higher dimensions the convergence of QMC is likely to be adversely affected, except for very large values of $N_{\text{points}}$.

In this case the sparse grid method fails to converge, and performs significantly worse than both MC and QMC sampling, as expected from the regularity tests. The size of the error also remains $O(10^{-1})$ or worse, similar to or greater the size of $\hat{E}$ in most directions, meaning the results it produces cannot be reliably used.
4.4.4 Scattering profile

Figure 4.22: Scattering profile for an array of 4 bubbles using MC sampling with 2000 sampling points
Figure 4.23: Scattering profile for an array of 4 bubbles using QMC sampling with 2000 sampling points

Figure 4.24: Scattering profile for an array of 4 bubbles using sparse grid integration at the third level
In the figure 4.24, the issue with the sparse grid method for this case becomes obvious. Aside from the calculated expected values of $\bar{E}$ varying significantly from the correct solution, the calculated variance was negative for several angles, which can be seen as angles where there is no dotted red or blue line present. This indicates that as well as having a higher computational cost due to the increased number of dimensions, the sparse grid method does not produce reliable results for arrays of bubbles as defined above.

From figures 4.22 and 4.23 (for MC and QMC sampling respectively), the variance appears to be large for all measured angles, while the expected value of $\bar{E}$ was lower in front of the object and higher behind. To get a better idea of the distribution, a probability density plot generated by QMC sampling is included below:

![Probability density plot of $E(\theta)$ for an array of 4 bubbles](image)

Figure 4.25: Probability density plot of $E(\theta)$ for an array of 4 bubbles

Figure 4.25 indicates that while the variance was high for all angles, values at the lower end of the possible range were more probable in front of the object,
whereas values closer to 1 became marginally more probable behind the array, however the shadow is distinctly less visible than in the corresponding plots for the perturbed bubble or even shell. This is likely due to the finite probability that there is no part of $\Omega$ lying on the $x$-axis, meaning the wave is not directly blocked from the measurement points around $\theta = 0$.

Another feature of figure 4.25 is that the calculated density distribution appears to be less smooth than in the previous two cases, with the plot having a 'wrinkled' appearance. This is likely due to the higher dimensionality requiring more sample points for convergence.

### 4.5 Case 4: Array of many small bubbles

This test case is a generalisation of the previous to the case where there are many bubbles in the region of a smaller size, to see whether the results will converge as the number of bubbles increases while the size decreases. For practical purposes only MC and QMC methods were applied in this section, due to the high dimensionality of the problem creating excessively high computational costs for sparse grid methods.

#### 4.5.1 Precise description of shape

The bubble arrays used in this section were generated with the same algorithm presented in the previous section. The difference was that while before the number of bubbles used was $n_b = 4$, a relatively low quantity, here the number of bubbles used is 20 or higher. Four examples of what this shape will look like with $n_b = 20$, including the numerical solution at $t \approx 3$ are given below:
Figure 4.26: Plots of the bubble array shape when 20 bubbles are generated

4.5.2 Expected results

Due to the nature of the algorithm, as $n_b$ increases, the maximum size of the bubbles decreases. Overall however, the expected energy measured at various points around the measurement ring should converge for large values of $n_b$ to a non-zero value. The derivation of this can be found in the appendix.

4.5.3 Tests with 20 bubbles

For this case, the number of dimensions is 61. This meant that the number of grid-points required for even the second level of sparse grids became impractically large, even without considering the requirement of more grid-points due to a more detailed shape. However, since QMC sampling has some regularity requirements, tests of regularity were still relevant. The same test performed for an array of 20 bubbles as for the previous test cases yielded the following:
Figure 4.27: Plots of $\vec{E}(\theta)$ and its derivatives along a randomly selected line in sample space

In this case, the oscillations seen for the first two test cases are almost negligible compared to the calculated values of the derivatives, leading to graphs that appear smoother by comparison. However, this means the variations seen in figure 4.27 are most likely due to the system itself being irregular. Even for the values of $\vec{E}(\pi)$, there is significant change between adjacent values of $\mu$ in the discretisation, which could potentially indicate either discontinuities or changes rapid enough to pose an issue, at least for $\theta = \pi$.

