Sparsified discrete wave problem involving radiation condition on a prolate spheroidal surface

Hélène Barucq, M'Barek Fares, Carola Kruse, Sébastien Tordeux

IMA Journal of Numerical Analysis (2019) Page 1 of 29
doi:10.1093/imanum/drn000

Sparsified discrete wave problem involving radiation condition on a prolate spheroidal surface

HÉLÈNE BARUCQ †
INRIA Bordeaux Sud-Ouest, 200 Avenue de la Vieille Tour, 33405 Talence, France
M'BAREK FARES ‡, CAROLA KRUSE §
CERFACS, 42 Avenue Gaspard Coriolis, 31100 Toulouse, France
SÉBASTIEN TORDEUX ¶
CNRS, Université de Pau et des Pays de l'Adour, Avenue de l'Universit, 64012 Pau, France

[Received on 6 September 2019]

We develop and analyze a high-order outgoing radiation boundary condition for solving three-dimensional scattering problems by elongated obstacles. This Dirichlet to Neumann (DtN) condition is constructed using the classical method of separation of variables that allows to define the scattered field in a truncated domain. It reads as an infinite series which is truncated for numerical purposes. The radiation condition is implemented in a finite element framework represented by a large dense matrix. Fortunately, the dense matrix can be decomposed into a full block matrix which involves the degrees of freedom on the exterior boundary and a sparse finite element matrix. The inversion of the full block is avoided by using a Sherman-Morrison algorithm which reduces the memory usage drastically. Despite being of high-order, this method has only a low memory cost.

Keywords:

Absorbing boundary condition, prolate spheroidal coordinates, Dirichlet-to-Neumann operator, scattering problems.

Many applications such as sonar, radar or geophysical exploration involve exterior Helmholtz problems which raise the question of handling the unbounded propagation domain. One accurate approach consists in using boundary element methods (BEM), see Bendali & Fares (2008) and their references. The linear partial differential equation is reformulated as an integral equation which is set on the boundary of the scatterer. BEMs have thus the advantage of providing an exact reformulation of the exterior problem in a bounded region without introducing any kind of approximation before discretization. However, boundary element discretizations lead to dense matrices which increase both storage requirements and computational costs of the solution algorithm, especially in three dimensions. A popular alternative is the usage of volume finite element methods (FEM) Harari & Hughes (1992b) once the computational domain has been truncated. The truncation is performed by introducing an artificial boundary surrounding the scatterer which requires transforming the exterior problem into a mixed one corresponding to the initial volume formulation of the problem set in the bounded domain and coupled with a suitable boundary condition set on the external boundary. To get an accurate representation of the wave field into the bounded domain, the external boundary should not generate any reflection. This point is managed with

†Email: helene.barucq@inria.fr
‡Email: fares@cerfacs.fr
§Corresponding author. Email: carola.kruse@cerfacs.fr
¶Email: sebastien.tordeux@inria.fr

the boundary condition which is supposed to govern waves outgoing to the artificial boundary. The literature on this subject is abundant and it turns out difficult to combine accuracy with low computational costs.

Regarding three dimensional scattering problems, the artificial boundary is often a sphere, independently of the geometry of the scatterer. Besides being simple to introduce, the sphere is a surface for which absorbing boundary conditions are quite easy to construct. A widely used boundary condition is the BGT one developed by Bayliss, Gunzburger and Turkel (see Bayliss *et al.* (1982) and Antoine *et al.* (1999) for an extension to arbitrarily shaped surfaces). Artificial boundary conditions can also be obtained by approximating the Dirichlet to Neumann (DtN) operator Harari & Hughes (1992a); Antoine *et al.* (1999); Barucq *et al.* (2007, 2009b) and in the case of a sphere, the BGT and the DtN conditions are identical when considering a second-order formulation. A spherical artificial boundary might however not be optimal for certain types of objects like elongated scatterers. Indeed, in that case, the introduction of a sphere that surrounds scatterers of elongated shape (e.g. a submarine) would define an unnecessarily large computational domain. As a remedy, elliptical domains in 2D as well as spheroidal domains in 3D have been considered. In Reiner *et al.* (2006) following former works in Givoli & Keller (1990); Harari & Hughes (1992a); Grote (1995), local absorbing boundary conditions of BGT type have been developed for elliptical surfaces and it turns out that they perform better in the mid and high frequency domain than for low frequencies. In Barucq *et al.* (2007, 2009a), new local DtN boundary conditions have been proposed for elliptical and prolate spheroidal-shaped boundaries based on the decomposition of the scatterer field onto a prolate spheroidal function basis. The construction principle is based upon the decomposition of the scattered field and the condition depends on two parameters which are computed to ensure the radiation condition is exact for the two first modes of the ansatz. It is shown in Barucq *et al.* (2007, 2009a) that the corresponding DtN condition outperforms the BGT one in the low frequency regime, regardless of the slenderness of the boundary. It is worth noting that the DtN condition is easy to implement and to parallelize. A similar ABC was proposed in Zarmi & Turkel (2013). High-order ABC has been considered and numerically tested in Medvinsky *et al.* (2008).

In this paper, we revisit the truncation of the exterior domain by a prolate spheroid, following the ideas of Barucq *et al.* (2007, 2009a). We first develop an exact Dirichlet to Neumann (DtN) boundary operator using the method of separation of variables for the Helmholtz equations which is set in the unbounded outer domain of a spheroidal manifold. We provide a decomposition of the wavefield which is used to represent the DtN operator exactly by an infinite series. The obtained operator is non local and owns the interesting property of being decomposable into a compact and positive part. Based on this observation, we show that the continuous variational problem associated to the corresponding mixed problem has a unique solution.

When coupling the Helmholtz equation with the exact DtN condition that is imposed on the bounded computational domain limited by the prolate spheroidal boundary, we obtain a boundary value problem. To get an implementable version of the DtN operator, the exact DtN series is truncated to a finite sum with an upper bound parameter N representing the order of truncation and thus the approximation. We finish the study of the continuous problem by proving that the wave field u_N, obtained by solving the mixed problem in the bounded domain and with the approximate DtN of order N, converges to the solution u of the exact variational problem when N tends to infinity.

The discretization of the approximate mixed problem leads to solving a matrix which is the sum of a classical sparse finite element matrix and a dense block whose size is determined by the number of elements on the boundary. The size of this block increases when the boundary elements are refined and sparse solvers are no longer efficiently applicable. Fortunately, with the observation that the full matrix is low-rank, we can employ a Sherman-Morrison algorithm that transforms the problem of inverting the

full matrix into the inversion of a sparse matrix having a few number of additional entries.

Index of notations

- A rigid body $\Omega' \subset \mathbb{R}^3$ and its complement, the propagation domain $\Omega = \mathbb{R}^3 \setminus \Omega'$.
- The computational domain D with $\Omega' \subset \mathsf{D}$ and its boundary Γ.
- We denote respectively by $H^s(\Omega)$ and $H^s(\Gamma)$ the Sobolev space of integer and fractional order s of Ω and Γ.
- The shared boundary $\partial\Omega$ of Ω and Ω'.
- The normal unit vector $\mathbf{n}$ is exterior to the considered computational domain Ω.
- The wave vector $\mathbf{k} = (\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ with norm $k = \sqrt{\mathbf{k}_1^2 + \mathbf{k}_2^2 + \mathbf{k}_3^2}$.
- The velocity c and the pulsation of the time harmonic wave $\omega = \frac{k}{c}$.
- The incident plane wave $u_{\text{inc}} : \mathbb{R}^3 \longrightarrow \mathbb{C}$, $u_{\text{inc}}(\mathbf{x}) = \exp(i\mathbf{k}\cdot\mathbf{x})$.
- The set of regular compactly supported functions $\mathscr{D}(\mathbb{R}^3)$.
- The set of functions which are locally H^1 up to the boundary

$$H^1_{\text{loc}}(\Omega) \;=\; \Big\{ u : \mathbb{R}^3 \longrightarrow \mathbb{C}; \chi u \in H^1(\Omega) \quad \forall \chi \in \mathscr{D}(\mathbb{R}^3) \Big\}. \tag{0.1}$$

The considered problem

In the context of time-harmonic linear acoustic wave propagation, the scattering of an incident plane wave by an impenetrable bounded rigid body Ω' has been widely studied. The total pressure field p can be deduced from the phasor of the scattered pressure field $u : \Omega = \mathbb{R}^3 \setminus \Omega' \longrightarrow \mathbb{C}$

$$p(\mathbf{x},t) = \Re\Big(\big(u(\mathbf{x}) + u_{\text{inc}}(\mathbf{x})\big) \exp(-i\omega t) \Big) \quad \forall \mathbf{x} \in \Omega \text{ and } t \in \mathbb{R}. \tag{0.2}$$

The complex function u is governed by a second order problem composed of a Helmholtz equation equipped with Neumann boundary conditions of the form

$$\begin{cases} u \in H^1_{\text{loc}}(\Omega), \\ \Delta u(\mathbf{x}) + k^2 u(\mathbf{x}) = 0 \text{ on } \Omega, \\ \dfrac{\partial u}{\partial \mathbf{n}}(\mathbf{x}) \;=\; -\dfrac{\partial u_{\text{inc}}}{\partial \mathbf{n}}(\mathbf{x}) \text{ on } \partial\Omega, \\ u(\mathbf{x}) \text{ is outgoing.} \end{cases} \tag{0.3}$$

This problem is well posed under suitable hypothesis on the regularity of Ω (see Lecture 4 in Wilcox (1984)). To compute a numerical approximation with a finite element discretization, it is rather classical to restrict the computational domain to

$$\big\{ \mathbf{x} \in \Omega : \|\mathbf{x}\| < R \big\}$$

where R is chosen large enough so that $\Omega' \subset \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| < R\}$. The radiation condition is then taken into account by an artificial boundary condition posed on the exterior boundary

$$\Big\{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = R\Big\}.$$

This impedance boundary condition can either be an exact non local Dirichlet-to-Neumann map, see Keller & Givoli (1989), or an approximate local differential operator which can be derived by a microlocal analysis, see Engquist & Majda (1977, 1979).

When the rigid body is elongated, it is not optimal to truncate the rigid body with a sphere. The computational cost can actually be reduced by considering a prolate spheroidal computational domain:

$$\mathsf{D} \;=\; \Big\{\mathbf{x} \in \Omega : \frac{x^2+y^2}{b^2} + \frac{z^2}{a^2} \leqslant 1\Big\} \quad \text{with } b < a. \tag{0.4}$$

where a and b are chosen large enough to ensure that the obstacle Ω' is included in the prolate spheroidal domain with semi-major axis a and semi-minor axis b

$$\Omega' \;\subset\; \Big\{\mathbf{x} \in \mathbb{R}^3 : \frac{x^2+y^2}{b^2} + \frac{z^2}{a^2} \leqslant 1\Big\}$$

In Saint-Guirons (2008); Barucq *et al.* (2009a), the authors have derived a second order radiation condition suitable for

$$\Gamma \;=\; \Big\{\mathbf{x} \in \Omega : \frac{x^2+y^2}{b^2} + \frac{z^2}{a^2} = 1\Big\}.$$

In this paper, we propose and analyze an approximate radiation condition which combines the efficiency of a low order absorbing boundary condition and the accuracy of a high order condition.

The paper is organized as follows: Section 1 is devoted to preliminary materials. We introduce the Sobolev spaces with fractional order of the prolate spheroidal manifolds. In section 2, we propose to approximate the outgoing condition by the truncation of an exact Dirichlet-to-Neumann operator derived by separation of variables. In Section 3.2, an implementation of our method is proposed by resorting to a Sherman-Morrison algorithm in order to improve the accuracy of the numerical method.

