Embedding Methods for the Numerical Solution of Convolution Equations

James Atkinson

August 31, 2004

Abstract

In this report we study linear integral operators generated by a difference kernel. The embedding results of Porter [8] for such operators are extended to Toeplitz matrices which have analogous structure. This structure is preserved by a basic collocation discretisation of such integral equations in the sense that it results in a Toeplitz matrix equation. It follows that for this discretisation embedding can be employed at the computational level.

This application of the theory is used in the numerical solution of a simple two dimensional diffraction problem. When Green's method is used to reduce the dimension of the problem, embedding results can be applied to the resulting one dimensional equation. Our computations show how a thorough study can be completed numerically with embedding as a useful tool.

Acknowledgements

This project was funded by an EPSRC studentship.

The project was supervised by Dave Porter to whom I am very grateful for the first hand account of an interesting subject and for the engaging supervision meetings. I would also like to thank Simon Chandler-Wilde for help with a problem in Analysis.

Declaration

This report is entirely the work of the stated author.

.....

Contents

Chapter 1

Background

In this chapter we review some developments which provide context and motivation for this work. Our review is not comprehensive, we focus on the particular theory and application pertinent here.

1.1 Embedding

In the Theory section of this report we are able to be quite precise about what we mean by embedding. In a wider context embedding has come to mean that the solution of a problem can be acquired directly from the solutions of other problems. Although vague this implies a structural result for the mapping in question. A class of equations which exhibit a lot of structure involve convolution operators which lead directly to embedding results.

1.2 Convolution Equations

Gokhberg and Feldman [2] studied extensively convolution structure in a broad sense, they derive results in both continuous and discrete systems. Here we are concerned with integral operators generated by a difference kernel on a finite interval which are referred to in [2] as truncated Wiener-Hopf integral operators. They prove an embedding result for such operators which is suggested by an analogous result for Toeplitz matrices (or truncated discrete Winer-Hopf operators). More extensive embedding results of Porter [8] for the same integral operators were motivated by a particular application which we now describe.

1.3 Wave Diffraction

Many physical phenomena are described reasonably well by the scalar wave equation, for example surface waves on water, pressure waves in a fluid and electromagnetic waves. These correspond to increasing magnitudes of wave-number in

ratio to the dimensions of the geometry. One can take advantage of approximations that can be made for large wave-numbers (short wavelengths) which are well established in the theory of electromagnetic wave propagation and referred to as Fresnel (second-order) and Fraunhofer (first-order) analysis. The use of exact, or near-field analysis, is necessary when the wave-number is smaller, like for water-wave and acoustic scattering problems.

A fundamental problem in this area, which can be considered in 2D, is the description of the diffracted field due to a plane wave incident on a thin barrier of infinite extent in which there is a single gap. This problem has an exact solution because the underlying boundary value problem is separable. However the solution is expressed in terms of Mathieu functions, for example see Carr and Stelzriede [1], which is unsatisfactory for computation because their evaluation is difficult. An alternative computation is available by first reducing the problem using Green's method. This results in a one dimensional equation which involves an integral operator generated by a difference kernel on a finite interval (the gap). In fact the converse problem, diffraction around a finite thin straight barrier, can be reduced in a similar fashion and results in an equation which involves the same operator. These equations were shown to be useful for computation by Gilbert and Brampton [6]. The application of embedding in these problems has been well established.

An early embedding result for the far-field pattern was given in the case of the second problem by Lebedev and Polishchuk [4]. Further embedding results, which relied on playing the problems off against each other, have been established by Porter and Chu [7]. Embedding for the total field of the first problem in isolation can now be verified as an instance of the more general embedding results established by Porter [8] as a property of the integral operator with difference kernel.

1.4 Recent Developments

In fact this 2D problem lends itself to interpretation as the surface of the ocean, so numerical simulations would produce information useful in Engineering problems such as the design of off-shore wind-farms, oil-rigs and coastal defences. Clearly such an enterprise would require analysis of geometries more complex than the single gap and its converse. This has motivated the study of various scattering problems which have led to embedding results applicable to a wider class of equations which govern more complex diffracted fields. Biggs et al [11] study an infinite thin straight barrier with several gaps and find embedding results for operators with a difference kernel on a union of finite intervals. This work is paralleled by Sakhnovich [9] who derives other embedding formula for such operators in a more abstract setting. This theme is extended by Biggs and Porter when they consider a perforated barrier with non-zero thickness [12] and two parallel thin perforated

barriers [13]. The class of equations for which embedding results are known to apply are expanded by these papers, to those involving integral operators with a matrix difference kernel and operators generated by sum-and-difference kernels.

1.5 Toeplitz Matrices

The embedding results considered in this report are connected to those of Sakhnovich [9], which was shown by Porter in [8]. The results of Sakhnovich apply to integral operators however he also presents a discrete version applicable to Toeplitz matrices which is attributed to A.L.Sakhnovich [3]. This result bears the same relationship to the discrete case presented here as it does in the continuous case.

1.6 This Project

It can be beneficial to preserve structure when a continuous model is discretised for computation. For example one is often able to determine salient information about a solution, like symmetries or conserved quantities. We can place additional demands on our numerical method to ensure such features remain during computation. This project has that theme. The structure in this case is very rich, and as a result puts quite strict constraints on the discrete system.

In chapter 2, the theory section, we exhibit the structural components which lead to the embedding formula of Porter [8]. These are given discrete analogues so that we can derive an embedding result that applies to the analogous discrete system.

Chapter 3 concerns itself with the formulation of an example problem. This is the single gap diffraction problem and its converse which were mentioned earlier, these problems are ideal for our purposes because the embedding results are very well understood. The formulation is pretty standard, but the nature of this report means that our exposition is thorough. At the end of the chapter we directly apply the results of the theory section to the integral equation which governs the diffracted field, we can thus give the embedding results specific to this problem.

In chapter 4 we detail the implementation of a very simple collocation method for the numerical solution of the 1D integral equation which governs the diffracted field. We can apply the results of the theory section to the discrete system which results in a discrete embedding formula. For the purposes of comparison we also implement a second method which is equally as simple as the first, but makes an advantageous use of non-equal discrete intervals across the domain. This method does not preserve embedding so we complement it with an approximate embedding formula by discretising the continuous one. The performance of the methods is compared using numerical error analysis so that conclusions can be drawn. Some

pictures of diffraction are included to illustrate this application of numerical methods.

Chapter 5 concludes the report and suggests avenues of investigation which it has thrown up.

Chapter 2

Theory

This chapter establishes what we mean by convolution equations and embedding. The development is abstract but we follow closely our motivational examples of the integral equation with a difference kernel and the Toeplitz matrix equation.

Initially for convenience and later through necessity we assume the vector space in which we work is over the field $\mathbb{C}$.

2.1 The Generator of Convolutions

At the heart of our discussion is a definition which associates with each vector in our space a linear transform on that space.

Definition. *If V is a linear space then we refer to $T : V \longrightarrow \mathrm{Hom}(V, V)$ as a Generator of Convolutions when, given any $x, y \in V$ and $\lambda, \mu \in \mathbb{C}$,*

(i) $T(\lambda x + \mu y) = \lambda T(x) + \mu T(y)$,

(ii) $T(x)y = T(y)x$,

(iii) $T(T(x)y) = T(x)T(y)$.

For example,

1. $h = T(f)g \quad \Rightarrow \quad h(x) = \displaystyle\int_0^x f(x-t)g(t)\,dt, \quad f, g, h \in L_2([0,1])$,

2. $u = T(v)w \quad \Rightarrow \quad u_i = \displaystyle\sum_{j=0}^{i} v_{i-j} w_j, \quad u, v, w \in \mathbb{R}^n$,

3. $h = T(f)g \quad \Rightarrow \quad h(x) = \displaystyle\int_{-\infty}^{\infty} f(x-t)g(t)\,dt, \quad f, g, h \in L_2(\mathbb{R})$ and

4. $z = T(z_0)z_1 \quad \Rightarrow \quad z = z_0 z_1 \quad z, z_0, z_1 \in \mathbb{C}$

generate convolutions on the indicated spaces.

Convolution is usually encountered as a binary operation defined on V, for example here we would write

$$x \otimes y = T(x)y.$$

In this context (*ii*) and (*iii*) are axioms of commutativity and associativity so that V forms a commutative ring with respect to addition and convolution of vectors. Note that the final axiom to get from commutative ring to field is invertibility which we have illustrated by example 4.

Given T is non-standard we should explain our purpose in its introduction. We define T because it provides a natural way to explore the interaction of convolution with its adjoint when V is a Hilbert space.

It turns out that $\mathrm{Im}(T)$ forms a commutative ring with respect to addition and composition of linear transforms. This statement is entirely contained within the definition of T except for commutativity which is straightforward to verify. In fact given (*ii*) it follows that (*iii*) is equivalent to

(iv) $T(x)T(y) = T(y)T(x)$,

(*iii*) $\Rightarrow$ (*iv*) is trivial, (*iv*) $\Rightarrow$ (*iii*) because for any $z \in V$, $T(T(x)y)z = T(z)T(x)y$ $= T(x)T(z)y = T(x)T(y)z$. This equivalence is useful in verifying that our examples fit the definition.

The examples given of T make it appear less pervasive than it actually is. In finite dimensions one can solve $T(x) = K$ for T given x and suitable K. This is achieved by choosing a basis set and writing $T(e_i) = A_i$, the resulting linear system consisting of the axioms for T together with $\sum x_i A_i = K$. More general existence and uniqueness results are not immediately obvious. However this does point a direction for development which we do not pursue here.

2.2 Basic Embedding

The return we get for identifying convolution structure includes embedding formula for the solutions of convolution type equations. Here we demonstrate the ideas involved by giving some simple examples.

By solving the linear system $A\phi = y$ we determine a vector in the inverse image of y under A. We say that other solutions are embedded when we can use such solutions to directly acquire vectors in the inverse image of vectors belonging to a larger subset of V. Note that here we refer to A as of convolution type when $A = \kappa - T(x)$ for $\kappa \in \mathbb{C}$ and $x \in V$.

1. Suppose that A and B are of convolution type, then we get a straightforward embedding result from the fact that A and B commute. If $A\phi = y$ it follows that $AB\phi = By$ so that $B\phi$ is in the inverse image of By under A.

2. If y is such that $T(y) = 1$ then $A\phi = y$ implies $AT(\phi) = T(y) = 1$ so that A is invertible and its inverse is given by $T(\phi)$. Here we get the inverse image of every vector in V.

3. If $A = \kappa - T(x)$ and ϕ is such that $A\phi = x$ then $AT(\phi) = T(x) = \kappa - A$ so that $A(1 + T(\phi)) = \kappa$. If $\kappa \neq 0$, A is invertible and its inverse is given by $\frac{1}{\kappa}(1 + T(\phi))$.

In the next section the convolution structure is somewhat narrowed by a further assumption whilst the type of equation considered is broadened.

2.3 The Adjoint of Convolution

If we suppose T is defined on a Hilbert space, we can consider $T^* : V \longrightarrow \mathrm{Hom}(V, V)$ which maps x to the adjoint of $T(x)$. This is a mild abuse of notation, for although T is a linear transform, T^* is not the adjoint of T.

It is straightforward to verify, through the definition, that

(v) $T^*(\lambda x + \mu y) = \overline{\lambda} T^*(x) + \overline{\mu} T^*(y)$,

(vi) $T^*(T(x)y) = T^*(x)T^*(y)$.

T^* does not generate convolutions on V. However noting that (vi) implies closure under composition as well as commutativity (use (ii) in (vi)) we see that $\mathrm{Im}(T^*)$ forms a commutative ring. Although much structure is preserved in T^*, our next assumption replenishes it fully.

From this point we assume there exists a self-adjoint, unitary conjugate linear map (see appendix A), P, such that

$$PT^*(x) = T(x)P$$

for all $x \in V$. This is not true in general; however for examples 1 and 2 of section 2.1 it is easily verifiable that P defined by

1. $g = Pf \quad \Rightarrow \quad g(x) = \overline{f}(1 - x)$ and

2. $v = Pu \quad \Rightarrow \quad v_i = \overline{u}_{n-i}$

satisfy the above. The following equivalence reveals how the convolution structure returns to T^* when such a P exists.

Theorem. If T is injective and generates convolutions on V and if P is an invertible conjugate linear map on V,

$$PT^*(x) = T(x)P, \quad \forall x \in V \quad \Longleftrightarrow \quad Q(x) = T^*(Px) \Rightarrow \begin{array}{l} Q \text{ generates} \\ \text{convolutions on } V. \end{array}$$

Proof.

$\Rightarrow$) Axioms (i) and (iv) are trivial, it remains to verify axiom (ii) .

$$\begin{aligned} Q(x)y &= T^*(Px)y \\ &= P^{-1}T(Px)Py \\ &= P^{-1}T(Py)Px \\ &= T^*(Py)x \\ &= Q(y)x. \end{aligned}$$

$\Leftarrow$) Assuming Q is a generator of convolutions we can use (vi) to start the following string of implication

$$\begin{aligned} & Q^*(Q(P^{-1}y)x)z = Q^*(P^{-1}y)Q^*(x)z \quad \forall x, y, z \in V \\ \Rightarrow \quad & T(PT^*(y)x)z = T(y)T(Px)z \\ \Rightarrow \quad & T(z)PT^*(y)x = T(z)T(y)Px \\ \Rightarrow \quad & T(z)Hx = 0 \ \ \forall x, z \in V \quad \text{where } H = PT^*(y) - T(y)P \\ \Rightarrow \quad & T(Hx)z = 0 \\ \Rightarrow \quad & T(Hx) = 0 \ \ \forall x \in V \\ \Rightarrow \quad & H = 0 \quad \text{by assuming the injectivity of } T \\ \Rightarrow \quad & PT^*(y) = T(y)P \quad \forall y \in V. \end{aligned}$$

□

Our move now is in the direction developed by Porter in [8]. The motivation was to find embedding formula for linear operators of the form

$$A = \kappa - T(l) - T^*(m).$$

Our previous examples are generalised in the following way,

1. $h = Ag \quad \Rightarrow \quad h(x) = \kappa g(x) - \displaystyle\int_0^1 f(x-t)g(t)dt \quad$ on $L_2([0,1])$,
   where $l(x) = f(x)$ and $m(x) = \overline{f}(-x)$, $x \in [0,1]$ and
2. $u = Aw \quad \Rightarrow \quad u_i = \kappa w_i - \displaystyle\sum_{j=0}^{n} v_{i-j}w_j \quad$ on $\mathbb{R}^n$,
   where $l_i = v_i$, $i \in \{0, 1 \ldots n\}$, $m_0 = 0$ and $m_i = \overline{v}_{-i}$, $i \in \{1, 2 \ldots n\}$.

