



Numerical Simulation of 3D Electromagnetic Scattering by Algebraic Fictitious Domain Method

Alexandre Bespalov

► To cite this version:

Alexandre Bespalov. Numerical Simulation of 3D Electromagnetic Scattering by Algebraic Fictitious Domain Method. [Research Report] RR-2729, INRIA. 1995. inria-00073965

HAL Id: inria-00073965

<https://inria.hal.science/inria-00073965>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

***Numerical Simulation of 3D Electromagnetic
Scattering by Algebraic Fictitious Domain
Method.***

Alexandre Bespalov

N° 2729

November 1995

PROGRAMME 6

Calcul scientifique,

modélisation

et logiciel numérique

 ***rapport
de recherche***

1995

Numerical Simulation of 3D Electromagnetic Scattering by Algebraic Fictitious Domain Method.

Alexandre Bespalov *

Programme 6 — Calcul scientifique, modélisation et logiciel numérique
Projet MENUSIN

Rapport de recherche n° 2729 — November 1995 — 43 pages

Abstract: New variants of algebraic fictitious domain method are proposed for solution of the 3D Helmholtz and Maxwell equations in unbounded domains with the Sommerfeld radiation condition at infinity. They are based on:

- the use of an infinite uniform Cartesian mesh (maybe, locally fitted to an obstacle) for finite-difference or finite-element approximation;
- nonsymmetric version of fictitious domain method for solution of a resulting mesh system;
- calculation of the partial solution during the iterative process via summation of mesh Green functions with corresponding weights, using a fast algorithm;
- a special way of construction of the approximate mesh Green function satisfying the radiation condition.

New ways of optimization of the fictitious domain method are also proposed. Results of numerical experiments are presented.

Key-words: fictitious domain method, diffraction problem, partial solution, mesh Green function, fast summation algorithm.

(Résumé : tsvp)

*INRIA, Domaine de Voluceau - Rocquencourt - B.P. 105 - Le Chesnay Cedex (France)

email: Alexander.Bespalov@inria.fr

This work has been done in 1994-1995 at INRIA. Special thanks are due to Profs. P. Le Tallec, J. Périaux and Dr. P. Nepomiashtch for their permanent support of the activity.

Simulation Numérique de la diffraction électromagnétique 3D par une méthode algébrique de domaines fictifs.

Résumé : On propose de nouvelles variantes de la méthode algébrique de domaines fictifs pour la résolution des équations de Helmholtz et de Maxwell 3D dans des domaines infinis avec la condition de radiation de Sommerfeld à l'infini. Elles sont basées sur:

- l'utilisation d'un maillage cartésien uniforme infini (s'ajustant éventuellement à un obstacle) pour des approximations par différences finies ou éléments finis;
- une version non symétrique de la méthode de domaines fictifs pour la solution du système discrétisé;
- le calcul de la solution partielle pendant le processus itératif via la sommation des fonctions de Green discrètes avec poids correspondant, utilisant un algorithme rapide;
- un moyen particulier de construction de la fonction de Green approchée satisfaisant la condition de radiation.

On propose également de nouvelles approches d'optimisation de la méthode de domaines fictifs. Les résultats des tests numériques terminent ce rapport.

Mots-clé : méthode de domaines fictifs, problème de diffraction, solution partielle, fonction de Green discrète, algorithme de sommation rapide.

1 Introduction

Calculation of a wave, acoustic or electromagnetic, scattered by an obstacle is a problem of great practical importance. It is also called “diffraction problem”. It can be two- or three-dimensional. The diffraction problem is mathematically described by Helmholtz wave equation (in a scalar case) or by Maxwell equations (in a vector 3D case) in an unbounded domain with the Sommerfeld radiation condition at infinity. Various aspects of this problem and methods of its solution were considered in [2, 3, 4, 8, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 25, 26, 28, 33, 34] and many others.

Application of fictitious domain method (in several treatments and under several names) for solution of the problem was considered in [2, 3, 4, 8, 12, 13, 19, 25, 26, 28]. It turned out to be an effective tool for this purpose. It can be said that this method combines advantages both of methods solving boundary integral equations (all handled vectors are non-zeroth at boundary nodes only) and of methods solving volume differential equations (sparseness of arising matrices).

This article continues the development of fictitious domain method for solution of the considered problem undertaken in [8, 12, 25, 26]. In that articles discretization of the problem was carried out by finite-difference [25] or by finite-element [8, 12, 26] method on a spherical (polar in 2D) mesh. This choice of a mesh type allows to reduce the problem in an unbounded domain to a problem inside a minimal sphere (circle in 2D) described around an obstacle and also to calculate an asymptotic form of a scattered wave in a very simple and precise way.

However, discretization on a spherical mesh has several obvious drawbacks especially important for the vector problem in 3D case. Among them are the following ones:

1. Such a mesh is too fine nearby the coordinate axis. This gives rise to a lot of excessive nodes not improving an approximation.
2. Finite-element approximation of the 3D Maxwell equations on a spherical mesh and separation of variables in a resulting discrete problem are very complicated for a practical implementation.
3. Local fitting of a 3D spherical mesh is very difficult for a practical implementation as well, especially nearby the coordinate axis.