1. Sobolev norm on a spheroidal manifold

Let us first parameterize the spheroidal manifold $\Gamma = \{\frac{x^2+y^2}{b^2} + \frac{z^2}{a^2} = 1\}$ by the two angular prolate spheroidal coordinates $\eta \in [-1,1]$ and $\varphi \in [0, 2\pi]$

$$\begin{cases} x(\eta,\varphi) = b\sqrt{1-\eta^2}\cos(\varphi) \\ y(\eta,\varphi) = b\sqrt{1-\eta^2}\sin(\varphi), \\ z(\eta,\varphi) = a\eta, \end{cases} \tag{1.1}$$

The angular spheroidal functions $Y_{m,n} : \Gamma \longrightarrow \mathbb{C}$ Abramowitz & Stegun (1964); Flammer (1957); Lebedev & Silverman (1972) can be defined as the eigenfunctions of the generalized eigenproblem

$$\begin{cases} \text{Find } Y \in H^1(\Gamma) \text{ and } \lambda \in \mathbb{R} \text{ such that for all } v \in H^1(\Gamma) \\ (Y,v)_{\Gamma,1} \;=\; \lambda\; (Y,v)_{\Gamma,0} \end{cases}$$

with the bilinear forms

$$\left\{\begin{array}{rcl}(u,v)_{\Gamma,0} &=& \displaystyle\int_{-1}^{1}\int_{0}^{2\pi} u(\eta,\varphi)\,\overline{v}(\eta,\varphi)\,d\varphi\,d\eta \\ (u,v)_{\Gamma,1} &=& \displaystyle\int_{-1}^{1}\int_{0}^{2\pi} \partial_\eta\, u(\eta,\varphi)\partial_\eta\,\overline{v}(\eta,\varphi)(1-\eta^2) \\ && \qquad +\partial_\varphi\, u(\eta,\varphi)\,\partial_\varphi\,\overline{v}(\eta,\varphi) \\ && \qquad\quad + c^2\,(\eta^2+1)\,u(\eta,\varphi)\,\overline{v}(\eta,\varphi)\,d\varphi\,d\eta\end{array}\right.$$

with $c>0$ a positive parameter which will be chosen later. Classically, these eigenfunctions are parametrized by two indices $(m,n)\in\mathbb{Z}\times\mathbb{N}$ with $-n\leqslant m\leqslant n$ and are given by

$$Y_{m,n}(\eta,\varphi)\;=\;\mathsf{E}_{m,n}(\eta)\,\exp(im\varphi)\quad\text{ and }\quad (Y_{m,n},Y_{m,n})_{\Gamma,0}=1$$

where $\mathsf{E}_{m,n}:[-1,1]\longrightarrow\mathbb{R}$ the eigenfunctions of the second order differential operator

$$-\frac{\partial}{\partial\eta}\Big((1-\eta^2)\frac{\partial \mathsf{E}_{m,n}}{\partial\eta}\Big)(\eta)+\Big(\frac{m^2}{1-\eta^2}+c^2(\eta^2+1)\Big)\,\mathsf{E}_{m,n}(\eta)=\lambda_{m,n}(c)\,\mathsf{E}_{m,n}(\eta). \tag{1.2}$$

The eigenvalues $\lambda_{m,n}(c)$ do not have a simple expression except for $c=0$ where $\lambda_{m,n}(0)=n(n+1)$. A comparison argument leads to the bound

$$n\,(n+1)+c^2\;\leqslant\;\lambda_{m,n}(c)\;\leqslant\;n(n+1)+2c^2$$

Moreover, the norms

$$\|u\|_{\Gamma,0}=\sqrt{(u,u)_{\Gamma,0}}\quad\text{ and }\|u\|_{\Gamma,1}=\sqrt{(u,u)_{\Gamma,1}}$$

are equivalent to the classical $L^2(\Gamma)$ and $H^1(\Gamma)$-norms

$$\left\{\begin{array}{rcl}\|u\|^2_{L^2(\Gamma)} &=& \displaystyle\int_\Gamma |u(\mathbf{x})|^2\,ds_{\mathbf{x}}=b^2\int_{-1}^{1}\int_{0}^{2\pi}\big|u(\eta,\varphi)\big|^2 g(\eta)\,d\varphi\,d\eta \\ \|u\|^2_{H^1(\Gamma)} &=& \displaystyle\int_\Gamma |\nabla_\Gamma u(\mathbf{x})|^2+|u(\mathbf{x})|^2\,ds_{\mathbf{x}} \\ &=& \|u\|^2_{L^2(\Gamma)}+\displaystyle\int_{-1}^{1}\int_{0}^{2\pi}\big|\partial_\eta\,u(\eta,\varphi)\big|^2\frac{(1-\eta^2)\,d\varphi\,d\eta}{g(\eta)} \\ && \qquad +\displaystyle\int_{-1}^{1}\int_{0}^{2\pi}\big|\partial_\varphi\,u(\eta,\varphi)\big|^2\,g(\eta)\,d\varphi\,d\eta\end{array}\right.$$

with $g(\eta)=\sqrt{\eta^2+\frac{a^2}{b^2}(1-\eta^2)}\in[1,\frac{a}{b}]$. Due to the spectral and interpolation theories (see Chap. 1, Remark 2.3 in Lions & Magenes (1968)), we have the Lemma

LEMMA 1.1 Let $s\in[0,1]$. For all $u\in H^s(\Gamma)$, we have

$$u(\mathbf{x})\;=\;\sum_{n=0}^{+\infty}\sum_{m=-n}^{n}u_{m,n}\,Y_{m,n}(\mathbf{x})\quad\text{ in }H^s(\Gamma) \tag{1.3}$$

with

$$u_{m,n} = \int_{-1}^{1}\int_{0}^{2\pi} u(\eta,\varphi)\,\overline{Y_{m,n}}(\eta,\varphi)\,d\varphi\,d\eta \;=\; \frac{1}{ab}\int_{\Gamma}\frac{u(\mathbf{x})\,\overline{Y_{m,n}}(\mathbf{x})\,ds_{\mathbf{x}}}{\sqrt{a^2/f^2-z^2/a^2}}. \tag{1.4}$$

Moreover, for all $s\in[0,1]$, $(\cdot,\cdot)_{H^s(\Gamma)}$ and $\|\cdot\|_{\Gamma,s}$ are respectively scalar product and norm on the Hilbert space $H^s(\Gamma)$

$$\|u\|^2_{\Gamma,s} \;=\; \sum_{n=0}^{+\infty}\sum_{m=-n}^{n}(\lambda_{m,n}(c))^s\,|u_{m,n}|^2. \tag{1.5}$$

2. Taking into account radiation condition with prolate spheroidal coordinates

2.1 *The Dirichlet-to-Neumann map*

The restriction of the scattered field u to D solves the following system of equations

$$\begin{cases} u\in H^1(\mathsf{D}),\\ \Delta u + k^2\,u \;=\; 0 \quad \text{in } \mathsf{D},\\ \dfrac{\partial u}{\partial \mathbf{n}}(\mathbf{x}) \;=\; -\dfrac{\partial u_{\rm inc}}{\partial \mathbf{n}}(\mathbf{x}) \quad \text{on } \partial\Omega. \end{cases} \tag{2.1}$$

This system should be supplemented by a boundary condition posed on Γ in order to define a well-posed problem

$$\frac{\partial u}{\partial \mathbf{n}} + \mathsf{DtN}_k u \;=\; 0 \quad \text{on } \Gamma. \tag{2.2}$$

This radiation condition is derived by exploiting the properties of u in the exterior domain

$$\mathsf{D}' \;=\; \Big\{\mathbf{x}\in\mathbb{R}^3 : \frac{x^2+y^2}{b^2}+\frac{z^2}{a^2}\geqslant 1\Big\}. \tag{2.3}$$

In D', the scattered field u is an outgoing solution of the Helmholtz equation

$$\begin{cases} u\in H^1_{\rm loc}(\mathsf{D}'),\\ \Delta u + k^2\,u \;=\; 0 \quad \text{in } \mathsf{D}',\\ u \text{ is outgoing.} \end{cases} \tag{2.4}$$

Let us recall that the Helmholtz equation is separable using the prolate spheroidal coordinates $\xi\in[\frac{a}{f},+\infty[$, $\eta\in[-1,1]$ and $\varphi\in[0,2\pi[$, see Claeys (2008); Saint-Guirons (2008); Lebedev & Silverman (1972)

$$\begin{cases} x = f\sqrt{(\xi^2-1)(1-\eta^2)}\cos(\varphi)\\ y = f\sqrt{(\xi^2-1)(1-\eta^2)}\sin(\varphi),\\ z = f\xi\eta, \end{cases} \tag{2.5}$$

where $f>0$ is the interfocal distance of the prolate spheroid Γ

$$f^2 = a^2-b^2.$$

The function u restricted to D' can be determined analytically and parametrized by a family of scalars $u_{m,n} \in \mathbb{C}$, with $n \in \mathbb{N}$ and $-n \leqslant m \leqslant n$

$$u(\mathbf{x}) \;=\; \sum_{n=0}^{+\infty} \sum_{m=-n}^{n} u_{m,n} \frac{\Psi_{m,n}(\xi)}{\Psi_{m,n}(a/f)}\, Y_{m,n}(\eta,\varphi) \quad \text{in } \mathsf{D}' \tag{2.6}$$

with $\Psi_{m,n}$ the prolate radial spheroidal functions of third type, see Abramowitz & Stegun (1964); Flammer (1957); Lebedev & Silverman (1972)

$$\Psi_{m,n} : [1,+\infty[\longrightarrow \mathbb{C} \quad \xi \longmapsto \Psi_{m,n}(\xi);$$

which are defined by the ordinary differential equation, with $c = kf$,

$$\frac{d}{d\xi}\Big((\xi^2-1)\frac{d\psi_{m,n}}{d\xi}\Big)(\xi) \;+\; \Big(c^2\,\xi^2 \;-\; \frac{m^2}{\xi^2-1} \;-\; \lambda_{m,n}(c)\Big)\,\psi_{m,n}(\xi) = \;0,$$

completed by the asymptotic behavior at infinity

$$\psi_{m,n}(\xi) \;=\; \frac{\exp(ic\xi)}{\xi}\Big(1 + \underset{\xi\to+\infty}{o}(1)\Big). \tag{2.7}$$

Evaluating the Neumann traces of (2.6) on $\Gamma = \{\mathbf{x} \in \mathbb{R}^3 : \xi(\mathbf{x}) = a/f\}$

$$\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x}) \;=\; \sum_{n=0}^{+\infty} \sum_{m=-n}^{n} \frac{Z_{m,n}\, u_{m,n}\, Y_{m,n}(\eta(\mathbf{x}),\varphi(\mathbf{x}))}{\sqrt{a^2/f^2 - z^2/a^2}} \text{ on } \Gamma.$$

where $Z_{m,n} \in \mathbb{C}$ is defined by

$$Z_{m,n} \;=\; \frac{b}{f^2}\frac{\dfrac{\partial \psi_{m,n}}{\partial \xi}(\dfrac{a}{f})}{\psi_{m,n}(\frac{a}{f})}. \tag{2.8}$$

This leads to the next proposition

PROPOSITION 2.1 The non local Dirichlet-to-Neumann operator $\mathsf{DtN}_k : H^{\frac{1}{2}}(\Gamma) \longrightarrow H^{-1/2}(\Gamma)$ takes the form

$$\mathsf{DtN}_k : v \longmapsto -\sum_{n=0}^{+\infty} \sum_{m=-n}^{n} \frac{Z_{m,n}\, v_{m,n}\, Y_{m,n}(\mathbf{x})}{\sqrt{a^2/f^2 - z^2/a^2}} \tag{2.9}$$

with $v_{m,n}$ defined by (1.4) replacing u by v. Moreover, there exists a constant C_k such that:

$$\langle \mathsf{DtN}_k u, v\rangle_\Gamma \;\leqslant\; C_k\, \|u\|^2_{\Gamma,1/2}\, \|v\|^2_{\Gamma,1/2} \tag{2.10}$$

In the sequel, the next lemma will be helpful.