In this context our demand that P be self-adjoint and unitary means that $P = P^{-1}$ which enables us to write $PA = A^*P$, this relation being critical in the following theory.

2.4 The Projection Property

The subset of V to which our embedding method applies is defined as follows. If there exists $\sigma \in \mathbb{R}$ so that $X = \sigma + T(x) + T^*(x)$ is the projection onto x, i.e.

$$X\phi = \langle\, \phi \,|\, x \,\rangle x$$

for all $\phi \in V$, then we say that x has the projection property. It turns out that we can find all such vectors for our core examples, although the arguments are slightly lengthy.

1. Suppose that f has the projection property, then for all $\phi \in L_2([0,1])$

$$\sigma\phi(x) + \int_0^x f(x-t)\phi(t)dt + \int_x^1 \overline{f}(t-x)\phi(t)dt = f(x)\int_0^1 \phi(t)\overline{f}(t)dt.$$

We first establish that f is continuous and $\sigma = 0$. Choosing ϕ continuous and not orthogonal to f and noting that $f \in L_2([0,1])$ we see the two indefinite integrals, and hence the LHS, are continuous. This then implies the continuity of f on the RHS. Now choose ϕ with a jump discontinuity in $(0,1)$, again the indefinite integrals are continuous and we have also established that f is, the only way to reconcile this with the discontinuity in ϕ is if $\sigma = 0$. With this information we can write the above equation as

$$\int_0^1 k(x,t)\phi(t)dt = 0, \quad x \in [0,1] \quad \forall \phi \in L_2([0,1]), \quad \text{where}$$

$$k(x,t) = \begin{array}{ll} f(x-t) - f(x)\overline{f}(t), & 1 \geq x \geq t \geq 0 \\ \overline{f}(t-x) - f(x)\overline{f}(t), & 1 \geq t \geq x \geq 0. \end{array}$$

This together with the continuity of f implies that $k = 0$ everywhere. Putting $k = 0$ in the above, when $t = x$ we see that $f(0) = \overline{f}(0) = |f(x)|^2$ for all $x \in [0,1]$ which implies, for $f \neq 0$, that $|f(x)|^2 = 1$, $x \in [0,1]$. Writing $y = x - t$ we can rearrange to find

$$f(y+t) = f(y)f(t), \quad (y+t) \in [0,1].$$

This tells us that, on a restricted domain, f is a group homeomorphism between $\mathbb{R}$ under addition and the unit circle in $\mathbb{C}$ under multiplication. This is uniquely satisfied by the exponential map

$$f(x) = e^{-i\gamma x}, \quad x \in [0,1]$$

to within some constant $\gamma \in \mathbb{R}$.

2. If v has the projection property then for all $\phi \in \mathbb{R}^n$

$$\sigma\phi_k + \sum_{j=0}^{k} v_{k-j}\phi_j + \sum_{j=k}^{n} \overline{v}_{j-k}\phi_j = v_k \sum_{j=0}^{n} \phi_j \overline{v}_j.$$

It is sufficient that this be true for the canonical basis so that choosing $\phi = e_l$ we find that v must satisfy

$$v_{k-l} = \overline{v}_l v_k, \qquad l < k, \tag{2.1}$$

$$\sigma + v_0 + \overline{v}_0 = \overline{v}_k v_k, \qquad l = k \quad \text{and} \tag{2.2}$$

$$\overline{v}_{l-k} = \overline{v}_l v_k, \qquad l > k, \tag{2.3}$$

for $k \in \{0, 1 \ldots n\}$. Clearly from 2.2 $|v_0|^2 = |v_1|^2 = \ldots = |v_n|^2$ so that $v_k = 0 \Rightarrow v = 0$. Pursuing non-trivial v we can assume $v_k \neq 0$ for all k, hence choosing $l = 0$ in 2.1 we can write $v_k = \overline{v}_0 v_k$ which implies $v_0 = 1$ therefore $|v_0|^2 = |v_1|^2 = \ldots = |v_n|^2 = 1$. This in fact fixes σ because 2.2 now implies $\sigma + 1 + 1 = 1$ so that $\sigma = -1$. Choosing $l = 1$ in 2.1 and $l = k + 1$ in 2.3 gives the recurrence $v_k = v_1 v_{k-1}$ and $v_{k+1} = v_1 v_k$ which both imply that v necessarily satisfies $v_k = v_1^k$. In fact such a v is also sufficient to satisfy 2.1 and 2.2, which is easy to verify. Finally, v_1 can be any number with unit magnitude, we choose $v_1 = e^{-i\gamma/n}$ where $\gamma \in \mathbb{R}$ is arbitrary. Thus we have determined that

$$v_j = e^{-i\gamma j/n}, \quad j \in \{0, 1 \ldots n\}.$$

We can establish some elementary facts about vectors which have the projection property. First we notice that X is self-adjoint so, from the end of the last section, it commutes with P and hence X and P share eigenvectors. Clearly then, because x is the sole eigenvector of X, it is also an eigenvector of P. Appendix A documents properties of such eigenvectors which we shall use freely in the following. Note that in both our examples we have contrived to write the projection type vectors so that their eigenvalue under P is $e^{i\gamma}$.

If v and w have the projection property then under suitable conditions we can find $\alpha \in \mathbb{C}$ so that

$$[1 + \alpha(\sigma + T(v))]w = v. \tag{2.4}$$

To see this suppose that v and w have distinct eigenvalues under P of λ and μ respectively. The following manipulation then hinges on the fact that $T^*(v)w =$

$T^*(Pw)Pv = \lambda\overline{\mu}T^*(w)v$.

$$\begin{aligned}(1-\lambda\overline{\mu})(\sigma+T(v))w &= (\sigma+T(v)+T^*(v))w - \lambda\overline{\mu}(\sigma+T(v))w - T^*(v)w \\ &= \langle\, w \,|\, v \,\rangle v - \sigma\lambda\overline{\mu}w - \lambda\overline{\mu}T(w)v - T^*(Pw)Pv \\ &= \langle\, w \,|\, v \,\rangle v - \lambda\overline{\mu}(\sigma+T(w)+T^*(w))v + \sigma\lambda\overline{\mu}(v-w) \\ &= \langle\, w \,|\, v \,\rangle v - \lambda\overline{\mu}\langle\, v \,|\, w \,\rangle w + \sigma\lambda\overline{\mu}(v-w) \\ &= (\langle\, w \,|\, v \,\rangle + \sigma\lambda\overline{\mu})(v-w).\end{aligned}$$

Assuming $\langle\, v \,|\, w \,\rangle + \sigma\lambda\overline{\mu} \neq 0$ we get the desired result finding that

$$\alpha = \frac{1-\lambda\overline{\mu}}{\langle\, w \,|\, v \,\rangle + \sigma\lambda\overline{\mu}}.$$

Operating on equation 2.4 with P we get the parallel result

$$\overline{\lambda}\mu[1+\overline{\alpha}(\sigma+T^*(v))]w = v. \tag{2.5}$$

These results are applicable directly in the context of example 1 in section 2.2. However, it turns out these operators *almost* commute with the more general convolution type operator; in fact the commutator is of finite rank. It is this which enables us to construct embedding formula for the more general equation.

2.5 The Embedding Formula of Porter (1991)

Let us refer to the subset of V with the projection property under a common σ as S and suppose our linear operator $A = \kappa - T(l) - T^*(m)$ is injective. Our main embedding result is the construction of an operator which maps the inverse image under A of any two distinct vectors in S to the inverse image of a third.

To this end we suppose w_0, w_1 and $v \in S \cap \mathrm{Im}(A)$ with distinct eigenvalues under P of μ_0, μ_1 and λ. We construct V_0 and V_1 so that $V_0 w_0 = V_1 w_1 = v$ as we did at the end of the previous section (equation 2.4); we write $V_0 = 1+\alpha_0(\sigma+T(v))$ and $V_1 = 1+\alpha_1(\sigma+T(v))$.

Consider the commutator of V_0 with A,

$$\begin{aligned}V_0A - AV_0 &= T^*(m)V_0 - V_0T^*(m) \\ &= \alpha_0(T^*(m)T(v) - T(v)T^*(m)) \\ &= \alpha_0(T^*(m)[\sigma+T(v)+T^*(v)] - [\sigma+T(v)+T^*(v)]T^*(m)),\end{aligned}$$

which is the commutator of $T^*(m)$ with a projection operator and hence of finite rank as mentioned earlier. Suppose we label the solutions such that $A\psi_0 = w_0$, $A\psi_1 = w_1$ and $A\phi = v$, using the above result we see that

$$(V_0A - AV_0)\psi_0 = \alpha_0\langle\, \psi_0 \,|\, v \,\rangle T^*(m)v - \alpha_0\langle\, \psi_0 \,|\, T(m)v \,\rangle v.$$

Now, $V_0 A\psi_0 = V_0 w_0 = v$ and collecting together multiples of $v = A\phi$ in the above and for convenience writing the constant as c_0 we have that

$$A(c_0\phi - V_0\psi_0) = \alpha_0\langle\,\psi_0 \mid v\,\rangle T^*(m)v.$$

Notice here that if $\langle\,\psi_0 \mid v\,\rangle = 0$ we arrive straightaway at an embedding result. More generally we can use the parallel result

$$A(c_1\phi - V_1\psi_1) = \alpha_1\langle\,\psi_1 \mid v\,\rangle T^*(m)v,$$

to eliminate $T^*(m)v$. Doing this and again collecting multiples of ϕ, then using the injectivity of A we find

$$c\phi = \alpha_1\langle\,\psi_1 \mid v\,\rangle V_0\psi_0 - \alpha_0\langle\,\psi_0 \mid v\,\rangle V_1\psi_1$$

for some $c \in \mathbb{C}$. This is our main result. We have left c in the formula although it is found quite easily by the less direct route of taking the inner product of this with w_0 or w_1, so that

$$c\langle\,\phi \mid w_0\,\rangle = \alpha_1\langle\,\psi_1 \mid v\,\rangle\langle\,\psi_0 \mid V_0^* w_0\,\rangle - \alpha_0\langle\,\psi_0 \mid v\,\rangle\langle\,\psi_1 \mid V_1^* w_0\,\rangle.$$

It is easy to verify that $V_0^* w_0 = \lambda\overline{\mu}_0 v$ and $\overline{\alpha}_0 V_1^* w_0 = (\overline{\alpha}_0 - \overline{\alpha}_1)w_0 + \overline{\alpha}_1\lambda\overline{\mu}_0 v$ and on substitution this reveals

$$c\langle\,\phi \mid w_0\,\rangle = (\alpha_1 - \alpha_0)\langle\,\psi_0 \mid v\,\rangle\langle\,\psi_1 \mid w_0\,\rangle.$$

Then $\langle\,\phi \mid w_0\,\rangle = \mu_0\langle\,\phi \mid PA\psi_0\,\rangle = \mu_0\langle\,\phi \mid A^*P\psi_0\,\rangle = \mu_0\langle\,v \mid P\psi_0\,\rangle = \mu_0\langle\,\psi_0 \mid Pv\,\rangle = \overline{\lambda}\mu_0\langle\,\psi_0 \mid v\,\rangle$ so that assuming v is such that $\langle\,\psi_0 \mid v\,\rangle \neq 0$ (or $\langle\,\psi_1 \mid v\,\rangle \neq 0$) we have found

$$c = \lambda\overline{\mu}_0\langle\,\psi_1 \mid w_0\,\rangle(\alpha_1 - \alpha_0) = \lambda\overline{\mu}_1\langle\,\psi_0 \mid w_1\,\rangle(\alpha_1 - \alpha_0).$$

The second result can be deduced directly because $\langle\,\psi_1 \mid w_0\,\rangle = \mu_0\overline{\mu}_1\langle\,\psi_0 \mid w_1\,\rangle$ using the same reasoning as above. We can now write the embedding formula in full, as

$$\lambda\overline{\mu}_0\langle\,\psi_1 \mid w_0\,\rangle(\alpha_1 - \alpha_0)\phi = \alpha_1\langle\,\psi_1 \mid v\,\rangle V_0\psi_0 - \alpha_0\langle\,\psi_0 \mid v\,\rangle V_1\psi_1. \tag{2.6}$$

As a corollary to the main result, we also give an embedding formula for the inner products between vectors in S and their inverse image under A. We suppose additionally that $A\psi_2 = w_2 \in S$ and take the inner product of 2.6 with w_2. By reasoning exactly as for the evaluation of c we find that

$$\begin{aligned} & \alpha_2(\alpha_1 - \alpha_0)\overline{\mu}_0\langle\,\psi_1 \mid w_0\,\rangle\langle\,\psi_2 \mid v\,\rangle \\ + \;& \alpha_0(\alpha_2 - \alpha_1)\overline{\mu}_1\langle\,\psi_2 \mid w_1\,\rangle\langle\,\psi_0 \mid v\,\rangle \\ + \;& \alpha_1(\alpha_0 - \alpha_2)\overline{\mu}_2\langle\,\psi_0 \mid w_2\,\rangle\langle\,\psi_1 \mid v\,\rangle \quad = 0. \end{aligned} \tag{2.7}$$

Equations 2.6 and 2.7 enable us to find ϕ and $\langle\,\phi \mid w_2\,\rangle$ given ψ_0 and ψ_1 such that $\alpha_1 - \alpha_0 \neq 0$ and $\langle\,\psi_0 \mid w_1\,\rangle \neq 0$.

2.6 Remarks

To summarise, we have introduced convolution as a set of commutative linear transforms in one-to-one correspondence with an underlying inner product space, V. Our main supposition is the existence of P, a conjugate linear similarity transform between each convolution map and its adjoint. The idea of a generalised convolution operator, A, being a linear combination of the identity, a convolution map and the adjoint of a convolution map then becomes our focus. We define a set $S \subseteq V$ of vectors with the projection property and suppose V is such that S is non-empty. Finally we use convolution structure and projection property to construct an embedding formula for vectors in the inverse image of S under A.

It should be noted that the inverse image of vectors in the span of S are also embedded in the same sense, the calculation being a little more cumbersome.

Chapter 3

A Diffraction Problem

Here we study a simple two-dimensional diffraction problem which we follow from an elementary beginning through to computation. To set the scene, we consider $u(t) : \mathbb{R}^2 \longrightarrow \mathbb{R}$ which satisfies the wave equation

$$\nabla^2 u = \frac{\partial^2 u}{\partial t^2}\,.$$

We separate variables choosing the separation constant $-k^2$, $k \in \mathbb{R}$, and writing $u_k = \phi(x, y)\tau(t)$; ϕ and τ then satisfy

$$\left(\nabla^2 + k^2\right)\phi = 0 \quad \text{and} \quad (\tfrac{\partial^2}{\partial t^2} + k^2)\tau = 0.$$

It follows that $\tau = A_k \cos(kt) + B_k \sin(kt)$ for some A_k,B_k.