For these reasons the author undertook an attempt to develop a version of fictitious domain method for numerical solution of diffraction problems using a Cartesian

mesh. A main difficulty here consists in an approximation of the Sommerfeld radiation condition which would be correct, precise, economical¹ and suitable for the fictitious domain method.

This approximation could be carried out in the way proposed and considered in [17, 21, 22] and others. In this approach, a surface Γ described around an obstacle (“an artificial infinity”) is introduced, and then the exact Sommerfeld condition at infinity is approximated by a special local absorbing boundary condition on Γ with the use of pseudo-differential operators. For the case of the use of a Cartesian mesh, it is natural to choose Γ as a rectangle boundary.

In principle, this approach is workable and can be used in the context of fictitious domain method [2, 3, 4, 19]. However, to get a sufficient precision it is necessary to choose Γ rather far from the obstacle, resulting in a considerable increase of number of unknowns in an algebraic problem (especially in 3D case). Besides that, most of such approximations do not allow to separate variables inside the rectangle (which is necessary for fictitious domain method), so special expensive tricks should be used to override this. For these reasons such approximation of the radiation condition is not used in this article.

It is proposed here to approximate the problem on an infinite uniform Cartesian mesh (maybe, locally fitted to an obstacle boundary) by finite-difference or finite-element method, without introducing an artificial infinity. For approximation of the 3D Maxwell equation a special approach using staggered meshes is considered.

For solving a resulting infinite mesh problem, it is proposed to apply fictitious domain method, i.e. iterative process with preconditioner \mathcal{B}^{-1} , where mesh operator \mathcal{B} corresponds to the same approximation of the same differential operator on the infinite uniform Cartesian mesh in the whole “empty” space.

To be applied, the algebraic fictitious domain method requires so-called “enlargement” of a solved problem. This enlargement is arbitrary (within some equivalency conditions), so its choice is to be aimed at speeding the iterative convergence up. Besides that, an additional preconditioner H can be introduced into the iterative process. Thus, the problems of optimal choice of the enlargement and of the preconditioner arise. They are also considered in the article, and new possible approaches to their solving are proposed.

For implementation of the method, it is necessary to calculate mesh function $v^h = \mathcal{B}^{-1}\xi^h$ in each iterative step, in such a way that v^h satisfies some mesh radiation condition at infinity (the last one should approximate exact radiation condition). This algebraic problem is partial [23, 24, 27, 30], i.e. a right-hand side ξ^h can be

¹From the viewpoint of the algebraic dimension of a resulting problem.

non-zeroth only at boundary and near-boundary nodes (with respect to an obstacle), and solution $\mathcal{B}^{-1}\xi^h$ should be found at the same nodes only.

For this purpose, it is proposed to use the following obvious formulas:

$$(\mathcal{B}^{-1}\xi^h)_{i,j} = \sum_{l,m} \xi_{lm}^h G_{i-l,j-m}^h \quad \text{in 2D case,}$$

$$(\mathcal{B}^{-1}\xi^h)_{i,j,k} = \sum_{l,m,n} \xi_{lmn}^h G_{i-l,j-m,k-n}^h \quad \text{in 3D case,}$$

where G^h is the mesh Green function of the operator \mathcal{B} . These formulas are valid because all nodes of an infinite uniform Cartesian mesh are equivalent to each other. For the summation it is proposed to use a fast method using the fast Fourier transformation algorithm.

The Green function G^h (satisfying an approximate radiation condition at infinity) is constructed by means of Fourier analysis. Its value at each mesh node is represented as an integral of singular highly-oscillating function, but an effective method for its calculation has been worked out.

For reducing arithmetical expenses and required computer memory, it is also proposed to replace exact mesh Green function G^h in the algorithm by its approximation \tilde{G}^h . The latter one is built on a base of the exact Green function of the differential operator of the considered problem. A formula for the asymptotic form of a scattered wave (RCS) is derived using the same trick.

Implementation of the resulting algorithm in 3D case for a fixed problem requires $O(h^{-3} \log h^{-1})$ arithmetical operations per each iterative step and $O(h^{-2})$ computer memory locations (where h is a stepsize), both for scalar and vector cases.

Results of numerical experiments carried out in 3D case for verification and testifying of the approach are presented.