LEMMA 2.1 The operator $\mathsf{DtN}_k : H^{\frac{1}{2}}(\Gamma) \longrightarrow H^{-\frac{1}{2}}(\Gamma)$ can be decomposed into

$$\mathsf{DtN}_k \;=\; \mathsf{P} \;+\; \mathsf{K}$$

where $\mathsf{K} : H^{\frac{1}{2}}(\Gamma) \longrightarrow H^{-\frac{1}{2}}(\Gamma)$ is a compact operator and $\mathsf{P} : H^{\frac{1}{2}}(\Gamma) \longrightarrow H^{-\frac{1}{2}}(\Gamma)$ is a positive operator

$$\langle \mathsf{P}u, u\rangle_\Gamma \in \mathbb{R}^+ \quad \text{ for all } \quad u \in H^{\frac{1}{2}}(\Gamma).$$

Proof. Let us denote by $\mathsf{P} : H^{\frac{1}{2}}(\Gamma) \longrightarrow H^{-\frac{1}{2}}(\Gamma)$ and $\mathsf{K} : H^{\frac{1}{2}}(\Gamma) \longrightarrow H^{-\frac{1}{2}}(\Gamma)$ the operators

$$\mathsf{P} \;=\; \mathsf{DtN}_{ik} \quad \mathsf{K} \;=\; \mathsf{DtN}_k - \mathsf{DtN}_{ik}.$$

with DtN_k and DtN_{ik} the Dirichlet-to-Neumann operator

$$\left\{ \begin{array}{ll} \mathsf{DtN}_k : & H^{\frac{1}{2}}(\Gamma) \longrightarrow H^{-\frac{1}{2}}(\Gamma) \quad u|_\Gamma \longmapsto -\partial_{\mathbf{n}} u_k|_\Gamma, \\ \mathsf{DtN}_{ik} : & H^{\frac{1}{2}}(\Gamma) \longrightarrow H^{-\frac{1}{2}}(\Gamma) \quad u|_\Gamma \longmapsto -\partial_{\mathbf{n}} u_{ik}|_\Gamma, \end{array} \right. \tag{2.11}$$

associated with the exterior problems ($\mathbf{n}$ is interior to D')

$$\left\{ \begin{array}{ll} u_k \in H^1_{\rm loc}(\mathsf{D}') \text{ outgoing}, & \\ \Delta u_k + k^2 u_k = 0 & \text{in } \mathsf{D}', \\ u_k = u & \text{on } \Gamma, \end{array} \right. \qquad \left\{ \begin{array}{ll} u_{ik} \in H^1(\mathsf{D}'), & \\ \Delta u_{ik} - k^2 u_{ik} = 0 & \text{in } \mathsf{D}', \\ u_{ik} = u & \text{on } \Gamma. \end{array} \right.$$

In order to conclude we should prove that P is positive and K is compact. The function $u_{ik} - u_k : \mathsf{D}' \longrightarrow \mathbb{C}$ satisfies

$$\left\{ \begin{array}{l} u_{ik} - u_k \in H^1_{\rm loc}(\mathsf{D}'), \\ \Delta(u_{ik} - u_k) - k^2 (u_{ik} - u_k) = 2k^2 u_k \in H^1_{\rm loc}(\mathsf{D}'), \\ u_{ik} - u_k = 0 \text{ on } \Gamma. \end{array} \right.$$

The elliptic regularity theory McLean (2000) implies that $u_{ik} - u_k \in H^3_{\rm loc}(\overline{\mathsf{D}'})$ and consequently that $\partial_{\mathbf{n}} u_{ik} - \partial_{\mathbf{n}} u_k \in H^{\frac{3}{2}}(\Gamma)$. Therefore, the operator K is compact since its range is included in $H^{\frac{3}{2}}(\Gamma)$ which is compactly embedded in $H^{-\frac{1}{2}}(\Gamma)$.

The positivity of P follows from the Green formula:

$$\langle \mathsf{P} u, u \rangle_\Gamma = \langle \mathsf{DtN}_{ik} u, u \rangle_\Gamma = -\int_\Gamma \partial_{\mathbf{n}} u_{ik}\, \overline{u}_{ik} = \int_{\mathsf{D}'} \left| \nabla u_{ik} \right| + k^2 \left| u_{ik} \right| \geqslant 0.$$

This completes the proof. □

2.2 *A well-posed variational formulation*

We aim in this section at proving the existence and the uniqueness of the solution to (0.3) by resorting to the Fredholm alternative.

THEOREM 2.2 The restriction to D of the solution to (0.3) is the unique function $u \in H^1(\mathsf{D})$ satisfying

$$\mathsf{a}(u, v) + \mathsf{b}(u, v) = \ell(v) \quad \forall v \in H^1(\mathsf{D}) \tag{2.12}$$

where the forms $\mathsf{a} : H^1(\mathsf{D}) \times H^1(\mathsf{D}) \longrightarrow \mathbb{C}$, $\mathsf{b} : H^1(\mathsf{D}) \times H^1(\mathsf{D}) \longrightarrow \mathbb{C}$ and $\ell : H^1(\mathsf{D}) \longrightarrow \mathbb{C}$ are given

by

$$\left\{\begin{array}{l} \mathsf{a}(u,v) \;=\; \displaystyle\int_{\mathsf{D}} \Big(\nabla u(\mathbf{x})\cdot\overline{\nabla v(\mathbf{x})} - k^2 u(\mathbf{x})\,\overline{v(\mathbf{x})}\Big)d\mathbf{x}, \\ \\ \mathsf{b}(u,v) \;=\; \langle \mathsf{DtN}_k u,\overline{v}\rangle_{\Gamma} \;=\; -\displaystyle\sum_{n=0}^{+\infty}\sum_{m=-n}^{n} Z_{m,n}u_{m,n}\overline{v_{m,n}}, \\ \\ \ell(v) \;=\; -\displaystyle\int_{\partial\Omega} \frac{\partial u_{\rm inc}}{\partial \mathbf{n}}(\mathbf{x})\overline{v(\mathbf{x})}ds_{\mathbf{x}}. \end{array}\right.$$

with $u_{m,n}\in\mathbb{C}$, $Z_{m,n}$ given by (1.4) and (2.8).

Proof. We first derive the variational formulation. Then, we show that this variational formulation is well posed.

(i) We show that the solution u to (0.3) solves also (2.12). Assuming u regular enough, multiplying $\Delta u + k^2 u = 0$ by the conjugate of a test function $v\in H^1(\mathsf{D})$, integrating over D and using the Green formula we get

$$\int_{\mathsf{D}} \nabla u(\mathbf{x})\cdot\nabla\overline{v}(\mathbf{x}) \;-\; k^2\,u(\mathbf{x})\,\overline{v}(\mathbf{x})d\mathbf{x} \;=\; \left\langle \frac{\partial u}{\partial \mathbf{n}},\overline{v}\right\rangle_{\partial\Omega} + \left\langle \frac{\partial u}{\partial \mathbf{n}},\overline{v}\right\rangle_{\Gamma}$$

According to Eq. (2.2), this can be written as

$$\mathsf{a}(u,v) \;+\langle \mathsf{DtN}_k u,\overline{v}\rangle_{\Gamma} \;=\; \ell(v) \quad \forall v\in H^1(\mathsf{D}).$$

Using (2.9), we conclude that u satisfies (2.12).

(ii) Let us first remark that the bilinear form $\mathsf{a}+\mathsf{b}$ can be decomposed into $\mathsf{a}+\mathsf{b} = \mathsf{c}_{\rm coer}+\mathsf{c}_{\rm com}$ where $\mathsf{c}_{\rm coer}$ is a coercive bilinear form and $\mathsf{c}_{\rm com}$ is a compact bilinear form. This is a consequence of Lemma 2.1 and we have

$$\left\{\begin{array}{rcl} \mathsf{c}_{\rm coer}(u,v) & = & \displaystyle\int_{\mathsf{D}} \nabla u(\mathbf{x})\cdot\nabla\overline{v}(\mathbf{x}) \;+\; k^2\,u(\mathbf{x})\,\overline{v}(\mathbf{x})d\mathbf{x} \;+\; \langle \mathsf{P}u,\overline{v}\rangle_{\Gamma} \\ \\ \mathsf{c}_{\rm com}(u,v) & = & -\,2\,k^2\displaystyle\int_{\mathsf{D}} u(\mathbf{x})\,\overline{v}(\mathbf{x})d\mathbf{x} \;+\; \langle \mathsf{K}u,\overline{v}\rangle_{\Gamma} \end{array}\right.$$

Due to the Fredholm alternative (see Dautray & Lions (1988) Proposition 11 of Chap.VIII), the well-posedness will follow from the uniqueness of the solution. Let us show that $u=0$ in D if

$$\mathsf{a}(u,v)+\mathsf{b}(u,v)=0 \quad \text{for all } v\in H^1(\mathsf{D}).$$

Equivalently, the function u solves the following system

$$\left\{\begin{array}{l} u\in H^1(\mathsf{D}), \\ \Delta u \;+\; k^2\,u \;=\; 0 \quad \text{in } \mathsf{D}, \\ \displaystyle\frac{\partial u}{\partial \mathbf{n}} \;=\; 0 \quad \text{on } \partial\Omega, \\ \displaystyle\frac{\partial u}{\partial \mathbf{n}} \;+\; \mathsf{DtN}_k u \;=\; 0 \quad \text{on } \Gamma. \end{array}\right. \tag{2.13}$$

Picking $v = u$ as test function, we have

$$\Im\Big(\mathsf{a}(u,u)+\mathsf{b}(u,u)\Big) \;=\; \Im\Big(\mathsf{b}(u,u)\Big) \;=\; 0.$$

Evaluating this bilinear form, we have

$$\sum_{n=0}^{+\infty}\sum_{m=-n}^{n} \Im Z_{m,n}\big|u_{m,n}\big|^2 = 0$$

with $u_{m,n}$ given by (1.4).

Due to lemma 1 in Barucq *et al.* (2012), we have for all $n \in \mathbb{N}$ and $-n \leqslant m \leqslant n$

$$\Im Z_{m,n} < 0.$$

It follows that

$$u_{m,n} \;=\; 0 \quad \text{for all } n \in \mathbb{N} \text{ and } -n \leqslant m \leqslant n.$$

Due to (2.9) and the last line of system (2.13),we have

$$u \;=\; 0 \quad \text{and} \quad \frac{\partial u}{\partial \mathbf{n}} \;=\; 0 \quad \text{on } \Gamma$$

The unique continuation Theorem Protter (1960) ensures that

$$u \;=\; 0 \quad \text{in } D.$$

This concludes the proof. □

2.3 *Truncation of the variational formulation*

The bilinear form b involves a series which should be truncated for numerical purposes. Denote by $N \in \mathbb{N}$ the order of truncation, a first idea consists in truncating the bilinear form b in the following way

$$\widetilde{\mathsf{b}_N}(u,v) \;=\; -\sum_{n=0}^{N}\sum_{m=-n}^{n} Z_{m,n}\, u_{m,n}\, \overline{v_{m,n}}.$$

However, for some geometries, the associated variational formulation is not well-posed due to spurious modes. The approach of Lenoir & Tounsi (1988) is then reproduced in order to bypass this difficulty. Since $(Y_{m,n})_{m,n}$ is an orthonormal basis for the inner product $(\cdot,\cdot)_{\Gamma,0}$, we have the identity

$$\sum_{n=0}^{+\infty}\sum_{m=-n}^{n} u_{m,n}\overline{v_{m,n}} = \int_\Gamma \frac{u(\mathbf{x})\overline{v}(\mathbf{x})}{\sqrt{\frac{a^2}{f^2}-\frac{z^2}{a^2}}} ds_{\mathbf{x}}.$$

This leads to the expression

$$\mathsf{b}(u,v) \;=\; -\sum_{n=0}^{+\infty}\sum_{m=-n}^{n} \Big(Z_{m,n}+\alpha\Big) u_{m,n}\overline{v_{m,n}} + \alpha \int_\Gamma \frac{u(\mathbf{x})\overline{v}(\mathbf{x})}{\sqrt{\frac{a^2}{f^2}-\frac{z^2}{a^2}}} ds_{\mathbf{x}}.$$