We refer to u_k as a quasi steady-state solution; it is everywhere periodic in time with fundamental period ω given by $k\omega = 2\pi$. A more general initial value problem can be solved by integrating such solutions over k. This is not our aim here however, the extra layer demonstrating nothing further about embedding as well as being unnecessary for a myriad of interesting problems (i.e. those which are quasi steady-state by nature).

3.1 The Dual Scattering Geometries

We consider the spatial part of the diffracted field due to plane waves incident on two simple geometries. Our barriers are thin, straight and hard, in the first case of finite extent, in the second case of infinite extent with a finite gap. Figure 3.1 establishes the details. In the diagrams, I is an open interval in X (the x-axis) centred on the origin, ϕ_i is the incident plane wave and ϕ_r its reflection through X. ϕ_f and ϕ_g are the diffracted fields in the *finite barrier* and *gap* geometries. We have written the solution, ϕ, in each case as the sum of the diffracted field and the known plane-wave solution when $I = \emptyset$.

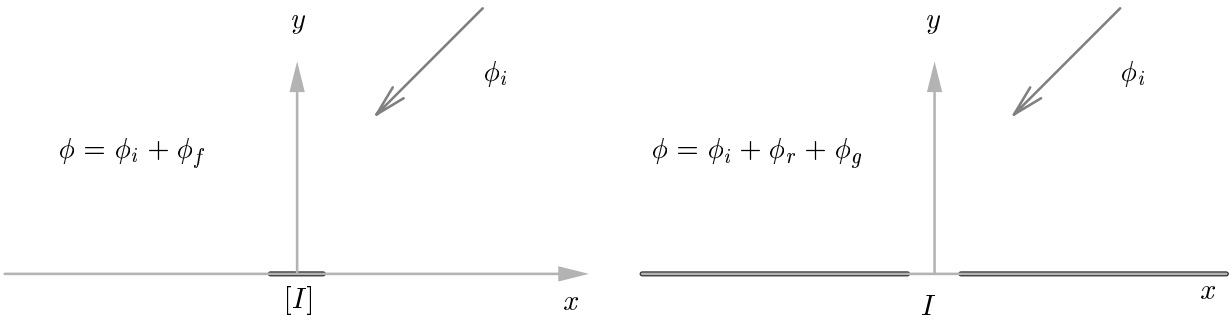

Figure 3.1: The complementary geometries.

3.2 The Solution in the Far-Field

In the far-field, I looks arbitrarily small. This gives physical plausibility to the important demand we place on the diffracted field, that it satisfy the Sommerfeld radiation condition,

$$\sqrt{r}\left(\frac{\partial}{\partial r} - ik\right)\phi_{f,g} \to 0 \quad \text{as } r \to \infty.$$

Solutions of the Helmholtz equation in two dimensions which satisfy this condition are travelling toward the far-field and decaying like cylindrical waves (i.e. they look as if they originate from a point source). Such solutions are also unique which is a fact with immediate consequence for our problems.

It turns out that ϕ_f and ϕ_g are anti-symmetric through X. Although this is not obvious it reduces the problem so is introduced early for the sake of simplicity. It is straightforward to verify that the gradient of either diffracted field is continuous everywhere through X. If we then write it as the sum of odd and even parts, both of which must satisfy the Helmholtz equation (because the operator $\left(\nabla^2 + k^2\right)$ conserves such symmetry), we notice the even part has zero derivative everywhere through X. But the function which is everywhere zero also has zero derivative through X, satisfies the Helmholtz equation and Sommerfeld radiation condition, by the uniqueness of this solution we see that the even part of either diffracted field is zero.

3.3 The Solution Close to a Barrier Edge

It is important later that we understand the behaviour of the diffracted field as we approach a barrier edge. Basic information is accessible to us by finding solutions of the Helmholtz equation separated in polar co-ordinates centred on the endpoint of a barrier in the limit $r \to 0$. Choosing the separation constant l^2, $l \in \mathbb{R}$, and

writing $\phi_l = \rho(r)\sigma(\theta)$, ρ and σ satisfy

$$\left(r^2\frac{\partial^2}{\partial r^2} + r\frac{\partial}{\partial r} + (kr)^2 - l^2\right)\rho = 0 \quad \text{and} \quad \left(\frac{\partial^2}{\partial\theta^2} + l^2\right)\sigma = 0.$$

Supposing the barrier lies along the line $\theta = 0$ we require that $\sigma'(0) = \sigma'(2\pi) = 0$ which leads to the constraint $\sigma = \cos(l\theta)$, $l \in \{0, \frac{1}{2}, 1, \frac{3}{2}\ldots\}$. The equation for ρ is Bessel's equation in the variable kr; for physically plausible solutions we demand that ρ remain bounded for all $r > 0$. This is violated by the Neumann functions so that ρ is the Bessel function of order l and we can conclude

$$\phi_l = A_l J_l(kr)\cos(l\theta), \quad l \in \{0, \tfrac{1}{2}, 1, \tfrac{3}{2}\ldots\},$$

from which it follows that

$$r \sim 0 \quad \Rightarrow \quad \phi_l = B_l(kr)^l\cos(l\theta) + O(r^{l+2}), \quad l \in \{0, \tfrac{1}{2}, 1, \tfrac{3}{2}\ldots\}$$

by using the expansion of J_l for small arguments (e.g. from [10]).

These are a subset of the harmonic functions (solutions of the equation of Laplace) in two dimensions. We have determined that the local behaviour of ϕ lies in the span of these functions and we note now a consequence of this which is important to us because it informs our method of computation. Specifically the gradient of $\phi_{\frac{1}{2}}$ becomes unbounded as $r \longrightarrow 0$ so that for $A_{\frac{1}{2}} \neq 0$ the gradient of ϕ is singular at $r = 0$. It is straightforward to confirm that the strength of the singularity in the gradient of $\phi_{\frac{1}{2}}$ past the barrier edge is characterised by

$$\frac{\partial\phi_{\frac{1}{2}}}{\partial y} \sim \frac{1}{\sqrt{x}}, \quad y = 0, x \to 0$$

and in fact for other directions the strength is the same.

3.4 The Elliptic Partial Differential Equations

By the symmetry of the diffracted field established in section 3.2, we can consider the upper half-plane, Ω, alone. Demanding that ϕ have zero derivative normal to the barrier and be continuous elsewhere we can write BVPs for the diffracted field in the two geometries as follows.

$$\begin{array}{rcll} \left(\nabla^2 + k^2\right)\phi_f & = & 0 & \text{in } \Omega \\ \sqrt{r}\left(\frac{\partial}{\partial r} - ik\right)\phi_f & \to & 0 & \text{as } r \to \infty \\ \phi_f & = & 0 & \text{on } X\backslash[I] \\ \frac{\partial}{\partial y}\left(\phi_f + \phi_i\right) & = & 0 & \text{on } [I] \end{array} \qquad \begin{array}{rcll} \left(\nabla^2 + k^2\right)\phi_g & = & 0 & \text{in } \Omega \\ \sqrt{r}\left(\frac{\partial}{\partial r} - ik\right)\phi_g & \to & 0 & \text{as } r \to \infty \\ \phi_g + \phi_i & = & 0 & \text{on } I \\ \frac{\partial}{\partial y}\phi_g & = & 0 & \text{on } X\backslash I \end{array}$$

It is worth the comment that ϕ_i, the incident wave which is the source of the diffracted field, appears only in the conditions on the finite interval in both the BVPs.

3.5 Reduction by Green's Method

We can establish an integral representation of the diffracted field using Green's method, for which we need a function G_0 which satisfies

$$\begin{array}{rcl} \left(\nabla^2 + k^2\right) G_0 & = & -\delta(x-x_0)\delta(y-y_0) \quad \text{in} \quad \Omega, \\ \sqrt{R}\left(\frac{\partial}{\partial R} - ik\right) G_0 & \to & 0 \quad \text{as} \quad R \to \infty, \end{array}$$

where $R^2 = (x-x_0)^2 + (y-y_0)^2$. It turns out that the first-kind Hankel function of order zero has the required properties and therefore

$$G_0(R) = \frac{i}{4} H_0^{(1)}(kR)\,.$$

G_0 is not unique in satisfying the above: if $\widetilde{R}^2 = (x-x_0)^2 + (y+y_0)^2$ then we can define

$$\begin{array}{rcl} G & = & G_0(R) + G_0(\widetilde{R}) \quad \text{and} \\ \hat{G} & = & G_0(R) - G_0(\widetilde{R}), \end{array}$$

which share the required properties with G_0 but are respectively symmetric and antisymmetric through X. We also note here the symmetry relations

$$\frac{\partial G}{\partial y} = -\frac{\partial \hat{G}}{\partial y_0}\,, \qquad \frac{\partial \hat{G}}{\partial y} = -\frac{\partial G}{\partial y_0}\,,$$

which are useful later. We aim to find an integral representation of the diffracted field with integration being over the finite interval I and this can be achieved in the complementary geometries with the converse choices G and $\hat{G}$.

At the heart of Green's method we substitute $\hat{G}$ and ϕ_f into Green's second identity to generate the integral representation. That is

$$\begin{array}{rcl} \displaystyle\int_\Omega \phi_f \nabla^2 \hat{G} - \hat{G}\nabla^2\phi_f \; d\mathbf{x} & = & \displaystyle\int_{\partial\Omega} \phi_f \frac{\partial \hat{G}}{\partial n} - \hat{G}\frac{\partial \phi_f}{\partial n} \; d\mathbf{x} \\ \Rightarrow \qquad -\phi_f & = & \displaystyle -\int_X \phi_f \frac{\partial \hat{G}}{\partial y} - \hat{G}\frac{\partial \phi_f}{\partial y} \; dx \quad \text{in } \Omega. \end{array}$$

The LHS follows from the delta-function property of $\hat{G}$. There is no contribution to the RHS integral from the boundary at infinity because both $\hat{G}$ and ϕ_f satisfy the Sommerfeld radiation condition, so the domain of integration becomes X on which $\frac{\partial}{\partial n} = -\frac{\partial}{\partial y}$. The representation further reduces because we chose $\hat{G}$ which is anti-symmetric and hence zero on X so the second term in the RHS integral does not contribute, then from the BVP we know that $\phi_f = 0$ on $X\backslash[I]$ so that we can restrict the domain to I, that is

$$\phi_f = \int_I \phi_f \frac{\partial \hat{G}}{\partial y} \; dx \quad \text{in } \Omega,$$

which is our desired integral representation of ϕ_f. Using G and ϕ_g in Green's second identity we find by similar reasoning that

$$\phi_g = -\int_I G\frac{\partial \phi_g}{\partial y}\ dx \quad \text{in } \Omega.$$

Of course, we do not know ϕ_f or $\frac{\partial \phi_g}{\partial y}$ on I; in fact we have reduced the total problem to one of finding these.

Note that the analysis of section 3.3 reveals its significance, informing us that $\frac{\partial \phi_g}{\partial y}$ is potentially singular close to the barrier edge which here means the endpoints of I.

3.6 The Reduced Equations

Let us defer ϕ_f for the moment; now we know ϕ_g on I from the BVP and the idea in this section is to take the limit of the integral representation of ϕ_g as we approach I and substitute its value there. This yields an integral equation for $\frac{\partial \phi_g}{\partial y}$ on I which we now denote by v to reflect its new found status at the centre of our problem. The integral equation for v is

$$\phi_i = \int_I G\, v\ dx \quad \text{on } I. \tag{3.1}$$

The route to an equation for ϕ_f on I is not so direct, the above procedure being unproductive essentially because what we know is $\frac{\partial \phi_f}{\partial y}$ on I which does not make an appearance in our integral representation of ϕ_f. Bearing this in mind we take the y_0 derivative of it and use the symmetry properties of G and $\hat{G}$ so that we can write

$$\frac{\partial \phi_f}{\partial y_0} = \int_I -\phi_f \frac{\partial^2 G}{\partial y_0^2}\ dx \quad \text{on } \Omega.$$

In the spirit of writing a one-dimensional equation we notice that $\frac{\partial^2}{\partial y_0^2} G = -(\frac{\partial^2}{\partial x_0^2} + k^2)G$ when $(x, y) \in X$ and $(x_0, y_0) \in \Omega$. Making this substitution, taking the differential operator outside the integral then taking the limit onto I we find that

$$-\frac{\partial \phi_i}{\partial y_0} = \left(\frac{\partial^2}{\partial x_0^2} + k^2\right) \int_I G\psi\ dx \quad \text{on } I. \tag{3.2}$$

We have given $\frac{\partial \phi_f}{\partial y_0}$ its value on I and changed notation from ϕ_f to ψ on I which like v is at our centre of focus. Boundary conditions for this equation can be found from the BVP, by the continuity of ϕ_f along X we require that $\psi = 0$ at the endpoints of $[I]$.

At this stage it is convenient to condense our notation and specify more precisely the incident plane wave. We suppose it arrives from a direction which makes an angle $\alpha \in [0, \pi]$ with the x-axis, i.e.

$$\phi_i = e^{-ik(x\cos\alpha + y\sin\alpha)}.$$

We define

$$f_\alpha(x) = e^{-ikx\cos\alpha}$$

which is the incident wave on I so that equations 3.1 and 3.2 are now written as

$$K\, v_\alpha = f_\alpha \tag{3.3}$$

$$\left(D^2 + k^2\right) K\, \psi_\alpha = ik\sin\alpha f_\alpha. \tag{3.4}$$

The subscripts on ψ and v indicate their dependence on α, $D = d/dx$ and K is the linear integral operator defined on I by the kernel

$$k(x, x_0) = \frac{i}{2} H_0^{(1)}(k|x - x_0|),$$

where we have used the fact that $R = \widetilde{R} = |x - x_0|$ when (x, y) and $(x_0, y_0) \in I$. We note here that the uniqueness of the diffracted field implies the injectivity of the operator K.

We now verify an elementary symmetry property of the solutions of these equations. Consider U, the reflection in the line $x = 0$ defined so that $g = Uf \Rightarrow g(x) = f(-x)$. Clearly the Helmholtz operator as well as our geometry is invariant under U so we can suppose that $UK = KU$, which is straightforward to confirm. Noting that $Uf_\alpha = f_{\pi-\alpha}$ and $\sin(\pi - \alpha) = \sin\alpha$ it follows from equations 3.3 and 3.4 that

$$Uv_\alpha = v_{\pi-\alpha} \quad \text{and} \quad U\psi_\alpha = \psi_{\pi-\alpha}. \tag{3.5}$$

So, we have established an integral equation for v_α and an integro-differential equation for ψ_α. The integral operator involved is of convolution type and it turns out, using the above symmetry, that solving for a particular angle of incident wave we can find embedded solutions for other angles. Along the embedding theme we also establish a link which enables us to find ψ_α from solutions to the v_α equation. Before doing this we introduce a practical definition to characterise the scattered wave in the far-field.