Below convergence analysis is performed for MC and QMC sampling. In this case, due to the shape of the graphs in the regularity checks, both MC and QMC sampling with 2000 points were used as model solutions, to better assess the validity of QMC:
Figure 4.28: Convergence graphs for the expected energy $\bar{E}$ a bubble array with 20 bubbles, using MC sampling with 2000 points as a model solution.
Figure 4.29: Convergence graphs for the expected energy $\bar{E}$ for a bubble array with 20 bubbles, using QMC sampling with 2000 points as a model solution.
Figure 4.30: Convergence graphs for the variance of $\bar{E}$ for a bubble array with 20 bubbles, using MC sampling with 2000 points as a model solution.
From the figures 4.28 and 4.30, the MC method shows convergence with the expected rate when the MC solution with $N_{\text{points}} = 2000$ is chosen as the model. In some of the cases shown in figures 4.29 and 4.31 this is also seen when QMC is used for the model solution, however there are also cases where the MC does not appear to converge when using this as a model. Since in these cases QMC also appears to offer similar accuracy or worse for many values of $N_{\text{points}}$ and taking into consideration the regularity tests, this is likely due to QMC being unreliable for the values of $N_{\text{points}}$ tested for this system. As such, to obtain greater accuracy for this system than seen in figures 4.28 - 4.31, many more sample points would be required.

4.5.4 Convergence tests with $n_b$ bubbles, $n_b \to \infty$

To test how the mean and variance of $\hat{E}$ change with $n_b$, tests were performed using $n_b = 4, 10, 20, 40, 60, 80$. Since in the tests performed for regularity and convergence for $n_b = 20$ it was seen that QMC sampling performed either similarly or worse than standard MC sampling, particularly when com-
pared to its performance for $n_b = 4$, MC sampling with $N_{\text{points}} = 2000$ was used for each value of $n_b$. For the numerical system, a $200 \times 200$ grid was used for $n_b = 4, 10$ and a $300 \times 300$ grid was used for $n_b \geq 20$ (grid size was limited due to the computational costs of calculating 2000 sample solutions). Below are graphs showing the trend in mean and variance of $\bar{E}(\theta)$ for $\theta = 0, \frac{1}{2}\pi, \pi$:

![Graph showing trends in expected value of scattered energy from an array of $n_b$ bubbles](image)

![Graph showing trends in the variance of scattered energy from an array of $n_b$ bubbles](image)

Figure 4.32: Convergence graphs for the mean and variance of $\bar{E}(\theta)$ for different $n_b$

From figure 4.32, it can clearly be seen that the mean converges to non-zero values for the chosen angles as $n_b$ increases. This is in keeping with the predictions made, based on the expected value of the radius of a given bubble in the array, and the number of bubbles within a certain distance of a given measurement point.

For the variance, it is harder to tell, however it is unlikely that the convergence is towards $\sigma^2 = 0$, given the possible variations in the distance of each bubble from the measurement ring for any value of $n_b$.

The convergence seen in figure 4.32 implies that the practical limitations on the size of the grid did not adversely affect the accuracy of the simulations up to $n_b = 80$, since if a significant number of bubbles were not being represented numerically by the grid then after some value of $n_b$ the number of bubbles actually modelled would start to decrease due to increasingly small radii allowing no grid-points to lie inside $\Omega$, and consequently less energy would be scattered. This would mean that the expected energy calculated would begin to converge to 0, which is not seen here. However, it is likely that for some large $n_b$ this would have an effect. Consequently, the simplest way to approximate the statistics of $\bar{E}(\theta)$ for large $n_b$ is to use the convergence seen above.
4.5.5 Scattering profiles

Finally, to provide an insight into what the scattering profile of an array of many bubbles looks like, below are the scattering profiles and probability density plots of the measured energy for $n_b = 10, 20, 80$, using MC with 2000 sample points:

Figure 4.33: Scattering profile and probability density plot of $\bar{E}(\theta)$ for an array of 10 bubbles

Figure 4.34: Scattering profile and probability density plot of $\bar{E}(\theta)$ for an array of 20 bubbles
From figures 4.33 - 4.35, it can be seen that generally the profile has different properties in front of and behind the array. Behind the array, as \( n_b \) increases the shadow becomes larger and less variable. This is likely due to larger \( n_b \) meaning the bubbles would typically be more evenly distributed, but also the definition of the bubble array meaning all bubbles laid in an annular region around the origin. This meant that for lower \( n_b \) it was more likely that some range of \( y \)-values close to 0 would experience greater variation in whether any part of \( \Omega \) had a \( y \)-coordinate in that range. In the probability plots the distribution of values appears to be roughly symmetric for each \( \theta \)-value.

In front of the object, the mean and variance converged quickly, and the variance remained significant even for high \( n_b \). In this region the recorded energy was due to reflections from the array, indicating that even for large numbers of bubbles, the distribution and size of those bubbles would have a significant effect on the exact profile recorded. From the probability plots, the probability density here is skewed towards smaller values of \( \bar{E}(\theta) \).
Chapter 5

Summary

In this section a brief summary of work conducted and conclusions drawn is made. In addition suggestions for further research based on questions and technical constraints that arose during the investigation are suggested.