The parameter $\alpha \in \mathbb{C}$ does not play any role as long as b is considered. In practice, b is replaced by an approximation which is obtained by truncating the series. We then define

$$\mathsf{b}_{N,\alpha}(u,v) = -\sum_{n=0}^{N}\sum_{m=-n}^{n}\Big(Z_{m,n}+\alpha\Big)u_{m,n}\overline{v_{m,n}} + \alpha\int_{\Gamma}\frac{u(\mathbf{x})\overline{v}(\mathbf{x})}{\sqrt{\frac{a^2}{f^2}-\frac{z^2}{a^2}}}ds_{\mathbf{x}},$$

which can also be written as:

$$\mathsf{b}_{N,\alpha}(u,v) = \langle \mathsf{DtN}_k \mathsf{Q}_N u, v\rangle_{\Gamma} + \alpha\,(u-\mathsf{Q}_N u, v)_{\Gamma,0} \tag{2.14}$$

where $\mathsf{Q}_N : L^2(\Gamma) \longrightarrow C^{\infty}(\Gamma)$ is the linear operator defined by

$$\mathsf{Q}_N \;:\; \left\{\begin{array}{ccc} L^2(\Gamma) & \longrightarrow & C^{\infty}(\Gamma) \\ u = \displaystyle\sum_{n=0}^{+\infty}\sum_{m=-n}^{n} u_{m,n}\,Y_{m,n} & \longmapsto & \displaystyle\sum_{n=0}^{N}\sum_{m=-n}^{n} u_{m,n}\,Y_{m,n}\end{array}\right.$$

or equivalently

$$(\mathsf{Q}_N u)_{m,n} \;=\; u_{m,n} \text{ if } n \leqslant N \quad \text{ and } \quad (\mathsf{Q}_N u)_{m,n} \;=\; 0 \text{ if } n > N. \tag{2.15}$$

LEMMA 2.2 The linear operators $\mathsf{Q}_N : H^s(\Gamma) \longrightarrow H^s(\Gamma)$ satisfy

(i) For all $s \in [0,1]$, Q_N is an orthogonal projection in $H^s(\Gamma)$ for the scalar product $(.,.)_{\Gamma,s}$ and we have

$$\|\mathsf{Q}_N u\|^2_{\Gamma,s} \;+\; \|(\mathsf{I}-\mathsf{Q}_N)u\|^2_{\Gamma,s} \;=\; \|u\|^2_{\Gamma,s}. \tag{2.16}$$

(ii) For all $s \in [0,1]$, the sequence of operators Q_N converges strongly to Identity

$$\lim_{N\to+\infty}\|Q_N u - u\|_{\Gamma,s} = 0 \quad \forall u \in H^s(\Gamma).$$

(iii) For all u and v in $H^{\frac{1}{2}}(\Gamma)$, we have the identity

$$\langle \mathsf{DtN}_k \mathsf{Q}_N u, \mathsf{Q}_N v\rangle_{\Gamma} \;=\; \langle \mathsf{DtN}_k u, \mathsf{Q}_N v\rangle_{\Gamma} \;=\; \langle \mathsf{DtN}_k \mathsf{Q}_N u, v\rangle_{\Gamma}\,. \tag{2.17}$$

Proof.

(i) By definition of $\|\cdot\|_{\Gamma,s}$, we have

$$\left\{\begin{array}{rcl} \|u\|^2_{\Gamma,s} & = & \displaystyle\sum_{n=0}^{+\infty}\sum_{m=-n}^{n}(\lambda_{m,n}(c))^s\,|u_{m,n}|^2 \\ & = & \displaystyle\sum_{n=0}^{N}\sum_{m=-n}^{n}(\lambda_{m,n}(c))^s\,|u_{m,n}|^2 \\ & + & \displaystyle\sum_{n=N+1}^{+\infty}\sum_{m=-n}^{n}(\lambda_{m,n}(c))^s\,|u_{m,n}|^2 \end{array}\right.$$

It follows from (2.15) that

$$\left\{\begin{array}{rcl} \|u\|^2_{\Gamma,s} & = & \displaystyle\sum_{n=0}^{+\infty}\sum_{m=-n}^{n}(\lambda_{m,n}(c))^s\,|(\mathsf{Q}_N u)_{m,n}|^2 \\ & + & \displaystyle\sum_{n=0}^{+\infty}\sum_{m=-n}^{n}(\lambda_{m,n}(c))^s\,|(u-\mathsf{Q}_N u)_{m,n}|^2 \\ & = & \|\mathsf{Q}_N u\|^2_{\Gamma,s} + \|u-\mathsf{Q}_N u\|^2_{\Gamma,s} \end{array}\right.$$

(ii) Since $u \in H^s(\Gamma)$, we have

$$\sum_{n=N+1}^{+\infty}\sum_{m=-n}^{n}(\lambda_{m,n}(c))^s\,|u_{m,n}|^2 < +\infty.$$

It follows that the remainder of this series converges to zero

$$\|u-\mathsf{Q}_N u\|^2_{\Gamma,s} = \sum_{n=N+1}^{+\infty}\sum_{m=-n}^{n}(\lambda_{m,n}(c))^s\,|u_{m,n}|^2 \underset{N\to+\infty}{\longrightarrow} 0.$$

(iii) Let us denote by I the expression

$$I = \sum_{n=0}^{N}\sum_{m=-n}^{n} Z_{m,n}\,u_{m,n}\,\overline{v_{m,n}}$$

The relation (2.17) follows from (2.15) and

$$\left\{\begin{array}{rcl} \langle \mathsf{DtN}_k \mathsf{Q}_N u, \mathsf{Q}_N v\rangle_\Gamma & = & \displaystyle\sum_{n=0}^{+\infty}\sum_{m=-n}^{n} Z_{m,n}\,(\mathsf{Q}_N u)_{m,n}\,\overline{(\mathsf{Q}_N v)_{m,n}} = I, \\ \langle \mathsf{DtN}_k u, \mathsf{Q}_N v\rangle_\Gamma & = & \displaystyle\sum_{n=0}^{+\infty}\sum_{m=-n}^{n} Z_{m,n}\,u_{m,n}\,\overline{(\mathsf{Q}_N v)_{m,n}} = I, \\ \langle \mathsf{DtN}_k \mathsf{Q}_N u, v\rangle_\Gamma & = & \displaystyle\sum_{n=0}^{+\infty}\sum_{m=-n}^{n} Z_{m,n}\,(\mathsf{Q}_N u)_{m,n}\,\overline{v_{m,n}} = I, \end{array}\right.$$

□

2.4 *Convergence*

In this section we aim at proving the following Theorem.

THEOREM 2.3 Let $\alpha \in i\mathbb{R}^{*-}$. There exists a unique $u_N \in H^1(\mathsf{D})$ such that

$$\mathsf{a}(u_N, v) + \mathsf{b}_{N,\alpha}(u_N, v) = \ell(v) \quad \forall v \in H^1(\mathsf{D}) \tag{2.18}$$

and

$$\lim_{N\to+\infty} \left\| u_N - u \right\|_{H^1(\mathsf{D})} = 0. \tag{2.19}$$

Proof. (i) The proof of well-posedness is similar to the part (ii) of the proof of Theorem 2.2. Indeed it suffices to remark that

$$\mathsf{b}_{N,\alpha}(u,v)=\sum_{n=0}^{+\infty}\sum_{m=-n}^{n}\ \widetilde{Z}_{N,m,n}\,u_{m,n}\,\overline{v_{m,n}}.$$

with

$$\widetilde{Z}_{N,m,n}=Z_{m,n}\quad\text{ for } n\leqslant N,\qquad \widetilde{Z}_{N,m,n}=\alpha\quad\text{ for } n>N.$$

(ii) Now we prove (2.19). Let A, B and $\mathsf{B}_{N,\alpha}:H^1(\mathsf{D})\longrightarrow H^1(\mathsf{D})$ be the linear operators related to the bilinear forms a, b, $\mathsf{b}_{N,\alpha}$

$$\begin{cases}(\mathsf{A}u,v)_{H^1(\mathsf{D})}\ =\ \mathsf{a}(u,v),\quad (\mathsf{B}u,v)_{H^1(\mathsf{D})}\ =\ \mathsf{b}(u,v),\\ (\mathsf{B}_{N,\alpha}u,v)_{H^1(\mathsf{D})}\ =\ \mathsf{b}_{N,\alpha}(u,v),\quad \text{for all } u \text{ and } v \text{ in } H^1(\mathsf{D})\end{cases}$$

where $(\cdot,\cdot)_{H^1(\mathsf{D})}$ denotes the inner product

$$(u,v)_{H^1(\mathsf{D})}\ =\ \int_{\mathsf{D}}\nabla u(\mathbf{x})\cdot\nabla\overline{v}(\mathbf{x})+u(\mathbf{x})\overline{v}(\mathbf{x})d\mathbf{x}.$$

We associate to ℓ the function $\mathsf{f}\in H^1(\mathsf{D})$ defined by the Riesz Lemma

$$(\mathsf{f},v)_{H^1(\mathsf{D})}\ =\ \ell(v),$$

Identifying each term, we then have

$$(\mathsf{A}+\mathsf{B})u=\mathsf{f}\quad\text{ and }(\mathsf{A}+\mathsf{B}_{N,\alpha})u_N=\mathsf{f}$$

By linearity, we then deduce that

$$(\mathsf{A}+\mathsf{B}_{N,\alpha})(u-u_N)=(\mathsf{B}_{N,\alpha}-\mathsf{B})u.$$

and therefore

$$u-u_N=(\mathsf{A}+\mathsf{B}_{N,\alpha})^{-1}(\mathsf{B}_{N,\alpha}-\mathsf{B})u.$$

The conclusion follows from the following lemmas whose proofs are postponed to section 2.5 and 2.6 □

LEMMA 2.3 We have the uniform stability estimate

$$\exists C>0:\quad \forall N>0\quad |||(\mathsf{A}+\mathsf{B}_{N,\alpha})^{-1}|||\ \leqslant\ C.$$

where $|||\cdot|||$ designates the operator norm on $H^1(\mathsf{D})$

$$|||\mathsf{C}|||\ =\ \sup_{v\neq 0,\ v\in H^1(\mathsf{D})}\frac{\|\mathsf{C}v\|_{H^1(\mathsf{D})}}{\|v\|_{H^1(\mathsf{D})}}$$

LEMMA 2.4 For all $u\in H^1(\mathsf{D})$, we have

$$\lim_{N\to+\infty}\big\|(\mathsf{B}-\mathsf{B}_{N,\alpha})u\big\|_{H^1(\mathsf{D})}\ =\ 0.$$

2.5 *Proof of Lemma 2.3*

Since the operator $\mathsf{A}+\mathsf{B}_{N,\alpha}$ is invertible for all $N \in \mathbb{N}$, it suffices to prove that

$$\exists (N,C) \in \mathbb{N}\times\mathbb{R}^+ \quad \forall \sigma > N : \qquad |||(\mathsf{A}+\mathsf{B}_\sigma)^{-1}||| \leqslant C.$$

or equivalently

$$\exists (N,C) \in \mathbb{N}\times\mathbb{R}^+ \quad \forall \sigma > N \ \forall v \in H^1(\mathsf{D}) : \qquad \|v\|_{H^1(\mathsf{D})} \leqslant C\,\|(\mathsf{A}+\mathsf{B}_\sigma)v\|_{H^1(\mathsf{D})}.$$

We act by contradiction. Let us suppose that

$$\forall (N,C) \in \mathbb{N}\times\mathbb{R}^+ \quad \exists \sigma > N \ \exists v \in H^1(\mathsf{D}) : \qquad \|v\|_{H^1(\mathsf{D})} > C\,\|(\mathsf{A}+\mathsf{B}_\sigma)v\|_{H^1(\mathsf{D})}.$$

For each $N \in \mathbb{N}$, we choose $C = N$. We are then able to construct a sequence $(\sigma_N)_{N\in\mathbb{N}}$ and a sequence of functions $(v_N)_{N\in\mathbb{N}}$ such that

$$\sigma_N > N, \quad v_N \in H^1(\mathsf{D}), \quad \|v_N\|_{H^1(\mathsf{D})} > N\|(\mathsf{A}+\mathsf{B}_{\sigma_N})v_N\|_{H^1(\mathsf{D})}$$

Denoting by $u_N = \dfrac{v_N}{\|v_N\|_{H^1(\mathsf{D})}}$, we have

$$\lim_{N\to+\infty} \sigma_N = +\infty \quad \|u_N\|_{H^1(\mathsf{D})} = 1, \quad \lim_{N\to+\infty} \|(\mathsf{A}+\mathsf{B}_{\sigma_N})u_N\|_{H^1(\mathsf{D})} = 0. \tag{2.20}$$