3.7 The Far-Field Diffraction Coefficient

If we take the view from the far-field the scatterer looks like a point source radiating cylindrical waves, but with amplitude dependent on the incident angle α and observation angle, θ. We refer to this angular variation of amplitude as the far-field diffraction coefficient. We find an expression for it by looking for the component of the diffracted field which falls off most gradually with distance from the scatterer. This involves a large argument asymptotic expansion of our integral representation of the diffracted field.

Taking polar co-ordinates in Ω and cartesian in the integration domain we can write

$$\begin{aligned} R^2 &= r^2 + x^2 + y^2 - 2r(x\cos\theta + y\sin\theta) \\ \widetilde{R}^2 &= r^2 + x^2 + y^2 - 2r(x\cos\theta - y\sin\theta). \end{aligned}$$

Asymptotically when r is much greater than $\sqrt{x^2+y^2} \sim 1$

$$r \sim \infty \quad \Rightarrow \quad \begin{array}{rcl} R & = & r - x\cos\theta - y\sin\theta + O(r^{-1}) \\ \widetilde{R} & = & r - x\cos\theta + y\sin\theta + O(r^{-1}). \end{array}$$

The asymptotic expansion of $H_0^{(1)}$ for large arguments (e.g. from [10]) leads directly to

$$R \sim \infty \quad \Rightarrow \quad G_0(R) = \frac{1}{2\sqrt{2\pi kR}} e^{i(kR+\frac{\pi}{4})} + O(R^{-\frac{3}{2}}).$$

We are now in a position to expand G and $\frac{\partial \hat{G}}{\partial y}$ asymptotically for $r \sim \infty$ and hence derive the far-field approximations to ϕ_f and ϕ_g, sparing the reader the details we find

$$r \sim \infty \quad \Rightarrow \quad \begin{array}{rcl} \phi_f & = & \dfrac{1}{\sqrt{2\pi kr}} e^{i(kr-\frac{3\pi}{4})} \left\{ ik\sin\theta \displaystyle\int_I \psi_\alpha \, e^{-ikx\cos\theta} \, dx \right\} + O(r^{-\frac{3}{2}}) \\ \phi_g & = & \dfrac{1}{\sqrt{2\pi kr}} e^{i(kr-\frac{3\pi}{4})} \left\{ \displaystyle\int_I v_\alpha \, e^{-ikx\cos\theta} \, dx \right\} + O(r^{-\frac{3}{2}}). \end{array}$$

The curly braces separate out the angular dependent part of the field which we now define as the far-field diffraction coefficients F and Γ corresponding to ϕ_f and ϕ_g. In terms of the $L_2(I)$ inner product,

$$\Gamma(\alpha,\theta) = \langle\, v_\alpha \,|\, f_{\pi-\theta} \,\rangle \qquad \text{and} \qquad F(\alpha,\theta) = ik\sin\theta \langle\, \psi_\alpha \,|\, f_{\pi-\theta} \,\rangle.$$

The symmetry property established in the previous section is also pertinent here. Consider U again: this operator is clearly unitary so following on from relations 3.5 we see that

$$\Gamma(\alpha,\theta) = \Gamma(\pi-\alpha,\pi-\theta) \quad \text{and} \quad F(\alpha,\theta) = F(\pi-\alpha,\pi-\theta).$$

3.8 The Dual Problems Linked

Here we give an algebraic equation relating ψ_α to v_α, v_0 and v_π. We start by establishing a property of K which we can state more generally for a linear integral operator generated by a kernel of the form $k(x,t) = s(x-t)$. We consider the effect of differentiating $K\phi$ where $\phi \in C^1(I)$ is arbitrary. Suppose, without loss of generality, $I = [-1,1]$ then we can manipulate as follows by a change of integration variable,

$$\begin{array}{rcl} \dfrac{d}{dx} \displaystyle\int_{-1}^{1} s(x-t)\phi(t)\, dt & = & \dfrac{d}{dx} \displaystyle\int_{x-1}^{x+1} s(q)\phi(x-q)\, dq \\ & = & \big[s(q)\phi(x-q) \big]_{x-1}^{x+1} + \displaystyle\int_{x-1}^{x+1} s(q)\phi'(x-q)\, dq \\ & = & s(x+1)\phi(-1) - s(x-1)\phi(1) + \displaystyle\int_{-1}^{1} s(x-t)\phi'(t)\, dt. \end{array}$$

If ϕ is zero at the endpoints of the domain this implies that

$$DK\phi = KD\phi.$$

By factoring the differential operator in the ψ_α equation (equation 3.4) we can commute one factor through K by using the above property which is valid because ψ_α is zero at the endpoints of I, and we find

$$(D + ik)K(D - ik)\psi_\alpha = ik \sin\alpha f_\alpha.$$

It is worth note that this manipulation is the first step in a subtle introduction of the boundary conditions on ψ_α. We can use the integrating factor e^{-ikx}, which we write here as f_0, to invert the leftmost differential operator resulting in

$$K(D - ik)\psi_\alpha = \cot(\tfrac{\alpha}{2})(f_\alpha - c_\alpha f_0)$$

where $c_\alpha \in \mathbb{C}$ is unknown. From the injectivity of K, the substitution of Kv_α for f_α and Kv_0 for f_0 leads directly to

$$(D - ik)\psi_\alpha = \cot(\tfrac{\alpha}{2})(v_\alpha - c_\alpha v_0). \tag{3.6}$$

The alternative choice of differential factors leads to the parallel equation

$$(D + ik)\psi_\alpha = \tan(\tfrac{\alpha}{2})(d_\alpha v_\pi - v_\alpha), \tag{3.7}$$

where again $d_\alpha \in \mathbb{C}$ is unknown. Now, we can eliminate $D\psi_\alpha$ between these to get the desired algebraic relation. Before doing this though, we pin down c_α and d_α by the second subtle use of our boundary conditions. Assuming ψ_α is zero at the endpoints of I it is straightforward to establish that

$$\begin{aligned} \langle\, (D \pm ik)\psi_\alpha \,|\, f_{\pi-\theta} \,\rangle &= \langle\, \psi_\alpha \,|\, (D^* \mp ik) f_{\pi-\theta} \,\rangle \\ &= ik(\cos\theta \pm 1)\langle\, \psi_\alpha \,|\, f_{\pi-\theta} \,\rangle \end{aligned} \tag{3.8}$$

which is equal to zero at $\theta = \pi, 0$. Hence by taking the inner product of equation 3.6 with f_0 and equation 3.7 with f_π we find

$$\begin{aligned} \langle\, v_\alpha - c_\alpha v_0 \,|\, f_0 \,\rangle &= 0 \quad \text{and} \\ \langle\, d_\alpha v_\pi - v_\alpha \,|\, f_\pi \,\rangle &= 0. \end{aligned}$$

Proceeding with the elimination of $D\psi_\alpha$ between equations 3.6 and 3.7 then substituting for c_α and d_α using these relations we arrive at the main result of this section, namely,

$$ik \sin\alpha \langle\, v_0 \,|\, f_0 \,\rangle \psi_\alpha = \cos^2(\tfrac{\alpha}{2})\langle\, v_\alpha \,|\, f_0 \,\rangle v_0 + \sin^2(\tfrac{\alpha}{2})\langle\, v_\alpha \,|\, f_\pi \,\rangle v_\pi - \langle\, v_0 \,|\, f_0 \,\rangle v_\alpha.$$

A similar result holds for the far-field diffraction coefficient $F(\alpha, \theta)$. Taking the inner product of equation 3.6 with $f_{\pi-\theta}$ and using relation 3.8 we find

$$\sin\alpha(\cos\theta - 1)F(\alpha,\theta) = \sin\theta(1+\cos\alpha)(\Gamma(\alpha,\theta) - c_\alpha\Gamma(0,\theta)), \qquad (3.9)$$

an algebraic equation for $F(\alpha, \theta)$ in terms of $\Gamma(\alpha, \theta)$, $\Gamma(0, \theta)$ and c_α. Clearly we might have started from equation 3.7 and in so doing found a parallel relation to this involving d_α. Between these we can eliminate $\Gamma(\alpha, \theta)$ and on substitution for c_α and d_α it turns out that

$$2(\cos\alpha + \cos\theta)\Gamma(0,\pi)F(\alpha,\theta) = \sin\alpha\sin\theta(\Gamma(\alpha,\pi)\Gamma(\pi,\theta) - \Gamma(\alpha,0)\Gamma(0,\theta)).$$

This relation has a deeper embedding property in that we get $F(\alpha, \theta)$ in terms of Γ for observation or incidence angles of 0 and π alone. In fact once $F(\alpha, \theta)$ has been established, 3.9 can then be used to calculate $\Gamma(\alpha, \theta)$ which leads to the embedding result for Γ. This can be decoupled from the equation for ψ_α as we show in the next section.

3.9 Embedding

We identify here the structural components which lead to embedding. The principal difference between this problem and example 1 of the theory section is that the interval, I, is taken here to be centred on the origin; in fact without losing generality we can choose $I = (-1, 1)$, I has been non-dimensionalised by the wave-number k.

A problem, which turns out to be benign, is that $v_\alpha \notin L_2([-1, 1])$ due to its singular nature. From the analysis of section 3.3 we know that $\sqrt{1-x^2}\ v_\alpha$, like ϕ_g on I, is a continuous function. This enables us to work around the issue by considering a larger space, E, which contains L_2, defined as the set of functions on $[-1, 1]$ for which all pairs f, $g \in E$ satisfy

$$\int_{-1}^{1} \sqrt{1-x^2}\ f(x)\overline{g}(x)\ dx \quad < \quad \infty.$$

We continue to understand inner product as that defined on L_2 only now we cannot assume it remains bounded for all pairs in E. In this context, we define T so that

$$h = T(f)g \quad \Rightarrow \quad h(x) = \int_{-1}^{x} f(x-t-1)g(t)\ dt,$$

for $x \in [-1, 1]$, that T generates convolutions on E is easily verifiable.

If P is such that $g = Pf \Rightarrow g(x) = \overline{f}(-x)$ then P is self-adjoint and unitary and that $PT^*(f) = T(f)P$ for all $f \in E$ is also straightforward. Finally, we restrict S, a non-trivial subset of E with the projection property, to lie within $L_2([-1, 1])$. It turns out that $S = \{e^{-i\gamma(x+1)}, x\in[-1:1] : \gamma \in \mathbb{R}\}$ with $\sigma = 0$. The

qualification here is that vectors in S have the projection property only for the subset $E_0 \subseteq E$ with which their inner product is defined. That everything in S is continuous means that, given our definition of E, $E_0 = E$.

The equation for v_α can then be written in the notation of chapter 2 as

$$(\kappa - T(l) - T^*(m))\phi = v \in S \tag{3.10}$$

where, for all $x \in [-1, 1]$,

$$\begin{aligned} \kappa &= 0, \\ l(x) &= \tfrac{-i}{2} H_0^{(1)}(k(x+1)), \\ m(x) &= \tfrac{i}{2} \overline{H}_0^{(1)}(k(x+1)), \\ \phi(x) &= e^{-ik\cos\alpha} v_\alpha(x) \quad \text{and} \\ v(x) &= e^{-ik\cos\alpha} f_\alpha(x). \end{aligned}$$

The eigenvalue of v under P is $e^{i2k\cos\alpha}$ and, if $w_0 \in S$ corresponds to an angle β, the constant α_0 is straightforward to evaluate, and we find

$$\alpha_0 = ik(\cos\beta - \cos\alpha).$$

By substituting these quantities into equation 2.6 we arrive at the embedding result for v_α. We can benefit from the symmetry relation 3.5 by choosing $w_1 = Uw_0$ which correspond to incident angles β and $\pi - \beta$ for some $\beta \neq \pi/2$. We find, after noting that $\langle\, \phi \,|\, w_0 \,\rangle = e^{-ik(\cos\alpha - \cos\beta)}\Gamma(\alpha, \pi - \beta)$ and so on, that

$$\begin{aligned} 2\cos\beta\Gamma(\beta,\beta)\ v_\alpha = {} & \Gamma(\beta,\alpha)(\cos\alpha + \cos\beta - ik(\cos^2\alpha - \cos^2\beta)T(v))\ v_\beta \\ & -\Gamma(\beta,\pi-\alpha)(\cos\alpha - \cos\beta - ik(\cos^2\alpha - \cos^2\beta)T(v))Uv_\beta. \end{aligned}$$

We can use equation 2.7 to get an embedding formula for Γ. We choose w_2 corresponding to angle $\pi - \theta$ and take the view that θ is the observation angle, we find

$$\begin{aligned} 2\cos\beta(\cos\theta + \cos\alpha)\ \Gamma(\beta,\beta)\ \Gamma(\alpha,\theta) \quad &= \\ (\cos\beta - \cos\alpha)(\cos\theta - \cos\beta)\ \Gamma(\beta,\pi-\alpha)\ \Gamma(\beta,\theta) & \\ + \quad (\cos\beta + \cos\alpha)(\cos\theta + \cos\beta)\ \Gamma(\beta,\pi-\theta)\ \Gamma(\beta,\alpha). & \end{aligned}$$

In this context, the embedding formula enable calculation of v_α for all α and $\Gamma(\alpha,\theta)$ for all α,θ directly from v_β for some $\beta \neq \pi/2$.

3.10 Remarks

In this chapter we have analysed the diffracted field due to plane waves incident on two simple geometries. Green's method has been employed to find an integral representation for the diffracted fields, reducing the problem in each case to the

solution of a one-dimensional equation on a finite interval. We find the link between solutions of the two problems so it is possible to compute one from the other, which begins the embedding theme. It is continued because the one-dimensional equations involve the same integral operator which is of generalised convolution type and we can apply the results of the theory section. This means we can express the diffracted field due to a plane wave incident at any angle in terms of only one. The corollary for inner products also has direct application here because it leads to embedding formula for the far-field diffraction pattern.

The computational benefit of embedding is made clear through this example problem. The solution of only one linear system is required in order to establish the diffracted field and far-field pattern for all angles of incident plane wave.

Chapter 4

The Numerical Solution

In this chapter we solve the diffraction problem numerically. This means that for a given k and α, we compute approximations to the far-field diffraction pattern on $[0, \pi]$ and the diffracted field on some subset of $\mathbb{R}^2$. The embedding results for α can be discretised to give a computational saving independent of the numerical method used. However, it is desirable that the discrete system preserves embedding. The following commutative diagram, in the notation of the theory section, clarifies what we mean by this,

$$\begin{array}{ccc} \psi_0, \psi_1 & \longrightarrow & \phi \\ \downarrow & & \downarrow \\ \widetilde{\psi}_0, \widetilde{\psi}_1 & \longrightarrow & \widetilde{\phi} \end{array}$$

where arrows to the right represent embedding and arrows down represent discretisation.