5.1 Review of work conducted and conclusions

Various sampling techniques were applied to scattering from an object with a Dirichlet boundary for the basic wave equation in 2D. It was found that in front of an object with a highly variable surface, the reflected energy was often highly variable as well, with lower values typically being more probable. It was found that the outline of an object not being smooth in and of itself did not mean the resulting system would be irregular, as long as the outline varied smoothly with the random parameters. It was found that if the object definitely occupied a sufficiently large region, then the shadow relatively close behind that region would have a low variance, even as the rest of the shape varied. The tests using the bubble array showed that when a shape $\Omega$ becomes complex enough for its scattering pattern to itself reflect off of $\Omega$, the regularity of the system would be reduced significantly enough to render sparse grid integration unreliable. Tests on the bubble array for large numbers of bubbles indicated that even as a shape becomes increasingly complex, as long as it retains certain characteristics observable properties of the resulting system can converge to finite and calculable quantities.
5.2 Suggestions for further research

The first extension of this work would be to investigate the source of the irregularity seen above. This could be achieved by applying different numerical methods, such as finite volume or discontinuous finite element methods, as well as using a wider range of grid resolutions to investigate the how grid-point transitions affect the solution.

Another extension would be to consider the related problem of wave equation in 3D. With the resources used for this paper this was impractical, due to the increased size of the numerical system. However the finite difference method used can be easily generalised to higher dimensions, and the sampling methods used can be easily parallelised with a parallel computing environment, making the increased computation time manageable.

Another potential extension based on having more computational power available is to use a larger domain $D$, and hence larger measurement rings. This would allow for a full investigation on how the radius of the measurement ring corresponds to the measured values $\bar{E}(\theta)$.

Finally, this paper only considered a narrow range of boundary conditions on $\partial \Omega$: Dirchlet conditions with a pre-determined function. Only a single norm for the scattered energy was considered as well. Further research could focus on how the measured energy changes with a generalised boundary condition on $\partial \Omega$, or a more arbitrary incoming wave than $U$. Since a clear relationship between different ways of measuring the scattered energy could not be assumed with this problem, this could also include an investigation of the profile obtained using an integral of energy over time instead of taking a maximum value.
Appendix A

A.1 Surface function for the perturbed bubble

To show that the generated function $f$ for the perturbed bubble is non-zero and less than 1.75 for all $\theta$, so that it can be used with radial coordinates and the object it represents remains within a circle of radius 1.75 centred at the origin, upper and lower limits for the function are calculated by replacing the trigonometric functions with combinations of their upper and lower limits. In order to represent all the different combinations, first define an $n \times 2^n$ matrix $B_n$ such that the $j^{th}$ column of $B_n$ is a representation in binary code of $j - 1$. That is, $B_n$ is the unique matrix satisfying:

(i) Every entry of $B_n$ is either 0 or 1,

(ii) $[2^{n-1} \; 2^{n-2} \; \ldots \; 2 \; 1] \; B_n = [0 \; 1 \; 2 \; \ldots \; 2^n - 1]$.

The resulting matrix $B_n$ can now be used to represent all the possible outcomes from a series of $n$ binary choices. In addition, let $1^{m \times n}$ be an $m \times n$ matrix in which every entry is 1.

To calculate upper and lower bounds of $f$, first consider that
\[ f(\theta) = \begin{bmatrix} 1 & \cos(\theta) & \sin(\theta) & \cos(2\theta) & \sin(2\theta) & \cos(3\theta) \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{bmatrix}. \]

Each of the trigonometric functions can take a value between \(-1\) and \(1\). Therefore, if both the minimum and maximum values of the function when the trigonometric functions are all replaced with either a \(1\) or \(-1\) are within the limits, then the function in general must always remain within the specified limits. The different combinations of these values can be represented using the matrix \(B_5\). In addition, let \(M\) be the \(6 \times 6\) matrix used to calculate \(f_0, \ldots, f_5\):

\[
M = \begin{bmatrix}
R_1 \\
R_2 \\
R_3 \\
R_4 \\
R_5 \\
R_6
\end{bmatrix},
\]

Then for a given set of values for \(R_1, \ldots, R_6\), the function \(f\) will be bounded by the largest and smallest entries in the vector given by:

\[
F_R = \begin{bmatrix} -1^{32 \times 1} & 1^{32 \times 1} \end{bmatrix} M^{-1} \begin{bmatrix} R_1 \\
R_2 \\
R_3 \\
R_4 \\
R_5 \\
R_6
\end{bmatrix} + \begin{bmatrix} B_5^T \\
0.4 B_6
\end{bmatrix}.
\]