Moreover, since u_N is bounded in $H^1(\mathsf{D})$, one can extract from (σ_N, u_N) a subsequence, still denoted (σ_N, u_N), such that u_N converges weakly to u

$$\underset{N\to+\infty}{\text{w-lim}} u_N = u \quad \text{ in } H^1(\mathsf{D}).$$

For all $v \in H^1(\mathsf{D})$ we first remark that:

$$\begin{aligned} |\mathsf{a}(u_N,v) + \mathsf{b}_{\sigma_N}(u_N,v)| &= \left|\left((\mathsf{A}+\mathsf{B}_{\sigma_N})u_N, v\right)_{H^1(\mathsf{D})}\right| \\ &\leqslant \|(\mathsf{A}+\mathsf{B}_{\sigma_N})u_N\|_{H^1(\mathsf{D})}\,\|v\|_{H^1(\mathsf{D})} \end{aligned}$$

Due to (2.20), we thus have:

$$\lim_{N\to+\infty} |\mathsf{a}(u_N,v) + \mathsf{b}_{\sigma_N}(u_N,v)| = 0 \tag{2.21}$$

Now we consider $v \in H^1(\mathsf{D})$ such that $v = 0$ on Γ. The bilinear form involves only the trace of v and so $\mathsf{b}_{\sigma_N}(u_N,v) = 0$. Thus, we get:

$$\mathsf{a}(u_N,v) + \mathsf{b}_{\sigma_N}(u_N,v) = \mathsf{a}(u_N,v) \tag{2.22}$$

Since u_N converges weakly to u in $H^1(\mathsf{D})$, we have

$$\lim_{N\to+\infty} \mathsf{a}(u_N,v) = \mathsf{a}(u,v)$$

Due to (2.21) and (2.22), we have

$$\mathsf{a}(u,v)=0 \quad \forall v\in H^1(\mathsf{D}) \text{ such as } v=0 \text{ on } \Gamma.$$

It follows from the Green formula that

$$\left\{\begin{array}{ll} \Delta u + k^2\, u = 0 & \text{in } \mathsf{D},\\ \dfrac{\partial u}{\partial \mathbf{n}}(\mathbf{x}) = 0 & \text{on } \partial\Omega. \end{array}\right. \tag{2.23}$$

Then, we consider as test function

$$v(\mathbf{x}) = Y_{m,n}(\eta(\mathbf{x}),\varphi(\mathbf{x}))\,\chi(\xi(\mathbf{x}))$$

where χ is a regular chop-off function satisfying $\chi(\xi)=0$ in a neighborhood of 1 and $\chi(\xi)=1$ in a neighborhood of a/f. For $\sigma_N \geqslant N$, we have

$$\mathsf{a}(u_N,v)+\mathsf{b}_{\sigma_N}(u_N,v) = \int_{\mathsf{D}} \nabla u_N\cdot\nabla\overline{v} - k^2 u_N\overline{v} - Z_{m,n}(u_N)_{m,n}.$$

Letting $N\to+\infty$, we get since u_N converges weakly to u in $H^1(\mathsf{D})$

$$\int_{\mathsf{D}} \nabla u\cdot\nabla\overline{v} - k^2 u\overline{v} - Z_{m,n}u_{m,n} = 0.$$

Integration by parts and taking into account (2.23), we obtain

$$\left\langle \frac{\partial u}{\partial \mathbf{n}}(\mathbf{x}), \overline{Y_{m,n}} \right\rangle_{\Gamma} - Z_{m,n}u_{m,n} = 0.$$

Comparing with (2.9), it follows

$$\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x}) + \mathsf{DtN}_k u(\mathbf{x}) = 0 \quad \text{on } \Gamma. \tag{2.24}$$

Due to (2.23) and (2.24), the weak limit u satisfies

$$\mathsf{a}(u,v) + \mathsf{b}(u,v) = 0 \quad \forall v\in H^1(\mathsf{D}).$$

Then, theorem 2.2 allows to conclude that $u=0$ in D.

Taking as test function $v=u_N$ in (2.18), we have

$$\|\nabla u_N\|^2_{L^2(\mathsf{D})} = k^2\|u_N\|^2_{L^2(\mathsf{D})} - \mathsf{b}_{\sigma_N}(u_N,u_N)_{H^1(\mathsf{D})} + \big((\mathsf{A}+\mathsf{B}_{\sigma_N})u_N,u_N\big)_{H^1(\mathsf{D})}. \tag{2.25}$$

Taking the real part of (2.25), we have

$$\|\nabla u_N\|^2_{L^2(\mathsf{D})} = k^2\|u_N\|^2_{L^2(\mathsf{D})} - \Re\mathsf{b}_{\sigma_N}(u_N,u_N)_{H^1(\mathsf{D})} + \Re\big((\mathsf{A}+\mathsf{B}_{\sigma_N})u_N,u_N\big)_{H^1(\mathsf{D})}.$$

Since α is a pure imaginary number, it follows from (2.14) and lemma 2.1

$$\left\{\begin{array}{rcl} \Re\mathsf{b}_{\sigma_N}(u_N,u_N) &=& \Re\langle \mathsf{DtN}_k\mathsf{Q}_{\sigma_N}u_N, \mathsf{Q}_{\sigma_N}u_N\rangle_{\Gamma}\\ &=& \Re\langle \mathsf{PQ}_{\sigma_N}u_N, \mathsf{Q}_{\sigma_N}u_N\rangle_{\Gamma} + \Re\langle \mathsf{KQ}_{\sigma_N}u_N, \mathsf{Q}_{\sigma_N}u_N\rangle_{\Gamma} \end{array}\right.$$

The Cauchy-Schwartz inequality allows to get the estimate

$$\left\|\nabla u_N\right\|^2_{H^1(\mathsf{D})} \leqslant k^2\|u_N\|^2_{L^2(\mathsf{D})} - \langle \mathsf{K}Q_{\sigma_N}u_N, Q_{\sigma_N}u_N\rangle_\Gamma + \left\|(\mathsf{A}+\mathsf{B}_{\sigma_N})u_N\right\|_{H^1(\mathsf{D})}\left\|u_N\right\|_{H^1(\mathsf{D})}. \quad (2.26)$$

Since u_N converges weakly to 0 in $H^1(\mathsf{D})$, it converges strongly in $L^2(\mathsf{D})$ due to the Rellich-Kondrachov Theorem

$$\lim_{N\to+\infty}\|u_N\|^2_{L^2(\mathsf{D})} = 0 \quad (2.27)$$

Moreover, since the operator Q_N converges strongly to Identity in $H^{\frac{1}{2}}(\Gamma)$, $Q_{\sigma_N}u_N$ converges weakly to 0 in $H^{\frac{1}{2}}(\Gamma)$. Since K is compact, we have

$$\lim_{N\to+\infty}\langle \mathsf{K}Q_{\sigma_N}u_N, Q_{\sigma_N}u_N\rangle_\Gamma = 0 \quad (2.28)$$

It follows from (2.20) that

$$\lim_{N\to+\infty}\left(\left\|(\mathsf{A}+\mathsf{B}_{\sigma_N})u_N\right\|_{H^1(\mathsf{D})}\left\|u_N\right\|_{H^1(\mathsf{D})}\right) = 0. \quad (2.29)$$

Letting $N \mapsto \infty$ in (2.26), equations (2.27), (2.28), (2.29) imply that

$$\lim_{N\to+\infty}\|\nabla u_N\|^2_{L^2(\mathsf{D})} = 0. \quad (2.30)$$

Finally, we have due to (2.27) and (2.30)

$$\lim_{N\to+\infty}\|u_N\|^2_{H^1(\mathsf{D})} = 0. \quad (2.31)$$

This contradicts (2.21) and the proof is completed.

2.6 *Proof of Lemma 2.4*

For this proof, we adopt the notation: $\widetilde{w}_N = (\mathsf{I}-\mathsf{Q}_N)w$, for $w = u$ or v. According to Lemma 2.2, we have

$$\lim_{N\to+\infty}\|\widetilde{w}_N\|_{\Gamma,\frac{1}{2}} = 0 \qquad \text{and} \qquad \|\widetilde{w}_N\|_{\Gamma,\frac{1}{2}} \leqslant \|w\|_{\Gamma,\frac{1}{2}} \quad (2.32)$$

By definition, we have

$$(\mathsf{B}u - \mathsf{B}_{N,\alpha}u, v)_{H^1(\mathsf{D})} = \langle \mathsf{DtN}_k\widetilde{u}_N, \widetilde{v}_N\rangle_\Gamma - \alpha\,(\widetilde{u}_N, \widetilde{v}_N)_{\Gamma,0}.$$

We then get from (2.10)

$$\left|(\mathsf{B}u - \mathsf{B}_{N,\alpha}u, v)_{H^1(\mathsf{D})}\right| \leqslant C_k\,\|\widetilde{u}_N\|_{\Gamma,\frac{1}{2}}\,\|\widetilde{v}_N\|_{\Gamma,\frac{1}{2}} + \alpha\,\|\widetilde{u}_N\|_{\Gamma,0}\,\|\widetilde{v}_N\|_{\Gamma,0}$$

It follows that

$$(\mathsf{B}u - \mathsf{B}_{N,\alpha}u, v)_{H^1(\mathsf{D})} \leqslant C\,\|\widetilde{u}_N\|_{\Gamma,\frac{1}{2}}\,\|\widetilde{v}_N\|_{\Gamma,\frac{1}{2}}$$

Due to (2.16), we have the identity $\|\widetilde{v}_N\|_{\Gamma,\frac{1}{2}} \leqslant \|v\|_{\Gamma,\frac{1}{2}}$. The trace theorem allows to get

$$(\mathsf{B}u - \mathsf{B}_{N,\alpha}u, v)_{H^1(\mathsf{D})} \leqslant C\,\|\widetilde{u}_N\|_{\Gamma,\frac{1}{2}}\,\|v\|_{\Gamma,\frac{1}{2}} \leqslant C\,\|\widetilde{u}_N\|_{\Gamma,\frac{1}{2}}\,\|v\|_{H^1(\mathsf{D})}$$

The conclusion follows from the characterization of the H^1-norm

$$\|\mathsf{B}u - \mathsf{B}_{N,\alpha}u\|_{H^1(\mathsf{D})} = \sup_{v\in H^1(\mathsf{D})}\frac{(\mathsf{B}u - \mathsf{B}_{N,\alpha}u, v)_{H^1(\mathsf{D})}}{\|v\|_{H^1(\mathsf{D})}} \leqslant C\,\|\widetilde{u}_N\|_{\Gamma,\frac{1}{2}}.$$

Equation (2.32) allows to conclude.

3. Localized finite elements

In order to discretize the variational formulation we follow the finite element discretization proposed in Lenoir & Tounsi (1988). The computational domain is discretized with a tetrahedral mesh composed of straight elements, see Figure 1 (the geometry is not exactly resolved). The discretized computational domain D and its boundary Γ are denoted by D_h and Γ_h.