This can be achieved by careful implementation of a basic collocation method, which has the additional benefit of being extremely simple. We write a collocation discretisation of equation 3.3 and it turns out we can identify the structural components which lead to embedding. This enables us to write the discrete equivalent of equation 3.10 and then direct substitution into 2.6 brings us to the discrete embedding formula.

4.1 Discretisation

We choose to approximate functions on the interval $[-1, 1]$ at the set of points

$$x_j = \frac{2j - n}{n + 1}, \ j \in \{0, 1 \ldots n\},$$

the half-width of a discrete interval being $\delta = \frac{1}{n+1}$. Equation 3.3 can then be approximated by the matrix system

$$\widetilde{K}\widetilde{v}_\alpha = \widetilde{f}_\alpha, \tag{4.1}$$

where

$$\begin{aligned}\widetilde{f}_{\alpha j} &= f_\alpha(x_j) \quad \text{and} \\ \widetilde{K}_{jl} &= \frac{i}{2}\int_{x_l-\delta}^{x_l+\delta} H_0^{(1)}(k|x_j-t|)\ dt + O(\delta^2),\end{aligned}$$

in the sense that if $\widetilde{v}_\alpha$ satisfies 4.1 then $\widetilde{v}_{\alpha j} = v_\alpha(x_j) + O(\delta)$. In choosing this discretisation we sidestep the singularities in v_α at the endpoints, at least in the purely theoretical sense that $\widetilde{v}_\alpha$ remains bounded for all finite n. The practicalities of floating point representation mean that we need to return to this issue later. We should comment that the approximation in the definition of $\widetilde{K}$ is only because some approximate method must be used for the evaluation of the integrals.

This discretisation has preserved, in a natural way, the system's invariance under reflection in x. If we define $\widetilde{U}$ by $\widetilde{U}v_i = v_{n-i}$ so that $\widetilde{U}x_i = x_{n-i} = -x_i$, then it is straightforward to verify that $\widetilde{U}\widetilde{K} = \widetilde{K}\widetilde{U}$, hence, assuming the injectivity of $\widetilde{K}$,

$$\widetilde{U}\widetilde{f}_\alpha = \widetilde{f}_{\pi-\alpha} \quad \Rightarrow \quad \widetilde{U}\widetilde{v}_\alpha = \widetilde{v}_{\pi-\alpha}.$$

4.2 Discrete Embedding

The matrix $\widetilde{K}$ has Toeplitz structure as can be seen by writing the elements as

$$\widetilde{K}_{jl} = \frac{i}{2}\int_{-\delta}^{\delta} H_0^{(1)}(k|2\delta(j-l)-t|)\ dt + O(\delta^2),$$

because the integrals, and hence their approximations, are dependent only on the difference $j-l$. The structural elements which lead to embedding in this system are exactly those of example 2 of the theory section, to summarise we define T, P and S so that

$$\begin{aligned} u = T(v)w \quad &\Rightarrow \quad u_j = \sum_{i=0}^{j} v_{j-i}w_i, \\ u = Pv \quad &\Rightarrow \quad u_j = \overline{v}_{n-j} \quad \text{and} \\ v \in S \quad &\Rightarrow \quad v_j = e^{-i\gamma j/n}, \quad \gamma \in \mathbb{R}. \end{aligned}$$

This enables us to write equation 4.1 as

$$(\kappa - T(l) - T^*(m))\phi = v \in S$$

where

$$\begin{aligned}
\kappa &= 0, \\
l_j &= \frac{-i}{2}\int_{-\delta}^{\delta} H_0^{(1)}(k(2\delta j - t))\ dt + O(\delta^2), \quad j \in \{0, 1 \ldots n\}, \\
m_j &= \frac{i}{2}\int_{-\delta}^{\delta} \overline{H}_0^{(1)}(k(2\delta j - t))\ dt + O(\delta^2), \quad j \in \{1, 2 \ldots n\}, \quad m_0 = 0, \\
v &= e^{-ik\cos\alpha\delta n}\ \widetilde{f}_\alpha \quad \text{and} \\
\phi &= e^{-ik\cos\alpha\delta n}\ \widetilde{v}_\alpha.
\end{aligned}$$

The eigenvalue of v under P is $e^{i2k\cos\alpha\delta n}$. In the notation of the theory section, we suppose $w_0 \in S$ corresponds to incident angle β, the constant α_0 can then be evaluated, we find

$$\alpha_0 = 1 - e^{-i2\delta k(\cos\beta - \cos\alpha)}.$$

As in the continuous case, we take advantage of the symmetry and choose $w_1 = \widetilde{U}w_0$ corresponding to angle $\pi - \beta$. Substitution of these quantities into equation 2.6 reveals, after a little manipulation,

$$\begin{aligned}
2\sin(2\delta k\cos\beta)\widetilde{\Gamma}(\beta,\beta)\widetilde{v}_\alpha &= \widetilde{\Gamma}(\beta,\alpha)[\sin(2\delta k\cos\alpha) + \sin(2\delta k\cos\beta) + \\
&\quad i(\cos(2\delta k\cos\alpha) - \cos(2\delta k\cos\beta))(2T(v) - 1)]\widetilde{v}_\beta \\
&- \widetilde{\Gamma}(\beta,\pi-\alpha)[\sin(2\delta k\cos\alpha) - \sin(2\delta k\cos\beta) + \\
&\quad i(\cos(2\delta k\cos\alpha) - \cos(2\delta k\cos\beta))(2T(v) - 1)]\widetilde{U}\widetilde{v}_\beta,
\end{aligned}$$

where we have written $\widetilde{\Gamma}(\alpha,\beta)$ for $2\delta\langle\, \widetilde{v}_\alpha \,|\, \widetilde{f}_{\pi-\beta}\,\rangle$ so that $\widetilde{\Gamma} = \Gamma + O(\delta)$. The embedding formula for $\widetilde{\Gamma}$ can be arrived at by similar substitution into equation 2.7. We choose w_2 corresponding to angle $\pi - \theta$ and calculation reveals

$$\begin{aligned}
&2i\sin(2\delta k\cos\beta)(e^{i2\delta k\cos\theta} - e^{-i2\delta k\cos\alpha})\ \widetilde{\Gamma}(\beta,\beta)\ \widetilde{\Gamma}(\alpha,\theta) \quad = \\
&\quad (e^{-i2\delta k\cos\alpha} - e^{-i2\delta k\cos\beta})(e^{i2\delta k\cos\theta} - e^{i2\delta k\cos\beta})\ \widetilde{\Gamma}(\beta,\pi-\alpha)\ \widetilde{\Gamma}(\beta,\theta) \\
+ \quad &(e^{i2\delta k\cos\beta} - e^{-i2\delta k\cos\alpha})(e^{i2\delta k\cos\theta} - e^{-i2\delta k\cos\beta})\ \widetilde{\Gamma}(\beta,\pi-\theta)\ \widetilde{\Gamma}(\beta,\alpha).
\end{aligned}$$

It is not difficult to check that these discrete embedding formula agree with the continuous case to first order in δ.

4.3 Implementation

Calculation of the Approximating Matrix

We calculate the elements of $\widetilde{K}$ using the approximation that the kernel function is constant (provided it is continuous) over each discrete interval. The kernel function

is continuous except at 0 so that if $j \neq l$, $H_0^{(1)}(k|2\delta(j-l)-t|) = H_0^{(1)}(k|2\delta(j-l)|) + O(\delta)$ for $t \in [-\delta, \delta]$. Hence we can write

$$\frac{i}{2}\int_{-\delta}^{\delta} H_0^{(1)}(k|2\delta(j-l)-t|)\ dt = i\delta H_0^{(1)}(k|2\delta(j-l)|) + O(\delta^2), \quad j \neq l.$$

We can approximate the integral when $j = l$ (where the integration domain straddles the logarithmic singularity) using the expansion of $H_0^{(1)}$ for small arguments e.g. from [10], $x \sim 0 \Rightarrow H_0^{(1)} = \frac{2i}{\pi}(\ln(\frac{x}{2}) + \gamma) + 1 + O(x^2)$ from which it follows that

$$\frac{i}{2}\int_{-\delta}^{\delta} H_0^{(1)}(k|t|)\ dt = 2\delta\left(\frac{i}{2} - \frac{1}{\pi}(\ln(\delta k/4) - 1 + \gamma)\right) + O(\delta^2).$$

Singularities in the Solution

As we have mentioned, the singularities in v_α at the endpoints are formally avoided because the discrete domain is strictly inside $[-1, 1]$. Computationally we can monitor $\widetilde{v}_{\alpha 0}$ and $\widetilde{v}_{\alpha n}$ to ensure that rounding errors do not begin to dominate our calculation as we increase n.

A more satisfactory way of eliminating this problem is available to us by a simple change of variables. This hinges on the fact established in chapter 3 that $u_\alpha(x) = \sqrt{1-x^2}\, v_\alpha(x)$ is a continuous function. For brevity we write the kernel as $s(x-t)$ and see that

$$\begin{aligned}\int_{-1}^{1} s(x-t)v_\alpha(t)\ dt &= \int_{-1}^{1} \frac{s(x-t)u_\alpha(t)}{\sqrt{1-t^2}}\ dt \\ &= \int_0^{\pi} s(x-\cos\theta)u_\alpha(\cos\theta)d\theta.\end{aligned}$$

Hence the singularity is removed by the $t = \cos\theta$ transform. It follows that the collocation method for equal θ intervals is the preferred method of solution for this equation. However, such a change of variables is non-linear so has no discrete analogue, arbitrary functional values at equal θ intervals not being any linear combination of those at equal x intervals. As a result the discrete embedding results cannot be applied to this system.

In the following we will refer to the equal x and θ discretisations as methods 0 and 1. Both are implemented and analysed in order to make a comparison.

We note that making the $\cos\theta$ transform but remaining with equal x intervals is equivalent to considering the function u_α and kernel $s(x-t)/\sqrt{1-x^2}$ for which the embedding results do continue to hold. Practically though this has no advantage over the equation for v_α, the change being only cosmetic.

Rounding Errors

There is some hazard due to rounding errors when we compute the constants in the discrete embedding formula. Consider evaluation of

$$\cos(2\delta k \cos\alpha) - \cos(2\delta k \cos\beta).$$

The result is $O(\delta^2)$ but is the difference between terms of $O(1)$. In this case the rounding error can be avoided by calculation in the alternative guise

$$-2\sin^2(\delta k \cos\alpha) + 2\sin^2(\delta k \cos\beta).$$

Similarly we note that

$$\begin{aligned} e^{-i2\delta k\cos\alpha} - e^{-i2\delta k\cos\beta} = &-2\sin^2(\delta k\cos\alpha) + 2\sin^2(\delta k\cos\beta) \\ &-i\sin(2\delta k\cos\alpha) + i\sin(2\delta k\cos\beta). \end{aligned}$$

4.4 Numerical Error Analysis

Given our ignorance of the true solution, the nature of the error in the computed solution is investigated numerically. We split our analysis into near-field and far-field parts. This is justified by the fact that the integral representation of the field is dominated by a logarithmic singularity in the near-field and tends to a plane wave as we move toward the far-field. Considering these two extremes separately we aim to be comprehensive without attempting an analysis of the complete field.

Error in the Near-Field

We choose to measure the near-field error by considering the RMS error on I where the solution is known. We can compute the approximate field at any point on I by virtue of the integral representation, and then comparison with the exact field enables us to compute the error. If we write $\hat{v}_\alpha$ for the piecewise constant function naturally extended from $\widetilde{v}_\alpha$ then our error measure can be written

$$E(k,\alpha,n) = ||K\hat{v}_\alpha - f_\alpha|| = ||\widetilde{\phi}_g - \phi_g||,$$

the norm being that on $L_2([0,1])$. In order to estimate E numerically we discretise I as usual, but now with m discrete intervals. A solid estimate can then be made by increasing m until the evaluations stop changing significantly.

As we expect, both methods exhibit first-order behaviour in the near-field, i.e.

$$E_i(k,\alpha,n) = a_i(k,\alpha)n^{-1} + O(n^{-2}), \quad i = 0,1, \quad n \gtrsim k.$$

Figure 4.1 illustrates such dependence by the example of $E_0(16, 0.3\pi, n)$ which is typical.

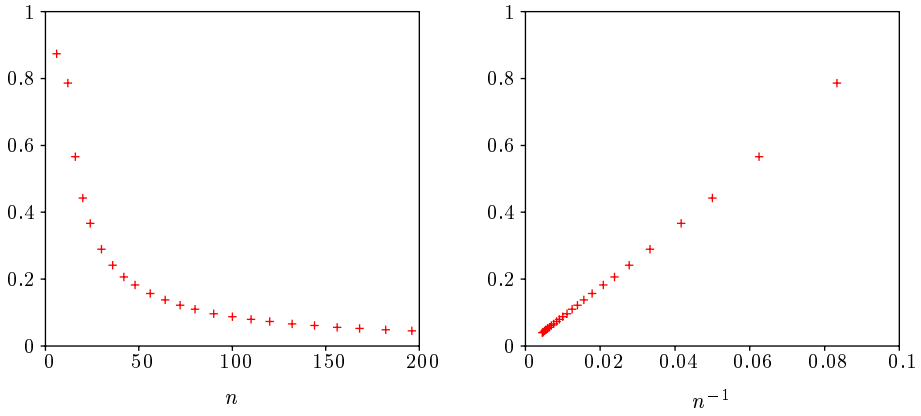

Figure 4.1: n dependence of the near-field RMS error, E_0, $k = 16$, $\alpha = 0.3\pi$.

The functional form of E_i enables us to compare the near-field error between the methods by comparing the error constant $a_i(k, \alpha)$. We estimate $a_i(k, \alpha)$ by the method of least-squares fitting. Figure 4.2 shows $a_i(k, \alpha)$ for $k \in (0, 100)$ and $\alpha = 0.3\pi$.