Finding the maximum and minimum values for each of these entries, and therefore for \(F\) in general, is a linear programming problem in \(\{R_1, R_2, R_3, R_4, R_5, R_6\}\). As the parameter space is a \(6^{th}\) dimensional hypercube, the maximum and minimum values of \(F\) will occur at one of the vertices of the parameter space. These vertices are represented by the matrix \(0.8 \times 1^{6 \times 64} + 0.4 B_6\). The function is thus bounded by the maximum and minimum values of the matrix:

\[
F = \begin{bmatrix} -1^{32 \times 1} & 1^{32 \times 1} \end{bmatrix} M^{-1} \left( 0.8 \times 1^{6 \times 64} + 0.4 B_6 \right).
\]
This matrix can easily be calculated, giving the upper and lower bounds:

\[
\min_{i,j} F_{ij} = 0.4357 \text{(4 d.p.)},
\]
\[
\max_{i,j} F_{ij} = 1.5643 \text{(4 d.p.)}.
\]

Therefore the zero contour of the level-set \(\Phi(r, \theta) = f(\theta) - r\) is always well-defined and confined within a circular region of radius 1.75 about the origin, as required for the problem.

### A.2 Expected results for many bubbles

To begin with, only one bubble is considered. For a Gaussian pulse of fixed amplitude \(A\) and fixed wave-speed \(c\), the energy \(E_{\text{scatter}}\) reflected by a circular Dirichlet scatterer of radius \(r\) will be proportional to \(A^2 \cdot 2r\), which is a measure of the amount of energy directly incident on the object. Assuming that only the radius of the scatterer \(r\) is allowed to vary, this implies that for circular scatterers, the total reflected energy is proportional to the radius of the scatterer. To use this to make predictions about the calculated scattering profile, this will be assumed to be proportional to the maximum energy density scattered in a given direction.

Let \(E_{\text{scatter}}(n_b)\) be the total energy scattered from a single bubble in an array of size \(n_b\). For simplicity of notation, call it the last bubble. Then we have, for some \(K\) (depending on the mathematical system being modelled):

\[
\mathbb{E} [E_{\text{scatter}}] = K \mathbb{E} [\rho_{n_b}],
\]
\[
= K \mathbb{E} [\rho_{\text{max}, n_b}] \mathbb{E} [S_{n_b}], \quad S_{n_b}, \rho_{\text{max}, n_b} \text{ are independent},
\]
\[
= \frac{3}{4} K \mathbb{E} [\rho_{\text{max}, n_b}].
\]

From the algorithm, we have that

\[
\rho_{\text{max}, n_b} = \min \left\{ 1.75 - R_{n_b}, R_{n_b} \sin (\Xi_{n_b}), R_{n_b} \sin \left( \frac{2\pi}{n_b} - \Xi_{n_b} \right) \right\}.
\]

By considering the last two entries in the list of values that \(\rho_{\text{max}, n_b}\) could take and allowing \(\Xi_{n_b}\) to vary, at least one of them must be less than or equal to
$R_{nb} \sin \left( \frac{\pi}{n_b} \right)$, which itself can be at most $1.5 \sin \left( \frac{\pi}{n_b} \right)$. Since the first entry on list is always greater than or equal to 0.25, $\rho_{\text{max,}nb}$ will always take one of the last two entries for $n_b > \frac{\pi}{\arcsin (\frac{1}{4})} \approx 18.7616 \ (4d.p)$, which covers the range of values of $n_b$ that are of interest ($n_b \to \infty$). We can therefore simplify the above expression:

$$
\rho_{\text{max,}nb} = R_{nb} \sin \left( \min \left\{ \left( \Xi_{nb} \right), \left( \frac{2\pi}{n_b} - \Xi_{nb} \right) \right\} \right).
$$

The expression can be simplified even further by considering the random variable $Y$, given by

$$
Y = \frac{n_b}{\pi} \min \left\{ \Xi_{nb}, \left( \frac{2\pi}{n_b} - \Xi_{nb} \right) \right\}.
$$