3.1 *Matrix formulation*

The discrete solution is given as

$$u_h(\mathbf{x}) \;=\; \sum_{i=1}^{I} \mathbf{u}_i \, w_i(\mathbf{x})$$

with w_i being the classical piece wise linear continuous shape functions ($\mathbb{P}_1$-continuous finite element) and I the number of basis function. The unknown column vector $\mathbf{U} = (\mathbf{u}_i)_{i\in[\![1,I]\!]}$, $\mathbf{u}_i \in \mathbb{C}$, is defined as the solution of

$$\Big(\mathbf{A} + \mathbf{B}_{N,\alpha}\Big)\mathbf{U} = \mathbf{F} \tag{3.1}$$

with $\mathbf{A} \in \mathbb{R}^{I\times I}$, $\mathbf{B}_{N,\alpha} \in \mathbb{R}^{I\times I}$ and $\mathbf{F} \in \mathbb{R}^{I}$ given by

$$\left\{\begin{array}{l} \mathbf{A}_{i,j} \;=\; \displaystyle\int_{\Omega_h} \nabla w_i(\mathbf{x}) \cdot \nabla w_j(\mathbf{x}) d\mathbf{x} - k^2 \int_{\Omega_h} w_i(\mathbf{x})\, w_j(\mathbf{x}) d\mathbf{x} \\ \qquad\qquad\qquad\qquad + \alpha \displaystyle\int_{\Gamma_h} \frac{w_i(\mathbf{x})\, w_j(\mathbf{x})}{\sqrt{a^2/f^2 - z^2/a^2}} ds_{\mathbf{x}}, \\ \mathbf{B}_{N,\alpha} := \; -\displaystyle\sum_{n=0}^{N} \sum_{m=-n}^{n} \Big(Z_{m,n} + \alpha\Big)\, \mathbf{M}_\Gamma\, \mathbf{Y}_{m,n}\, (\mathbf{Y}_{m,n})^*\, \mathbf{M}_\Gamma \end{array}\right.$$

with $\mathbf{M}_\Gamma \in \mathbb{R}^{I\times I}$ the weighted surface mass matrix and $\mathbf{Y}_{m,n}$ the vector composed of the nodal values of $Y_{m,n}$

$$\left\{\begin{array}{lcl} \mathbf{M}_\Gamma & = & \Big(\displaystyle\int_{\Gamma_h} \frac{w_i(\mathbf{x})\, w_j(\mathbf{x})}{\sqrt{a^2/f^2 - z^2/a^2}} ds_{\mathbf{x}}\Big)_{i,j\in[\![1,I]\!]}, \\ \mathbf{Y}_{m,n} & = & \Big(Y_{m,n}(\eta(\mathbf{x}_i), \varphi(\mathbf{x}_i)) 1_\Gamma(\mathbf{x}_i)\Big)_{i\in[\![1,I]\!]} \end{array}\right.$$

with $1_\Gamma(\mathbf{x}) = 1$ if $\mathbf{x} \in \Gamma$ and vanishes otherwise.

3.2 *Solution with the Sherman-Morrison algorithm*

One of the main drawbacks of the high order discretization of the radiation condition is that it involves full matrices which are generally the source of a numerical overcost.

In the proposed method, the matrix discretizing the radiation condition is also full but will not be responsible of a numerical burden. Remarking that this matrix is the sum of $(N+1)^2$ matrices with rank one, we can apply the Sherman-Morrison algorithm (Sherman & Morrison (1950)) which exploits the sparsity of the matrix $\mathbf{A}$. This method consists of two steps.

Step 1. We determine the scalar complex numbers

$$\mathbf{W}_{m,n} = (\widetilde{\mathbf{Y}}_{m,n})^* U \in \mathbb{C} \quad \text{ for } 0 \leqslant n \leqslant N \text{ and } -n \leqslant m \leqslant n.$$

Multiplying (3.1) by $(\widetilde{\mathbf{Y}}_{m,n})^*\mathbf{A}^{-1}$, with $\widetilde{\mathbf{Y}}_{m,n} = \mathbf{M}_\Gamma\, \mathbf{Y}_{m,n}$, we get

$$(\widetilde{\mathbf{Y}}_{m,n})^*\mathbf{U} + (\widetilde{\mathbf{Y}}_{m,n})^*\mathbf{A}^{-1}\mathbf{B}_{N,\alpha}\mathbf{U} = (\widetilde{\mathbf{Y}}_{m,n})^*\,\mathbf{A}^{-1}\mathbf{F}$$

Evaluating the left hand side, it follows that the $W_{m,n}$, for $0 \leqslant n \leqslant N$ and $-n \leqslant m \leqslant n$, solves the linear system

$$\mathbf{W}_{m,n} \;-\; \sum_{n'=0}^{N}\sum_{m'=-n'}^{n'} \mathbf{C}_{(m,n),(m',n')}\mathbf{W}_{m',n'} \;=\; (\widetilde{\mathbf{Y}}_{m,n})^*\mathbf{A}^{-1}\mathbf{F}.$$

with

$$\mathbf{C}_{(m,n),(m',n')} \;=\; \Big(Z_{m',n'} + \alpha\Big)(\widetilde{\mathbf{Y}}_{m,n})^*\mathbf{A}^{-1}\widetilde{\mathbf{Y}}_{m',n'}.$$

This system of size $(N+1)^2$ is solved to get the $\mathbf{W}_{m,n}$.

Step 2. It consists in remarking that

$$\left\{\begin{array}{rcl} \mathbf{B}_{N,\alpha}U & = & -\displaystyle\sum_{n=0}^{N}\sum_{m=-n}^{n}\Big(Z_{m,n}+\alpha\Big)\widetilde{\mathbf{Y}}_{m,n}(\widetilde{\mathbf{Y}}_{m,n})^*U \\ & = & -\displaystyle\sum_{n=0}^{N}\sum_{m=-n}^{n}\Big(Z_{m,n}+\alpha\Big)\widetilde{\mathbf{Y}}_{m,n}\mathbf{W}_{m,n} \end{array}\right. .$$

Consequently the solution U can be deduced by the formula

$$U \;=\; \mathbf{A}^{-1}\Big(\mathbf{F} + \sum_{n=0}^{N}\sum_{m=-n}^{n}\Big(Z_{m,n}+\alpha\Big)\mathbf{W}_{m,n}\widetilde{\mathbf{Y}}_{m,n}\Big).$$

REMARK 3.1 Note that

- No full matrix with large size has been inverted during both steps of the method.
- Only two matrices have been inverted: the matrix A, which is sparse and factorized by the public domain library MUMPS (MUlti-frontal Massively Parallel Solver) Amestoy *et al.* (2001), for $(N+1)^2+1$ different right hand sides and a small dense matrix with size $(N+1)^2$.

4. Numerical procedure and results

In this section we present numerical examples demonstrating the applicability of the method presented in this paper and we suggest an optimal choice for $N \in \mathbb{N}$. The elongated object is a submarine (borrowed from Hetmaniuk & Farhat (2003)). We first study the influence of the mesh size as well as the thickness of the computational domain D on the accuracy of the numerical solution. The prolate spheroidal radial functions of the first kind and their first derivatives are computed with the Fortran code developed in Burn & Boisvert (2002, 2015). We use the abbreviation SRBC to denote the newly proposed prolate spheroidal radiation boundary conditions.

4.1 *The model problem*

We consider the acoustic scattering by a hard-sound obstacle. The tube of this mock-up submarine is of length $L_A = 5\lambda$ and of diameter $D_A = \lambda$. The tower is of length $L_T = \lambda$ and of height $H_T = 0.3\lambda$, with

λ the wavelength. Our main interest is to predict the scattering of the time-harmonic incident wave in the direction

$$(-\sqrt{3}/3, -\sqrt{3}/3, -\sqrt{3}/3)$$

in the frequency regime corresponding to $k = \frac{2\pi}{\lambda}$.

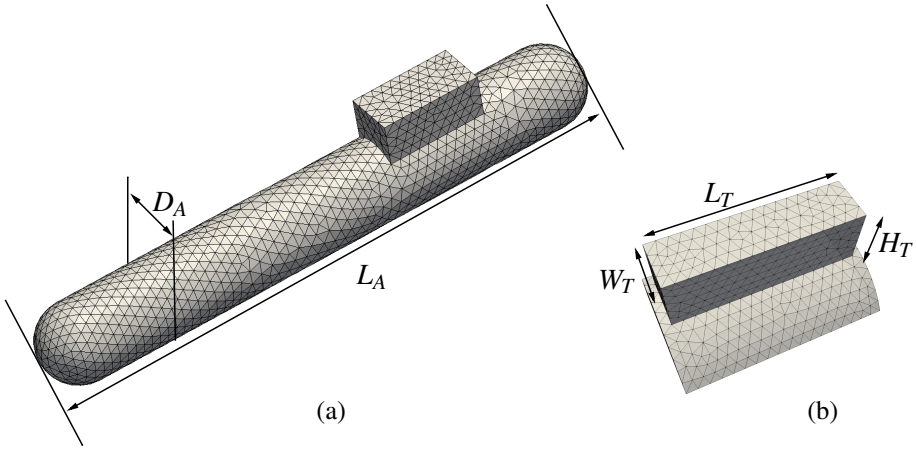

FIG. 1. The considered geometry: (a) mockup submarine, (b) the tower

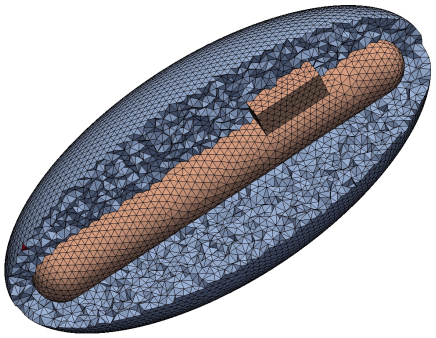

FIG. 2. Geometry with mesh: prolate spheroid of diameters $a = 3.5\lambda$, $b = 1.5\lambda$. Mesh contains around $4 \cdot 10^4$ tetrahedra

4.2 *Tools for the numerical analysis*

4.2.1 *The reference solution.* As for the considered complex geometry there is no exact solution to our equations available, we compute an accurate solution by the Boundary Element Method and use it as reference solution. This is justified, as the BEM discretization maps the far-field behavior onto

the boundary and produces highly accurate results. The calculations were performed using the in-house boundary element/finite element code CESC, a solver for acoustic scattering problems developed at CERFACS. If the reader likes to redo the shown experiments, an open-source alternative to CESC would be the code Bempp (Śmigaj *et al.* (2015)).

In the CESC package, the case of the hard-sound scatterer is treated by solving the integral equation (Bendali & Fares, 2008, p. 9, Eq. (1.36))

$$\mathscr{K}\xi = \frac{1}{2}\xi I + K\xi = -u^{inc}|_{\partial\Omega}$$

with unknown potential ξ and

$$K\xi(x) = \int_{\partial\Omega} \partial_{n_y} G(x,y)\xi(y)ds(y),\ x \in \partial\Omega \quad .$$

$G(x,y) = \dfrac{e^{ik|x-y|}}{4\pi|x-y|}$ is the fundamental solution to the Helmholtz equation. The numerical procedure is based upon a triangular meshing of the surface and uses linear basis functions. We seek an approximate solution ξ_h, such that

$$\begin{cases} \dfrac{1}{2}\displaystyle\int_{\partial\Omega_h} \xi_h(x)\xi_h'(x) + \int_{\partial\Omega_h}\int_{\partial\Omega_h} \partial_n G(x,y)\xi_h(y)\xi_h'(x)ds(y)ds(x) = \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad -\displaystyle\int_{\partial\Omega_h} u^{inc}(x)\xi_h'(x)ds(x), \\ \text{for all test function } \xi_h'(x) \end{cases}$$

This discretization leads to solving a linear system with a dense complex symmetric (non-hermitian) matrix whose size is determined by the number of nodes of the mesh. For improved accuracy, the length of the longest edge in the discretization does not exceed a tenth of the wavelength. Furthermore, we paid special attention to properly taking into account the singularity of the Green's function kernel during the assembly process Bendali (1984). The LU decomposition of the matrix is performed by means of a set of ScaLAPACK parallel routines. Once ξ_h has been obtained, the associated far fields are easily deduced by applying the integral representation formula

$$u_h(x) = u^{inc}(x) + K\xi_h(x) \quad x \notin \partial\Omega.$$

4.2.2 *Coupled FEM-BEM discretization.* Another alternative to SRBC is to couple the finite element and boundary element methods as described in Bendali & Fares (2008) and Johnson & Nédélec (1980). This is often done when the boundary element method encounters difficulties to take local heterogeneities into account. In this method, the exterior problem is treated by an integral formulation and the interior solution is approximated by FEM. The integral equations can then be seen as exact boundary conditions on the prolate spheroidal.