For small k we see that method 1 gives smaller errors than method 0. As we increase k the error constant increases approximately linearly, the gradient being

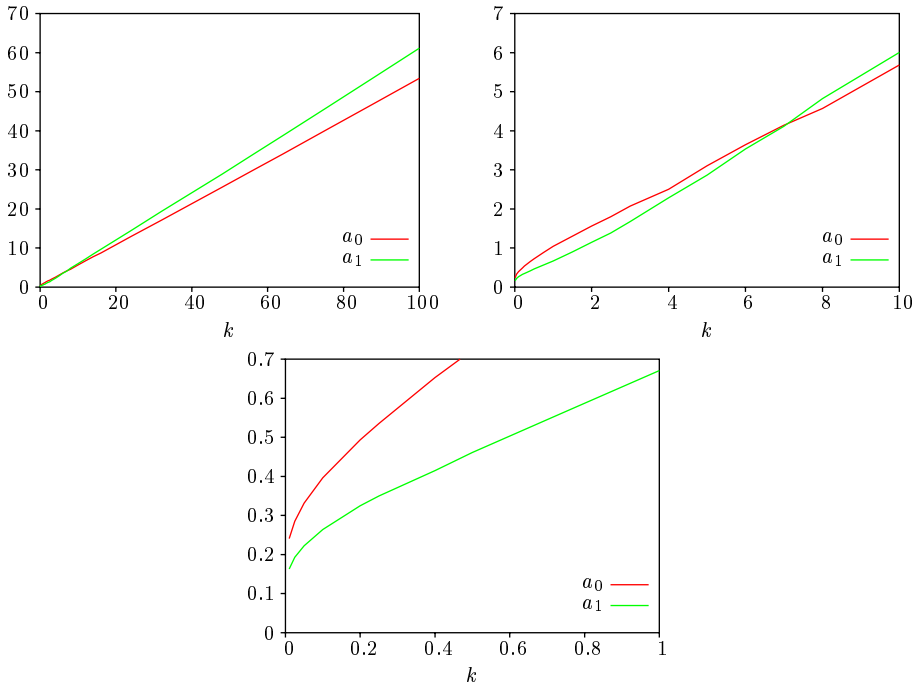

Figure 4.2: k dependence of the near-field error constant, $\alpha = 0.3\pi$.

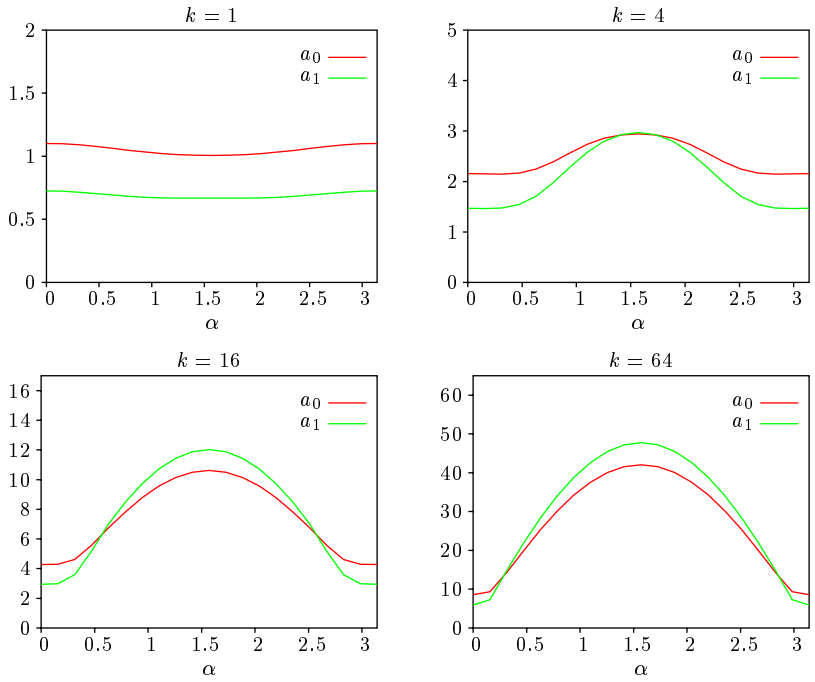

Figure 4.3: α dependence of the near-field error constant.

dependent on α. The α dependence is illustrated by Figure 4.3, for small k we see that $a_i(k, \alpha)$ is approximately constant for $\alpha \in [0, \pi]$. Method 0 exhibits less dependency on α and as a result gives smaller errors than method 1 for large k.

Error in the Far-Field

The diffracted far-field is given, in the limit, by Γ. In this section we investigate the error in Γ to complete our comparative study of the numerical methods. There is no analytic solution here, so in order to investigate the behaviour of the error we compute the solution for large n, say N, and treat it as if it were exact. Here we use N of about 5 times the largest n considered; using larger N does not effect the results presented in any significant way.

Given a particular α, Γ is defined for observation angle between 0 and π and we consider the RMS error in Γ on this interval. In terms of the $L_2([0, \pi])$ norm we can write this as

$$E(k, \alpha, n) = \frac{||\Gamma - \widetilde{\Gamma}||}{\sqrt{2\pi k}}.$$

E can be estimated numerically in the same way as for the near-field, the primary difference here is that we do not know Γ.

The first order behaviour is again very clearly exhibited by method 0. However, although method 1 is first order asymptotically for large n, the dependence is not

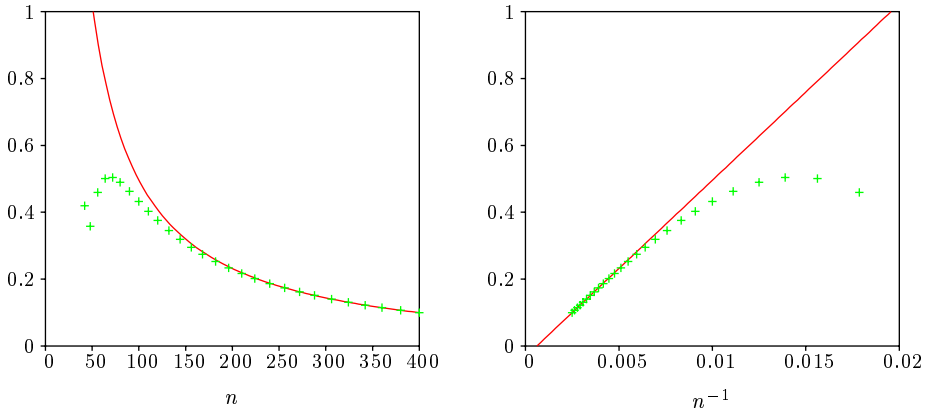

Figure 4.4: n dependence of the far-field RMS error, E_1, $k = 32$, $\alpha = 0.3\pi$.

so clear as can be seen from Figure 4.4 which is typical. This deviation from first order error grows with k. We can still compare the error constants between the two methods,

$$E_i(k, \alpha, n) = b_i(k, \alpha)n^{-1} + O(n^{-2}), \quad i = 0, 1, \quad n \gtrsim k.$$

But now for $k \gtrsim 1$ and moderate n we note that E_1 produces an overestimate of

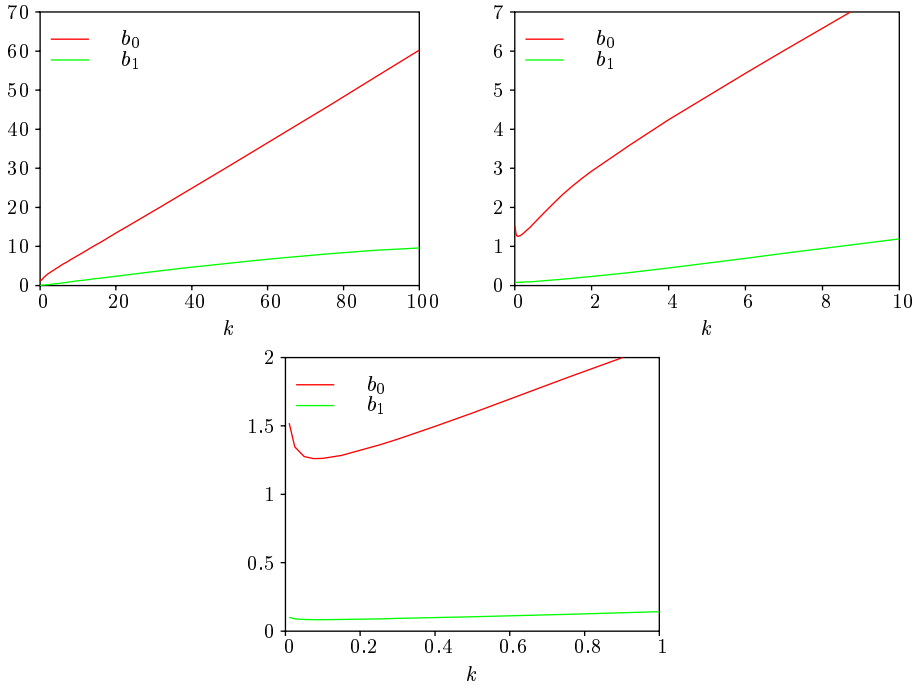

Figure 4.5: k dependence of the far-field error constant, $\alpha = 0.3\pi$.

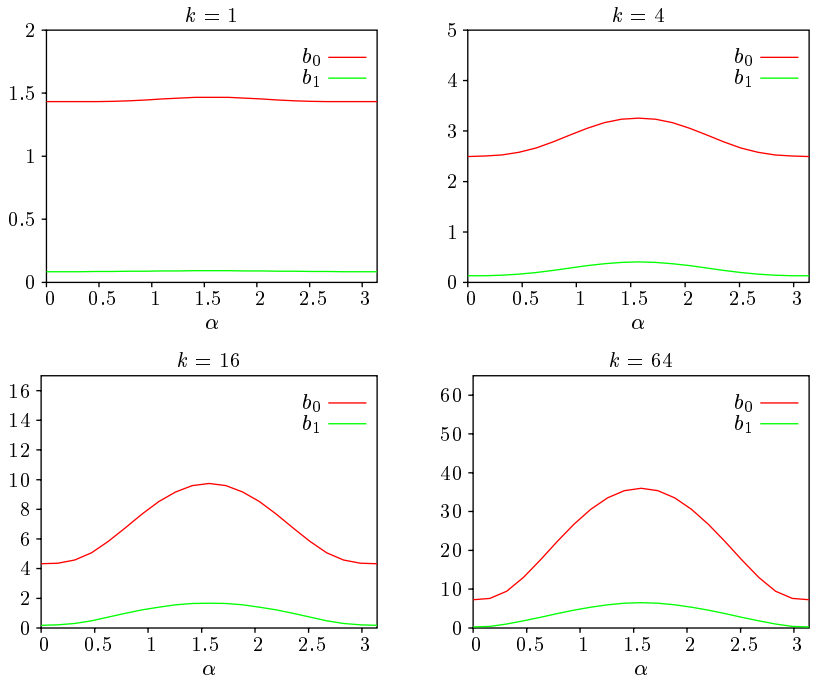

Figure 4.6: α dependence of the far-field error constant.

the error.

The general trends in the far-field error are similar to the near-field which is clear from Figures 4.5 and 4.6. The most immediate difference is that b_1 is about an order of magnitude smaller than a_1. This tells us that method 1 produces significantly more accurate results in the far field.

Approximate Embedding

Here we consider how embedding effects the accuracy of the results obtained. Method 0 is of course unaffected by the use of the discrete embedding formula. However, we expect that discretising the continuous embedding formula, which is required for method 1, will introduce additional error in the numerical solution. To this end we consider method 2 which uses the solution calculated using method 1 at $\alpha = 0$ together with the approximate embedding formula to obtain the solution for other α.

With $k \lesssim 1$ the approximate embedding formula makes no significant contribution to the near-field or far-field error, we find method 2 is first order with error constant almost exactly that of method 1. In the near-field we find that $k \gtrsim 1$ implies, again, the same first-order error behaviour as method 1, there is however noticeable growth in the second-order error term as the example in Figure 4.7

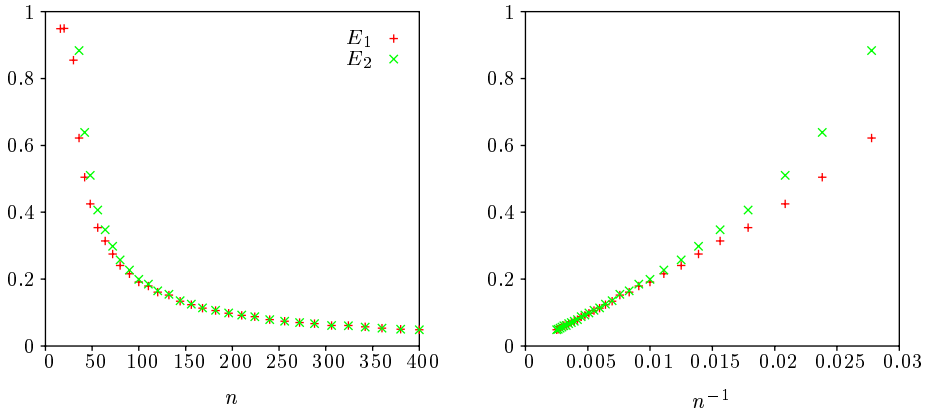

Figure 4.7: n dependence of the near-field RMS error, $k = 32$, $\alpha = 0.3\pi$.

illustrates.

As we increase k past 1 the far-field error changes more dramatically. Method 2 does not exhibit the deviation from first order found in method 1 (which we illustrated in Figure 4.4). Instead the first-order error term diminishes with k and, as in the near-field, the second-order term increases, and in fact here it starts to dominate. This behaviour is illustrated at $k = 32$ in Figure 4.8 which is typical, we have plotted E_1 and E_2 so that comparison can be drawn.

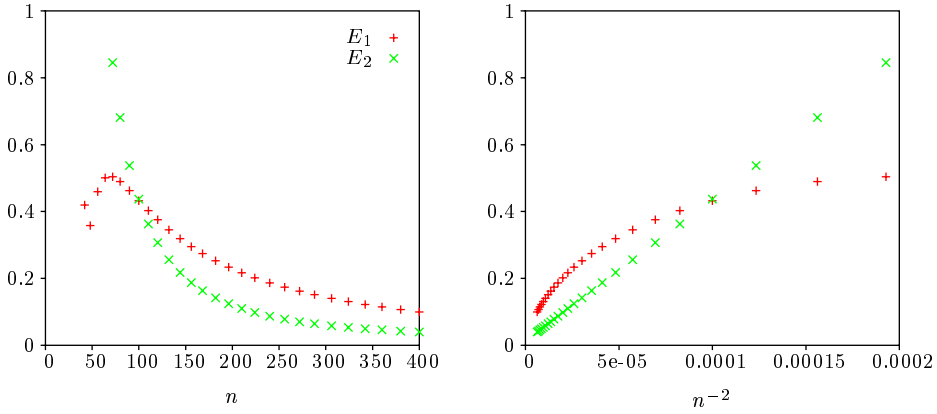

Figure 4.8: n dependence of the far-field RMS error, $k = 32$, $\alpha = 0.3\pi$.

The main point we take from these results is that the approximate embedding formula complicate the far-field error behaviour when $k \gtrsim 1$, but the scale of impact is quite small when compared with the differences between methods 0 and 1 in the far-field. For $k \lesssim 1$ and in the near-field the impact of approximate embedding is very small.

4.5 Remarks

To summarise, we have discretised the integral equation for v_α in order to approximate it numerically. This has been done using a basic collocation method, but in two ways; one which draws on results of the theory section and preserves embedding, and one which is more satisfactory because it naturally removes the singularities from the computations, but which relies on approximate embedding. When embedding is used in the computations, the two methods are indistinguishable in terms of computational cost, and error analysis shows that the second discretisation produces less error in its solution even though the embedding formula is approximate.