Firstly, $\frac{1}{4} \leq Y \leq 1$. Secondly, by considering the probability function $F_Y(y) = \mathbb{P}(Y \leq y)$ for $y \in \left[ \frac{1}{4}, 1 \right]$, it can be shown that $Y$ is uniformly distributed on this interval:

$$
F_Y(y) = \mathbb{P}(Y \leq y) = 1 - \mathbb{P}(Y > y),
$$

$$
= 1 - \mathbb{P} \left( \frac{n_b}{\pi} \Xi_{nb} > y, 2 - \frac{n_b}{\pi} \Xi_{nb} > y \right),
$$

$$
= 1 - \mathbb{P} \left( \frac{\pi}{n_b} y < \Xi_{nb} < \Xi_{nb} < \frac{(2-y)\pi}{n_b} \right),
$$

$$
= 1 - \frac{\left( \frac{2(1-y)\pi}{n_b} \right)}{\left( \frac{7\pi}{4n_b} - \frac{\pi}{4n_b} \right)},
$$

$$
= 1 - \frac{4}{3} \left( 1 - y \right),
$$

$$
= \frac{4}{3} y - \frac{1}{3} = \frac{y - \frac{1}{4}}{1 - \frac{1}{4}}.
$$

Therefore $\rho_{\text{max,}nb}$ can be written in the simplified form:

$$
\rho_{\text{max,}nb} = R_{nb} \sin \left( \frac{\pi}{n_b} \left( \frac{3}{4} Y^* + \frac{1}{4} \right) \right), \quad Y^* \sim U[0, 1], Y^* = \frac{4}{3} Y - \frac{1}{3}
$$

Returning to the scattered energy $E_{\text{scatter}}$, this expression for $\rho_{\text{max,}nb}$ gives:

$$
\mathbb{E} [E_{\text{scatter}}] = \frac{3}{4} K \mathbb{E} \left[ R_{nb} \sin \left( \frac{\pi}{n_b} \left( \frac{3}{4} Y^* + \frac{1}{4} \right) \right) \right],
$$
\[= \frac{3}{4} K \mathbb{E} [R_{nb}] \mathbb{E} \left[ \sin \left( \frac{\pi}{4nb} (3Y^{*} + 1) \right) \right], \]

\[= \frac{21}{32} K \left( \sin \left( \frac{\pi}{4nb} \right) + \int_{\sin \left( \frac{\pi}{4nb} \right)}^{\sin \left( \frac{\pi}{nb} \right)} \left( \frac{4nb}{3\pi} \arcsin s - \frac{1}{3} \right) ds \right), \]

\[= \frac{21}{32} K \left( \frac{4}{3} \sin \left( \frac{\pi}{4nb} \right) - \frac{1}{3} \sin \left( \frac{\pi}{nb} \right) + \frac{4nb}{3\pi} \left[ s \arcsin s + \sqrt{1 - s^2} \sin \left( \frac{\pi}{4nb} \right) \right] \right), \]

\[= \frac{21}{32} K \left( \sin \left( \frac{\pi}{4nb} \right) + \sin \left( \frac{\pi}{nb} \right) + \frac{4nb}{3\pi} \left( \cos \left( \frac{\pi}{nb} \right) - \cos \left( \frac{\pi}{4nb} \right) \right) \right), \]

\[= \frac{21}{32} K \left( \frac{5\pi}{4nb} - \frac{2\pi}{3nb} + \frac{\pi}{24nb} + \mathcal{O} \left( n_{b}^{-3} \right) \right). \]

Therefore as \(n_{b}\) increases, the expected energy scattered by a single bubble in the array decreases with order \(\mathcal{O} \left( n_{b}^{-1} \right)\).

To connect the measured energy intensity to the total energy scattered by each individual bubble, it is assumed that the main contribution to the maximum energy density at each point on the measurement ring is due the bubbles in the array closest to it; that all the bubbles within a range of distances \(a \in [2 - r^{*}, 2]\), where \(r^{*}\) is the radial coordinate of the centre of the bubble closest to the measurement point. This assumption is based on the idea that scattered energy from other bubbles will itself be scattered by the bubbles closer to the measurement point, and become more diffuse as a consequence.

Due to the radial coordinate of the centre of the bubbles being independent of the value of \(n_{b}\), as \(n_{b}\) becomes large the expected distance from any given bubble to the nearest point in the measurement ring will converge to a fixed value. This gives an approximately constant range for \(a\) as defined above. In addition, the resulting area from which a bubble can affect a given measurement point will cover a fixed range of angular coordinates, so the expected number of bubbles within this area will be proportional to \(n_{b}\). This means for a given measurement point, there will be approximately \(\mathcal{O} \left( n_{b} \right)\) bubbles capable of contributing noticeably to the measured energy. This means the expected value of the measured energy will be approximately \(\mathcal{O} \left( n_{b} E_{\text{scatter}} \right)\) for large \(n_{b}\). Therefore, the expected value of the energy at the measurement ring \(\mathbb{E} [E(\theta)]\) should converge to a fixed value for large \(n_{b}\).
References