Following Bendali & Fares (2008) and Johnson & Nédélec (1980), the problem can be reduced to the following linear system

$$\begin{bmatrix} A_{\mathrm{II}} & A_{\mathrm{I}\Gamma} & 0 \\ A_{\Gamma\mathrm{I}} & A_{\Gamma\Gamma} & M_\Gamma \\ 0 & C & -S \end{bmatrix} \begin{bmatrix} u_{\mathrm{I}} \\ u_\Gamma \\ p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -b \end{bmatrix} \tag{4.1}$$

with

$$C = -\left(\tfrac{1}{2}M_\Gamma + K\right), \quad S = -V, \quad b = u^{\text{inc}}. \tag{4.2}$$

and

$$u = \begin{bmatrix} u_{\text{I}} \\ u_\Gamma \end{bmatrix} \tag{4.3}$$

where u_{I} stands for the interior nodal values of u and $u_\Gamma = u|_\Gamma$. Here, V and K are the single and double layers potentials and M_Γ denotes the mass matrix. The matrix blocks A_{II}, $A_{\text{I}\Gamma}$, $A_{\Gamma\text{I}}$ and $A_{\Gamma\Gamma}$ are sparse, while C and S are dense.

One way to avoid handling both sparse and dense blocks simultaneously is to use block Gaussian elimination to express (4.1) as follows

$$\begin{bmatrix} A_{\text{II}} & A_{\text{I}\Gamma} \\ A_{\Gamma\text{I}} & A_{\Gamma\Gamma} \end{bmatrix} \begin{bmatrix} u_{\text{I}} \\ u_\Gamma \end{bmatrix} = \begin{bmatrix} 0 \\ -M_\Gamma p \end{bmatrix} \tag{4.4}$$

$$\left(S + \begin{bmatrix} 0 & C \end{bmatrix} \begin{bmatrix} A_{\text{II}} & A_{\text{I}\Gamma} \\ A_{\Gamma\text{I}} & A_{\Gamma\Gamma} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ M_\Gamma \end{bmatrix} \right) p = b \tag{4.5}$$

However, despite its attractive ability to deal with the dense and the sparse blocks separately, the Schur complement technique presents some drawbacks when used in this context. Its construction can be excessively expensive both in memory storage and computational time, as we will present in Section 4.3.3.

4.2.3 *Standard solution.* In the following we introduce a boundary value problem which is widely used to compute the scattered field. It involves the classical outgoing Fourier Robin condition on the external boundary. We denote by u_1 the corresponding standard solution. It reads:

$$(\text{ABC}_1) \; : \; \begin{cases} \Delta u_1 + k^2 u_1 = 0 & \text{in D}, \\ \dfrac{\partial u_1}{\partial \mathbf{n}} = -\dfrac{\partial u_{\text{inc}}}{\partial \mathbf{n}}(\mathbf{x}) & \text{on } \partial\Omega, \\ \dfrac{\partial u_1}{\partial \mathbf{n}} - iku_1 \; = \; 0 & \text{on } \Gamma. \end{cases} \tag{4.6}$$

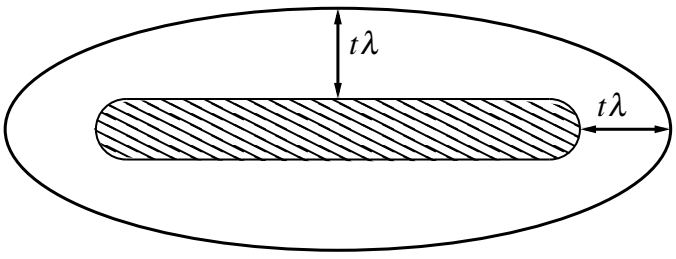

FIG. 3. Computational domain D delimited by a prolate spheroid (cut in the (x,z)-plane),$t \in \mathbb{R}+$

4.2.4 *Error estimate.* The acoustic scattered field pattern is defined by

$$u_\infty(\hat{x}) = \frac{1}{4\pi}\int_\Gamma \left(u(y)+u^{inc}(y)\right)\mathrm{e}^{-ik\hat{x}\cdot y}ds_y$$

for $\hat{x} = (\cos\varphi\sin\theta, \sin\varphi\sin\theta, \cos\theta) \in S$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \varphi \leqslant 2\pi$ and with $S = \{x \in \mathbb{R}^3 : \|x\| = 1\}$ being the unit sphere.

To measure the error of our method, we discretize the unit sphere S using 90 uniform points in the direction of θ and φ with $\theta_i = \frac{i}{90}\pi, i \in \{0,..,90\}$, $\varphi_j = \frac{j}{45}\pi, j \in \{0,..,90\}$. We thus obtain the points $\hat{x}_\ell$, $\ell = (i,j)$, that are uniformly distributed nodes on S.

We define the error ε on the surface of the radiator by

$$\varepsilon = \sqrt{\frac{\sum_\ell |u_\infty(\hat{x}_\ell) - u_\infty^{ref}(\hat{x}_\ell)|^2}{\sum_\ell |u_\infty^{ref}(\hat{x}_\ell)|^2}}. \tag{4.7}$$

$u_\infty^{ref}(\hat{x}_\ell)$ denotes the reference solution defined in Section 4.2.1. The scattered field pattern $u_\infty(\hat{x}_\ell)$ is computed using either the newly proposed method or, for comparison purposes, the standard one defined in Section 4.2.3.

4.3 *Numerical experiments*

In this section, we study the influence of the mesh size and the size of the computational domain on the error. Furthermore, we comment on the gain of memory.

4.3.1 *Influence of the mesh size on the error.* The computational domain D (Figure 3) is a prolate spheroid with $a = 3.0$, $b = 1.5$, $t = 1.0$. We choose the mesh step $h \in \{\lambda/10, \lambda/15, \lambda/20\}$. The numbers of corresponding elements and degrees of freedom are given in Table 1.

Table 1. Errors ε for different mesh sizes. We also report the number of tetrahedra n_{tetra} and the number of unknowns n_{dof} for each h.

$h = \lambda/10$

N	ABC_1	1	3	5	7	9	11	13	15
$Err(\%)$	12.8	12.8	12.4	10.6	5.8	3.4	0.8	0.7	0.6
$n_{tetra} \approx 23\cdot 10^6$; $n_{dof} \approx 4\cdot 10^6$									

$h = \lambda/15$

N	ABC_1	1	3	5	7	9	11	13	15
$Err(\%)$	12.9	12.8	12.5	10.7	5.9	3.4	0.7	0.5	0.4
$n_{tetra} \approx 51\cdot 10^6$; $n_{dof} \approx 8\cdot 10^6$									

$h = \lambda/20$

N	ABC_1	1	3	5	7	9	11	13	15
$Err(\%)$	12.9	12.8	12.5	10.7	5.9	3.5	0.6	0.37	0.36
$n_{tetra} \approx 59\cdot 10^6$; $n_{dof} \approx 9\cdot 10^6$									

The results for the different values of h and N are presented in Table 1 and Figure 4. The dotted lines in Figure 4 correspond to the error of the standard solution (4.6) and shall serve as a reference. For each mesh parameter h, the standard solution has the same bad accuracy (about 12.9%) as the one obtained with SRBC with $N = 1$. The gain of accuracy becomes visible from $N \geqslant 7$ and it is obvious for $N \geqslant 11$ and any h. The numerical results thus suggest that already a value of $N = 11$ provides a highly accurate solution.
For constant N and decreasing h, we can see that the error is almost constant. However, the number of degrees of freedom almost triples from $h = \lambda/20$ to $h = \lambda/10$. The choice of $h = \lambda/10$ is thus a good compromise, as any smaller h increases the computational cost while failing to provide a significantly smaller error. It suggests that a good SRBC is more important than refining the mesh.

Table 2. Errors ε for different sizes of D. We also report the number of tetrahedra n_{tetra} and the number of unknowns n_{dof} for each h.

$t = 1/4$

N	ABC_1	1	3	5	7	9	11	13
$Err(\%)$	24.0	23.9	20.4	17.4	4.8	2.9	0.5	0.4
$a = 2.75$ $b = 1$				$n_{tetra} \approx 3 \cdot 10^6$; $n_{dof} \approx 5 \cdot 10^5$				

$t = 1/2$

N	ABC_1	1	3	5	7	9	11	13
$Err(\%)$	22.9	22.9	22.1	15.7	9.6	4.7	1.0	0.5
$a = 3.$ $b = 1$				$n_{tetra} \approx 4 \cdot 10^6$; $n_{dof} \approx 6 \cdot 10^5$				

$t = 1$

N	ABC_1	1	3	5	7	9	11	13
$Err(\%)$	12.2	12.2	11.9	7.6	7.0	3.3	1.5	0.3
$a = 3.5$ $b = 1.5$				$n_{tetra} \approx 13 \cdot 10^6$; $n_{dof} \approx 2 \cdot 10^6$				

$t = 3/2$

N	ABC_1	1	3	5	7	9	11	13
$Err(\%)$	6.7	6.7	6.5	4.7	3.7	2.5	1.7	0.5
$a = 4.$ $b = 2$				$n_{tet} \approx 29 \cdot 10^6$; $n_{dof} \approx 5 \cdot 10^6$				

4.3.2 *Influence of the size of the computational domain on the error.* For the computation, we keep a constant mesh size of $h = \lambda/10$. The size of the computational domain evolves with the parameter $t\lambda$ with $t \in \{\frac{1}{4}, \frac{1}{2}, 1, \frac{3}{2}\}$, as displayed in Figure 3. The number of tetrahedra and degrees of freedom are presented in Table 2.

It is worth noting that the error reaches a plateau, which indicates that we can get a good level of accuracy by increasing the value of N and keeping t small.

As the matrix size increases for larger domains, the values $t = \frac{1}{4}$ and $N = 13$ seem to be a pratical choice.

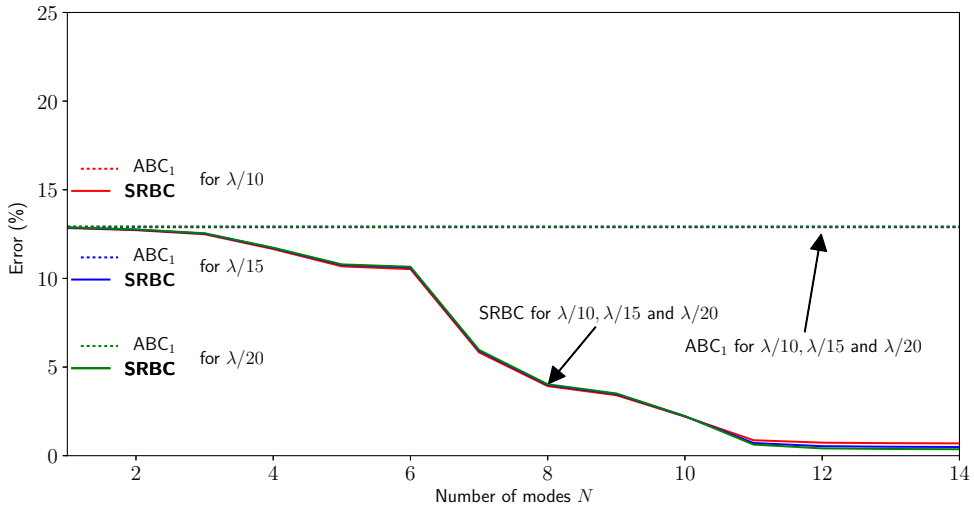

FIG. 4. Errors for mesh sizes $h = \lambda/10, \lambda/15, \lambda/20$

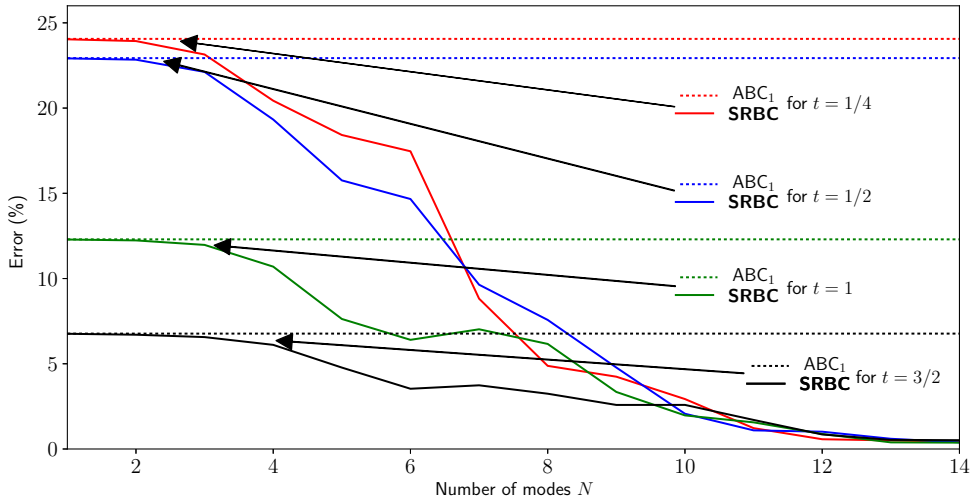

FIG. 5. Variation of computational domain.