The use of Green's method to reduce an elliptic BVP leads naturally to advantages in computation. Apart from a decrement in the dimension of the linear system, the method provides an integral representation of the solution. Such a representation can be used to reconstruct the solution at any point on its domain, this being done in a natural way rather than by interpolation through nodal values. It is this which enables us to analyse the error in these problems because the field reconstructed from a solution computed with a fine mesh can be successfully compared with one generated on a coarse mesh.

4.6 Diffraction Patterns

In this section we exhibit some numerical solutions of the diffraction problem. They illustrate a range of phenomena which are well described by our quasi steady-state near-field analysis of the wave equation in 2D.

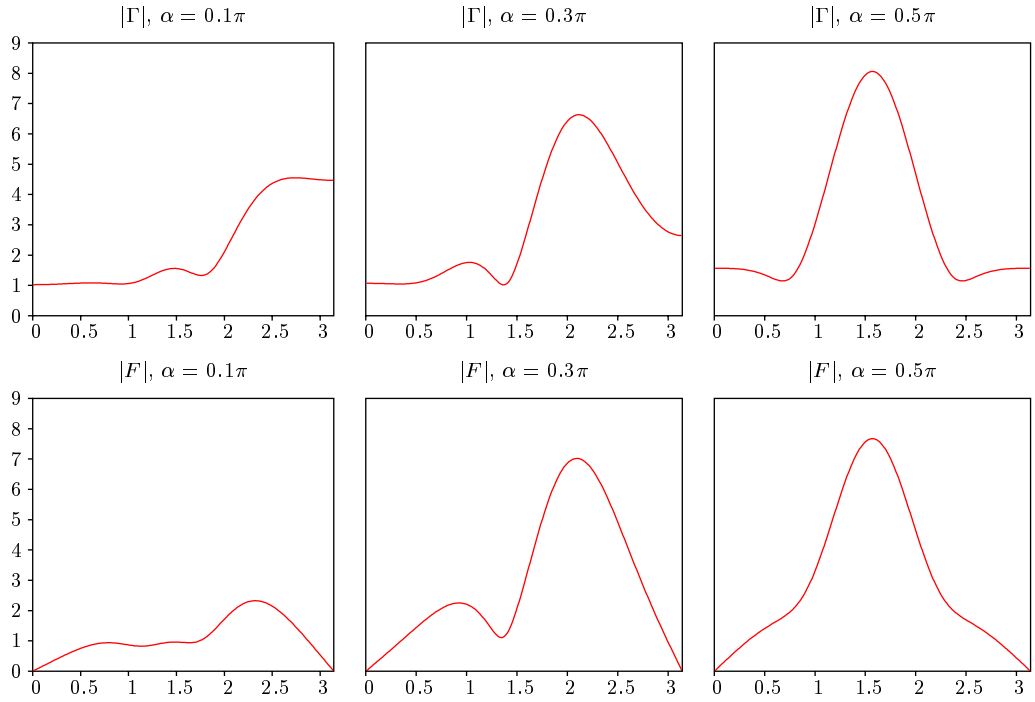

Figure 4.9: far-field diffraction pattern, $I = (-1, 1)$, $k = 4$

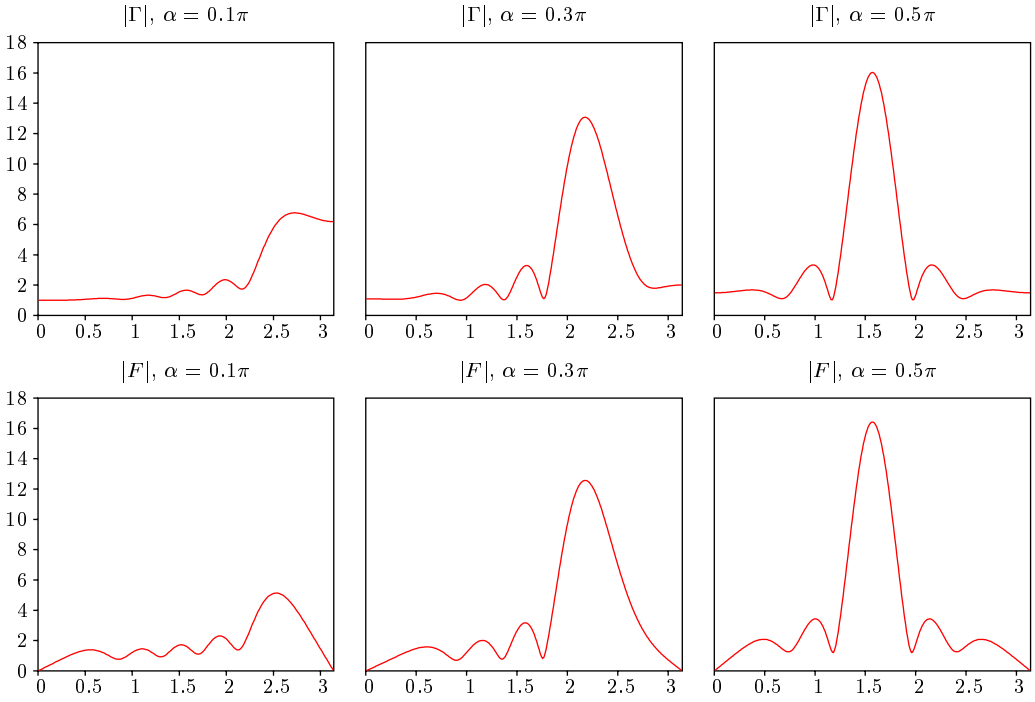

Figure 4.10: far-field diffraction pattern, $I = (-1, 1)$, $k = 8$

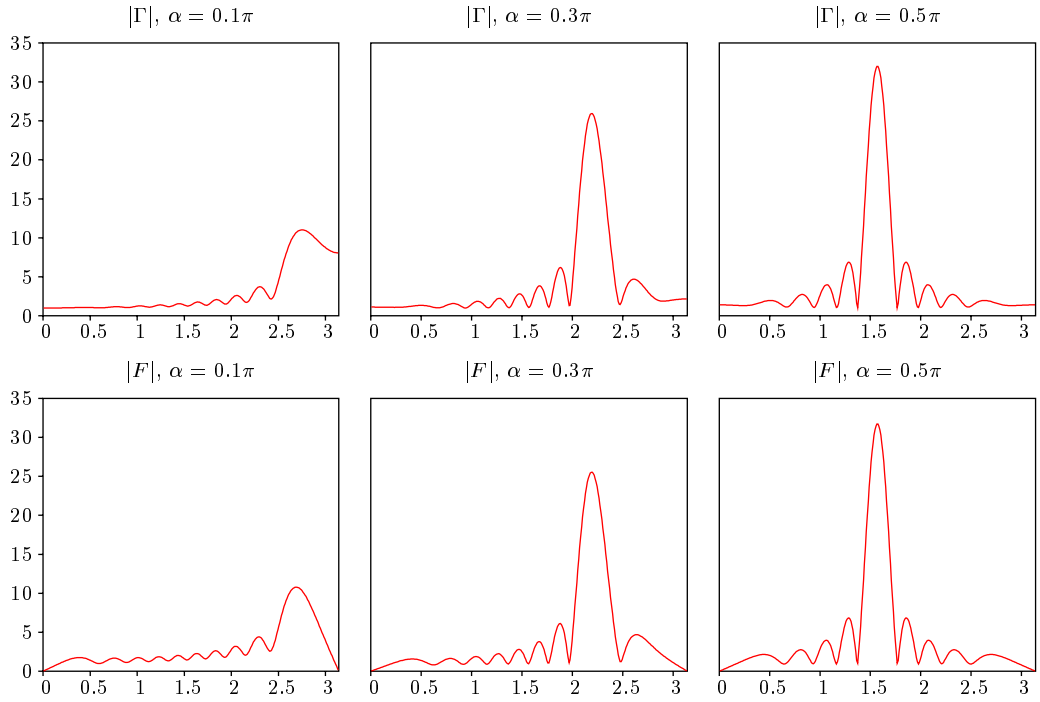

Figure 4.11: far-field diffraction pattern, $I = (-1, 1)$, $k = 16$

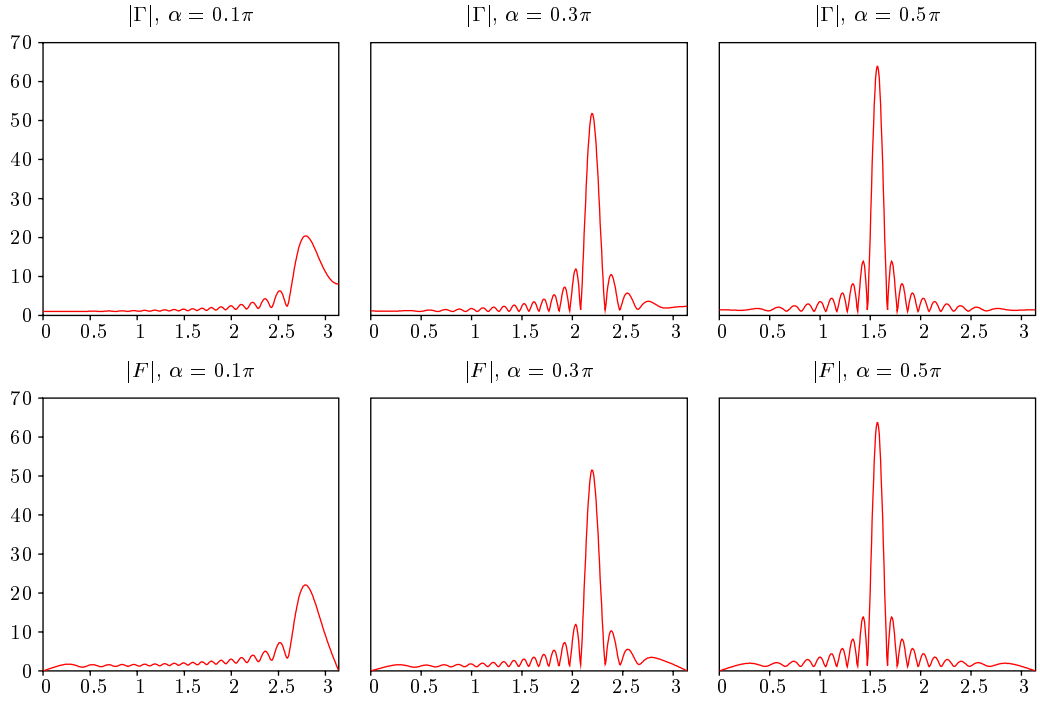

Figure 4.12: far-field diffraction pattern, $I = (-1, 1)$, $k = 32$

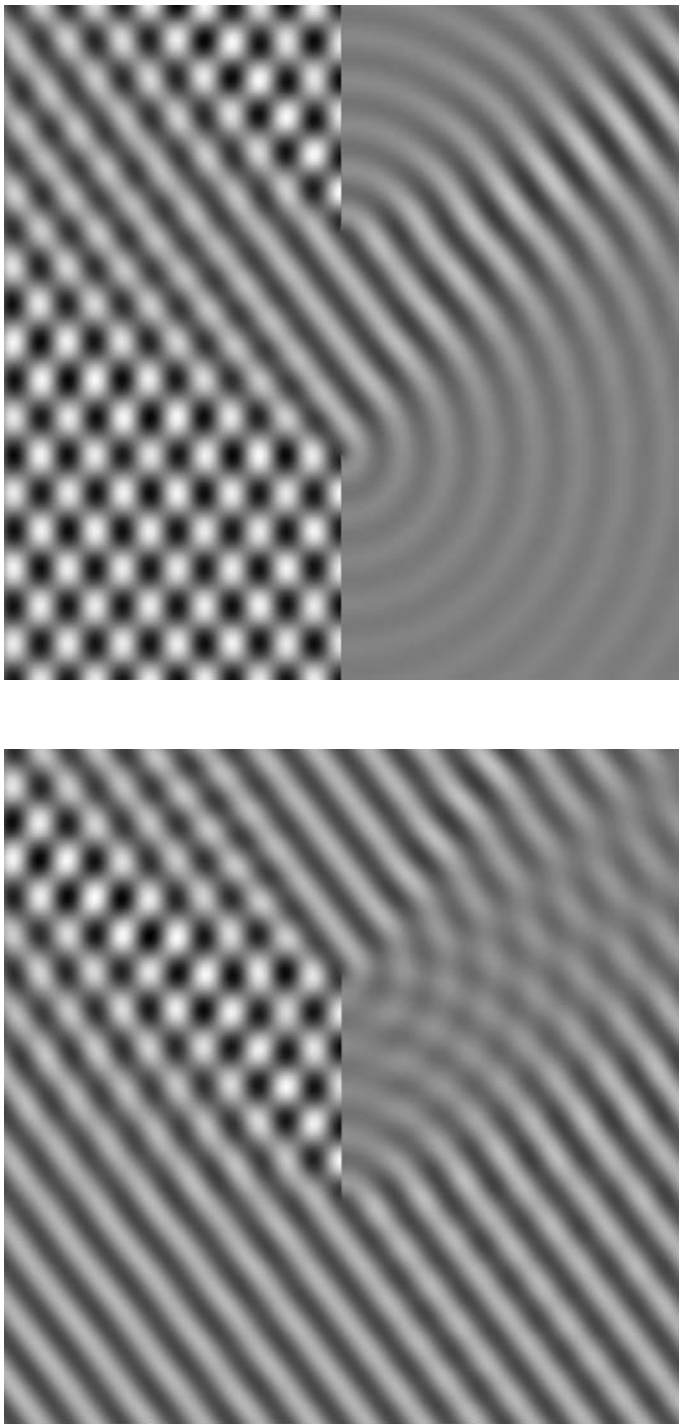

Figure 4.13: $\mathrm{Re}(\phi)$ on $[-3,3] \times [-3,3]$, $k = 16$, $\alpha = 0.7\pi$.

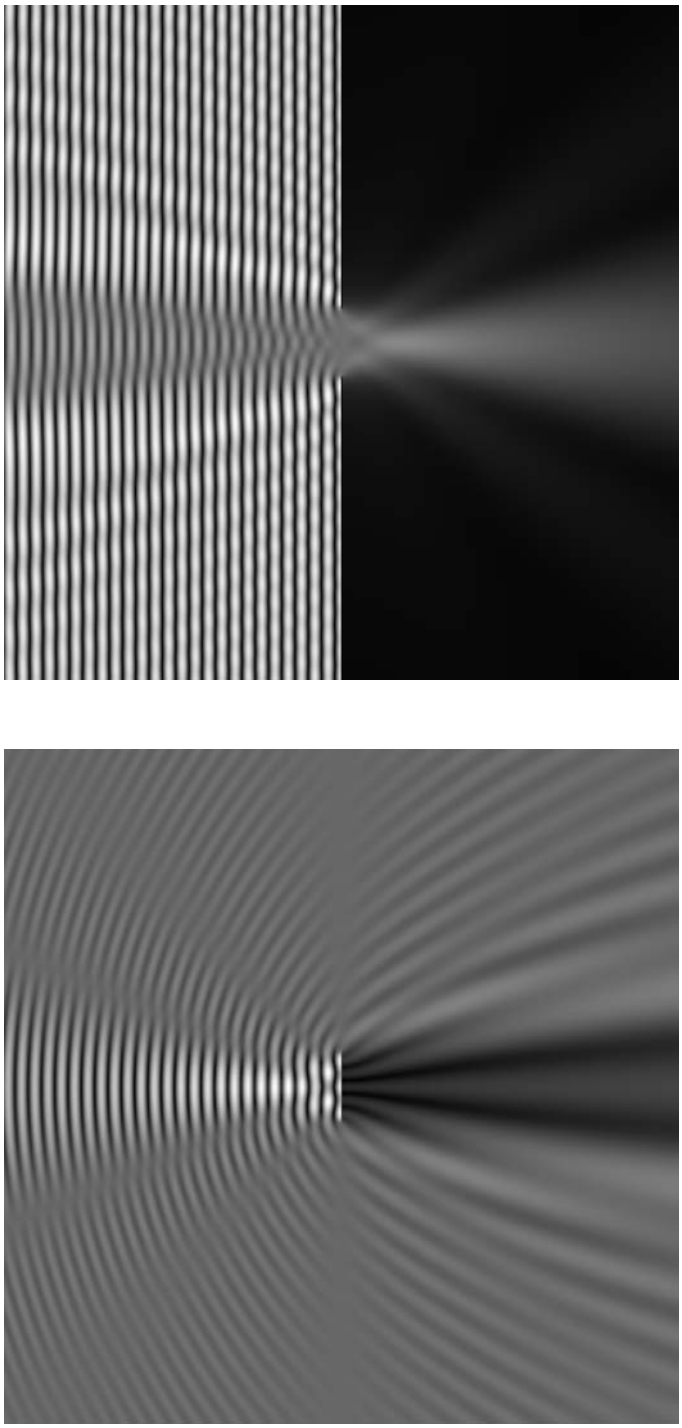

Figure 4.14: $|\phi|$ on $[-10, 10] \times [-10, 10]$, $k = 8$, $\alpha = 0.5\pi$.

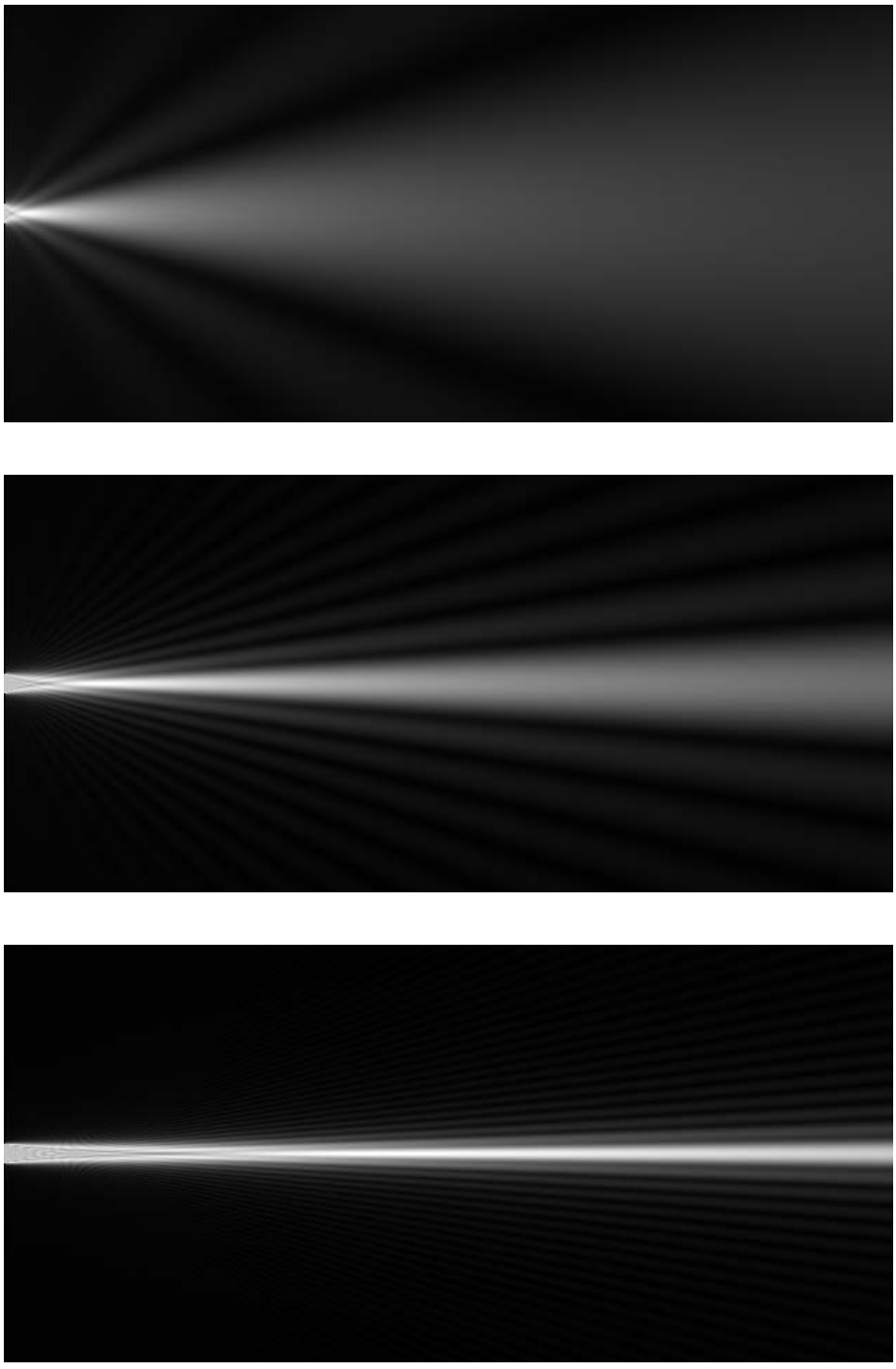

Figure 4.15: $|\phi_g|$ on $[-20, 20] \times [-43, 0]$, $k = 10, 40, 160$, $\alpha = 0.5\pi$.

Chapter 5

Conclusions

5.1 Overview

In this report we have derived an embedding result in an abstract setting which coincides with a result of Porter [8] in the continuous case and provides an analogous result for Toeplitz matrices in the discrete case. The application of this to the numerical solution of a simple diffraction problem illustrates how structure is preserved by a basic collocation discretisation of the governing integral equation. Our investigation into this numerical method has shown that, for this application, another discretisation is preferable in terms of accuracy, even though it relies on an approximate embedding formula.

5.2 Numerical Methods

We labour some important asides which have not been the focus of this work.

- The embedding formula lead to computational savings. Even with a fast matrix solver, the embedding formula (approximate or exact) used to generate new solutions is quicker compared to solving a new equation. The benefit is especially felt in the far-field calculations, the embedding formula require evaluation of a few inner-products compared with solving a linear system if they are not used.

- Green's method is a very useful front-end for numerical methods. Additionally to reducing the dimension of the problem, it enables computation, in a natural way, of the approximate solution at any point in its domain. The benefit of this has been highlighted here because it enables reliable numerical error analysis.

5.3 Further Work

This study has led to the identification of a few areas where further investigation may be useful.

- The core of the project was the theoretical work which stands alone now the intended application has been shown to be un-beneficial. As a general framework this is currently unsatisfactory because there exists a result for the principle examples (1 and 2 of the theory section) which it does not yet describe. Specifically, if $A = \kappa - T(l) - T^*(m)$ is invertible and $\kappa \neq 0$ then the inverse is *finitely generated* by, i.e. expressible in terms of, ψ_0 and ψ_1 which satisfy

  $$A\psi_0 = l \quad \text{and} \quad A^*\psi_1 = m.$$

  This is analogous to example 3 of section 2.2 which applies to the simpler convolution type operator. In the continuous case this result is attributable to Gokhberg and Feldman [2] (mentioned already in section 1.2), in fact they discover it by analogy with its discrete counterpart. Porter [8] shows, again in the continuous case, that this result can be derived from the main embedding formula we give in this report, which suggests the results are intimately connected.

  Unifying the discrete and continuous versions of this result thus seems a distinct possibility, and can only lead to a deeper insight into the nature of these embedding results and hence their extensions and applications.

- It may be that there exist other methods of discretisation which, when applied to integral operators generated by a difference kernel, lead to a Toeplitz matrix equation. If such a method also exhibited desirable error characteristics the discrete embedding formula could find application. If one considers implementing Galerkin's method using a finite set, say $\{w_i\}$, to approximate the solution of $A\phi = f$ where $A = \kappa - T(l) - T^*(m)$ in the notation of the theory section, one quickly arrives at the discrete equation

  $$\sum_{i=0}^{n} \left[\kappa \langle\, w_i \,|\, w_j \,\rangle + \langle\, l \,|\, T^*(w_i)w_j \,\rangle + \langle\, T^*(w_j)w_i \,|\, m \,\rangle\right] \phi_i = \langle\, f \,|\, w_j \,\rangle,$$

  $$\text{where} \quad j \in \{0, 1 \ldots n\} \quad \text{and} \quad \widetilde{\phi} = \sum_{i=0}^{n} \phi_i w_i.$$

  Which places quite specific demands on $\{w_i\}$ so that a search can be conducted in a systematic way.

- The numerical error analysis showed some interesting results for the approximate embedding formula, in some circumstances improving the accuracy

over the non-embedded calculation. Such results may be worth investigating analytically because once understood they could lead to numerical methods with improved error performance for such equations.

Appendix A

Conjugate Linear Maps

A.1 Introduction

We list here some elementary properties of a conjugate linear map defined on a unitary space. By conjugate linear we mean that for all $x, y \in V$ and $\lambda, \mu \in \mathbb{C}$

$$A(\lambda x + \mu y) = \overline{\lambda} Ax + \overline{\mu} Ay.$$

A.2 The Adjoint

We define the adjoint of A, which we write as A^*, so that

$$\langle\, Ax \,|\, y \,\rangle = \langle\, A^* y \,|\, x \,\rangle$$

for all $x, y \in V$. It is easy to show that A^* is also conjugate linear. If T is a linear transform on V then AT and TA are conjugate linear and $(AT)^* = T^* A^*$ because

$$\begin{aligned} \langle\, ATx \,|\, y \,\rangle &= \langle\, A^* y \,|\, Tx \,\rangle \\ &= \langle\, T^* A^* y \,|\, x \,\rangle. \end{aligned}$$

If T is conjugate linear, AT and TA are linear and $(AT)^* = T^* A^*$ because

$$\begin{aligned} \langle\, ATx \,|\, y \,\rangle &= \langle\, A^* y \,|\, Tx \,\rangle \\ &= \langle\, x \,|\, T^* A^* y \,\rangle. \end{aligned}$$

A.3 Unitary Conjugate Linear Maps

If P is conjugate linear and $P^*P = I$ we refer to P as unitary. If P has an eigenvector v with eigenvalue λ then

$$\langle\, v \,|\, v \,\rangle = \langle\, P^* P v \,|\, v \,\rangle = \langle\, Pv \,|\, Pv \,\rangle = \lambda \overline{\lambda} \langle\, v \,|\, v \,\rangle.$$

i.e. $|\lambda| \neq 1 \Rightarrow v = 0$.

If, then, $v \neq 0$, $Pv = \lambda v \Rightarrow P^*Pv = \overline{\lambda}P^*v \Rightarrow v = \overline{\lambda}P^*v \Rightarrow \lambda v = P^*v$. So that eigenvectors of P are eigenvectors of P^* with the same eigenvalue. A consequence of this is that if v, w are eigenvectors of P with eigenvalues λ, μ, then

$$\lambda\langle\, v \,|\, w \,\rangle = \langle\, Pv \,|\, w \,\rangle = \langle\, P^*w \,|\, v \,\rangle = \mu\langle\, w \,|\, v \,\rangle.$$

Bibliography

[1] **J.H.Carr and M.E.Stelzriede**, *Diffraction of water waves by breakwaters* US Nat. Bur. Stds. 521, 109-125, 1952.

[2] **I.Gohberg and I.Feldman** (translation from Russian by F.Goldware), *Convolution equations and projection methods for their solution*, Translations of mathematical monographs vol 41, American Mathematical Society, Providence, RI, 1974 (Russia 1971).

[3] **A.L.Sakhnovich**, *On a Method of Inverting Toeplitz Matrices*, Math. Issled., 8:4, 180-186 (Russian), 1973.

[4] **Y.N.Lebedev and I.M.Polishchuk**, *On Some Porperties of the Problem of Diffraction on a Strip*, Radio Engr. Electron Phys., 22:12, 27-30 (Russian), 1977.

[5] **D.Porter**, *Wave Diffraction around Breakwaters*, Technical Report, The Hydraulics Research Station - Wallingford, No IT 200, December 1979.

[6] **G.Gilbert and A.H.Brampton**, *The solution of two wave diffraction problems using integral equations*, Technical report, Hydraulics Research Station - Wallingford, December 1980.

[7] **D.Porter and K.Chu**, *The Solution of two Wave-Diffraction Problems*, Journal of Engineering Mathematics, Vol 20, 63-72, 1986.

[8] **D.Porter**, *The Solution of Integral Equations with Difference Kernels*, Journal of Integral Equations and Applications, Vol 3, No 3, Summer 1991.

[9] **L.A.Sakhnovich**, *Integral Equations with Difference Kernels on Finite Intervals*, Operator Theory Advances and Applications, Birkhauser, Vol 84, 1996.

[10] **M.Spiegel and J.Liu**, *Mathematical Handbook of Formulas and Tables - Second Edition*, Schaum's Outline Series, McGraw-Hill, 1999.

[11] **N.R.T.Biggs, D.Porter and D.S.G.Stirling** , *Wave Diffraction through a Perforated Breakwater*, Q.Jl Mech. appl. Math. 53(3), 375-391, 2000.

[12] **N.R.T.Biggs and D.Porter** , *Wave Diffraction through a Perforated Barrier of Non-Zero Thickness*, Q.Jl Mech. appl. Math. 54(4), 523-547, 2001.

[13] **N.R.T.Biggs and D.Porter** , *Wave Scattering by a Perforated Duct*, Q.Jl Mech. appl. Math. 55(2), 249-272, 2002.