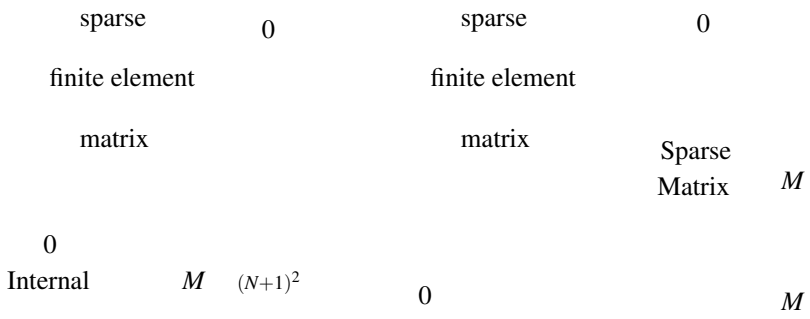

FIG. 6. Matrix structure for the SRBC and FEM-BEM methods

4.3.3 *Memory consumption assessment.* In this section, we compare the memory consumption of the coupled FEM-BEM of Section 4.2.2 with the one of the SRBC discrete formulation. First, it is worth noting that both methods involve the same finite element matrix as depicted in Figure 6.

- The size of the block corresponding to the BEM depends quadratically on the number M of nodes on the surface of the prolate spheroid, that is $2M^2$.

- The size of the dense matrix corresponding to the SRBC is $(N+1)^4+2M(N+1)^2$, $(N+1)^2$ being the number of prolate spheroidal functions.

In the following experiments, the mesh size is $h=\lambda/10$ and the computational domain is parameterized by $t\lambda$ and $t\in\{\frac{1}{2},1,\frac{3}{2}\}$. Table 3 displays the numbers of degrees of freedom for the sparse finite element matrix, the number M of nodes on the spheroid as well as the number nnz of non zero coefficients in the LU decomposition.

Table 4 displays the memory consumption for each case described in Table 3. In particular, we illustrate the percentage of the memory consumption of both the integral equation and SRBC solution with respect to the LU factorization of the sparse finite element matrix. First, the numbers show clearly that BEM requires a significantly larger amount of memory than the SRBC. Second, while BEM adds between 28.5% and 39.5% of additional memory to the sparse LU factorization, the SRBC formulation only adds between 1.875% and 3.125%. With respect to the already significant memory consumption of the sparse LU factorization, the SRBC method adds only a negligible computational overcost and is clearly advantageous.

Table 3. Statistics for each mesh

	n_{tetra}	M	N_{dof}	nnz	$(N+1)^2$
Mesh1	3835537	27664	616296	5107080	144
Mesh2	13271398	49085	2091984	17515738	144
Mesh3	29162092	99416	4588666	38458761	144

Table 4. Memory consumption

	LU Sparse	BEM	% BEM	SRBC	% SRBC
Mesh1	64 Go	24 Go	37.5%	2 Go	3.125%
Mesh2	280 Go	80 Go	28.5%	7 Go	2.5%
Mesh3	800 Go	316 Go	39.5%	15 Go	1.875%

Conclusion

The Boundary Element Method generates an accurate and efficient solution of scattering problems. However, it requires a coupling with volume finite elements that are able to treat possible heterogeneities of the scatterer. In that case, the discrete system is represented by a dense matrix which makes the method highly memory consuming.

Another technique consists in truncating the propagation domain by introducing an artificial boundary which fully surrounds the scatterer. The problem is then reduced to a mixed problem set in a bounded domain and its computational complexity depends on the boundary condition set on the external artificial boundary. The simplest idea is based upon the Fourier-Robin condition which is easy to implement and leads to a sparse matrix. In this case, the accuracy of the solution computed inside the truncated domain depends on the distance of the artificial boundary to the scatterer. For instance, the numerical error is below 8% for a distance of 1.5 wavelengths in between the scatterer and the external boundary.

In this paper, we propose a new condition based on the separation of variables, which leads to the exact characterization of the radiation operator on a basis of spheroidal functions. In practice, the series-like representation is truncated to end up with conditions that can be implemented easily. Although composed of a finite number of terms, the resulting condition is nonlocal but fortunately, it does not require to invert a dense matrix thanks to a suitable algebraic treatment involving the Sherman-Morrisson algorithm.

The numerical studies performed display that the numerical error becomes negligible for a manageable number of basis functions (around $N = 11$). Moreover, the discretization step h has only a slight influence on the numerical error as soon as it respects the rule of 10 points per wavelength. It is also worth noting that the error barely depends on the distance of the artificial boundary to the obstacle. These properties prove that the new condition, compared with a classical radiation condition set on a sphere, contributes to reduce the numerical costs.

Finally, we remark that thanks to the Sherman-Morrison approach, the proposed SRBC method combines the low cost of the ABC_1 method with the accuracy of a BEM discretization.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1964) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, ninth Dover printing, tenth GPO printing edn. New York: Dover.

AMESTOY, P. R., DUFF, I. S., KOSTER, J., & L'ÉXCELLENT., J.-Y. (2001) A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM Journal of Matrix Analysis and Applications,*, **23**, 15–41. http://mumps.enseeiht.fr/apo/MUMPS/.

ANTOINE, X., BARUCQ, H. & BENDALI, A. (1999) Bayliss–Turkel-like radiation conditions on surfaces of arbitrary shape. *Journal of Mathematical Analysis and Applications*, **229**, 184–211.

BARUCQ, H., DJELLOULI, R. & SAINT-GUIRONS, A.-G. (2007) Construction and performance analysis of local DtN absorbing boundary conditions for exterior Helmholtz problems. Part II : Prolate spheroid boundaries. *Research Report RR-6395*.

BARUCQ, H., DJELLOULI, R. & SAINT-GUIRONS, A.-G. (2009a) Construction and performance assesment of new local DtN conditions for elongated obstacles. *Applied Numerical Mathematics*, **59 (7)**, 1467–1498.

BARUCQ, H., DJELLOULI, R. & SAINT-GUIRONS, A.-G. (2009b) Performance assessment of a new class of local absorbing boundary conditions for elliptical- and prolate spheroidal-shaped boundaries. *Applied Numerical Mathematics*, **59**, 1467–1498.

BARUCQ, H., DUPOUY ST-GUIRONS, A.-G. & TORDEUX, S. (2012) Non-reflecting boundary condition on ellipsoidal boundary. *Numerical Analysis and Applications*, **5**.

BAYLISS, A., GUNZBURGER, M. & TURKEL, E. (1982) Boundary conditions for the numerical solution of elliptic equations in exterior regions. *SIAM Journal on Applied Mathematics*, **42**, 430–451.

BENDALI, A. (1984) Approximation par éléments finis de surface de problèmes de diffraction des ondes électromagnétiques. *Ph.D. thesis*, Université Pierre-et-Marie-Curie - Paris VI.

BENDALI, A. & FARES, M. (2008) *Boundary integral equations methods in acoustic scattering*. Computational Methods for Acoustics Problems, Saxe-Coburg edn.

BURN, A. L. V. & BOISVERT, J. E. (2002) Accurate calculation of prolate spheroidal radial functions of the first kind and their first derivatives. *Quarterly of Applied Mathematics*, **60**, 589–599.

BURN, A. L. V. & BOISVERT, J. E. (2015) Oblate and prolate spheroidal functions.

CLAEYS, X. (2008) Analyse asymptotique et numérique de la diffraction d'ondes par des films minces. *Ph.D. thesis*, Université de Versailles Saint-Quentin-en-Yvelines.

DAUTRAY, R. & LIONS, J.-L. (1988) *Mathematical analysis and numerical methods for science and technology*, vol. 5. Paris: Masson.

ENGQUIST, B. & MAJDA, A. (1977) Absorbing boundary conditions for the numerical simulation of waves. *Mathematics of Computation*, **31**, 629–651.

ENGQUIST, B. & MAJDA, A. (1979) Radiation boundary conditions for acoustic and elastic wave calculations. *Communications on Pure and Applied Mathematics*, **32 (3)**, 314–358.

FLAMMER, C. (1957) *Spheroidal wave functions*. Stanford, California: Stanford University Press.

GIVOLI, D. & KELLER, J. (1990) Nonreflecting boundary conditions for elastic waves. *Wave Motion*, **12**, 261–279.

GROTE, M. J. (1995) Nonreflecting boundary conditions. *Ph.D. thesis*, Stanford, CA, USA.

HARARI, I. & HUGHES, T. J. (1992a) Analysis of continuous formulations underlying the computation of time-harmonic acoustics in exterior domains. *Computer Methods in Applied Mechanics and Engineering*, **97**, 103 – 124.

HARARI, I. & HUGHES, T. J. R. (1992b) A cost comparison of boundary element and finite element methods for problems of time-harmonic acoustics. *Computer Methods in Applied Mechanics and Engineering*, **97**, 77–102.

HETMANIUK, U. & FARHAT, C. (2003) A fictitious domain decomposition method for the solution of partially axisymmetric acoustic scattering problems. Part 2: Neumann boundary conditions. *International Journal for Numerical Methods in Engineering*, **58**, 63–81.

JOHNSON, C. & NÉDÉLEC, J.-C. (1980) On the coupling of boundary integral and finite element methods. *Mathematics of Computation*, **35**, 1063–1079.

KELLER, J. B. & GIVOLI, D. (1989) Exact non-reflecting boundary conditions. *Journal of Computational Physics*, **82**, 172–192.

LEBEDEV, N. N. & SILVERMAN, R. A. (1972) *Special functions and their applications*. New York: Dover.

LENOIR, M. & TOUNSI, A. (1988) The localized finite element method and its application to the two-dimensional sea-keeping problem. *SIAM Journal on Numerical Analysis*, **25**, 729–752.

LIONS, J.-L. & MAGENES, E. (1968) *Problèmes aux limites non homogènes et applications - Volume 1*. Paris: Dunod.

MCLEAN, W. (2000) *Strongly Elliptic Systems and Boundary Integral Equations*. Cambridge, UK: Cambridge University Press.

MEDVINSKY, M., TURKEL, E. & HETMANIUK, U. (2008) Local absorbing boundary conditions for elliptical shaped boundaries. *Journal of Computational Physics*, **227**, 8254–8267.

PROTTER, M. H. (1960) Unique continuation for elliptic equations. *Transactions of the American Mathematical Society*, **95**, 81–91.

REINER, R., DJELLOULI, R. & HARARI, I. (2006) The performance of local absorbing boundary conditions for acoustic scattering from elliptical shapes. *Computer Methods in Applied Mechanics and Engineering*, **195**, 3622–3665.

SAINT-GUIRONS, A.-G. (2008) Construction et analyse de conditions absorbantes de type Dirichlet-to-Neumann pour des frontières ellipsoidales. *Ph.D. thesis*, Université de Pau et des Pays de l'Adour.

SHERMAN, J. & MORRISON, W. J. (1950) Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Annals of Mathematical Statistics*, **21**, 124–127.

ŚMIGAJ, W., ARRIDGE, S., BETCKE, T., PHILLIPS, J. & SCHWEIGER., M. (2015) Solving boundary integral problems with BEM++. *ACM Transactions on Mathematical Software*, **41(2)**, 6:16:40. http://www.bempp.org/docs.html.

WILCOX, C. H. (1984) *Scattering theory for diffraction gratings*. New York: Springer-Verlag.

ZARMI, A. & TURKEL, E. (2013) A general approach for high order absorbing boundary conditions for the Helmholtz equation. *Journal of Computational Physics*, **242**, 387–404.