2 Statement of the problems and discretization

2.1 The scalar problem

Scattering of a stationary longitudinal wave v_I by an obstacle Ω in a homogeneous medium is described by the scalar Helmholtz wave equation with the Sommerfeld (radiation) condition at infinity:

$$\begin{aligned} -\Delta u - \kappa^2 u &= 0 & \text{in } \mathbf{R}^m \setminus \bar{\Omega}, \\ \mathcal{L}u &= -\mathcal{L}v_I & \text{on } \partial\Omega, \\ \frac{\partial u}{\partial r} - i\kappa u &= O(r^{-\frac{m+1}{2}}) & \text{for } r \rightarrow \infty. \end{aligned} \tag{1}$$

Here $m = 2$ or $m = 3$ is the space dimension, i is the imaginary unity, Δ is the “scalar” Laplace operator:

$$\Delta = \sum_{i=1}^m \frac{\partial^2}{\partial x_i^2}, \quad (2)$$

$\kappa = \text{const} > 0$, $r = |\mathbf{x}|$, Ω is a bounded domain in \mathbf{R}^m with a piecewise-smooth boundary $\partial\Omega$ such that $\mathbf{R}^m \setminus \bar{\Omega}$ is a connected region, v_I is an incident wave (usually it is taken as $v_I = \exp(i\boldsymbol{\kappa} \cdot \mathbf{x})$, $|\boldsymbol{\kappa}| = \kappa$), u is a scattered wave, v_I and u are complex-valued scalar functions of \mathbf{x} , \mathcal{L} is some known linear boundary condition operator. We also introduce the wavelength $\lambda = 2\pi/\kappa$.

Results stating the existence and uniqueness of the solution to problem (1) can be found in [31, 34].

It is well known that a solution of problem (1) has the following asymptotic form:

$$\begin{aligned} u(r, \theta) &= A(\theta) \frac{e^{i\kappa r}}{\sqrt{r}} + O(r^{-\frac{3}{2}}) \quad \text{for } r \rightarrow \infty, \quad m = 2, \\ u(r, \theta, \varphi) &= A(\theta, \varphi) \frac{e^{i\kappa r}}{r} + O(r^{-2}) \quad \text{for } r \rightarrow \infty, \quad m = 3, \end{aligned} \quad (3)$$

where (r, θ) is polar and (r, θ, φ) is the spherical system of coordinates. Condition (3) is equivalent to the Sommerfeld one.

Problem (1) corresponds to scattering of acoustic waves. Besides, this 2D problem approximately describes scattering of TE- and TM-polarized electromagnetic waves by a 3D cylindrical obstacle with a constant cross-section Ω (if its length in z -direction is much greater than diameter of Ω and λ), see [34].

2.2 The vector (electromagnetic) problem

Scattering of a stationary electromagnetic wave \mathbf{E}_I in vacuum by an ideally conductive obstacle Ω is described by the following 3D problem:

$$\begin{aligned} \nabla \times \nabla \times \mathbf{E}_s - \kappa^2 \mathbf{E}_s &= 0 && \text{in } \mathbf{R}^3 \setminus \bar{\Omega}, \\ \mathbf{E}_s^\tau &= -\mathbf{E}_I^\tau && \text{on } \partial\Omega, \\ \frac{\partial \mathbf{E}_s}{\partial r} - i\kappa \mathbf{E}_s &= O(r^{-2}) && \text{for } r \rightarrow \infty. \end{aligned} \quad (4)$$

Here $\nabla \times \equiv \text{curl} \equiv \text{rot}$, \mathbf{E}_I and \mathbf{E}_s are electric fields of incident and scattered waves correspondingly, both are complex-valued vector functions of \mathbf{x} , \mathbf{E}^τ is a tangential component of \mathbf{E} .

Equation in (4) (the first line) is simply derived from the Maxwell equations in vacuum [34]. We note that an incident wave \mathbf{E}_I should also satisfy it. The scattering problem can also be formulated in terms of magnetic field \mathbf{H} but we are not considering that variant.

For problem (4), the existence and uniqueness theorem can also be formulated and proved [31, 34].

Electric fields \mathbf{E}_I and \mathbf{E}_s should also satisfy the divergency-free condition

$$\nabla \cdot \mathbf{E} \equiv 0 \text{ in } \mathbf{R}^3 \setminus \bar{\Omega}. \quad (5)$$

Applying operator $\nabla \cdot$ to the both sides of equation in (4) and using the well-known identity

$$\nabla \cdot (\nabla \times) \equiv 0, \quad (6)$$

we get that a solution \mathbf{E}_s of (4) satisfies (5) automatically.

Incident wave \mathbf{E}_I is taken in several ways. The simplest one is to choose it as a plane-polarized wave propagating in a direction $\boldsymbol{\kappa}$:

$$\mathbf{E}_I = \mathbf{E}_0 \exp(i\boldsymbol{\kappa} \cdot \mathbf{x}), \quad |\boldsymbol{\kappa}| = \kappa, \quad (7)$$

where \mathbf{E}_0 is a vector constant such that $\mathbf{E}_0 \cdot \boldsymbol{\kappa} = 0$. It is easy to check that \mathbf{E}_I from (7) truly satisfies both the equation from (4) and condition (5).

It is well known (see e.g. [34]) that a solution of problem (4) has the following asymptotic form:

$$\begin{aligned} E_{s,\theta}(r, \theta, \varphi) &= A_\theta(\theta, \varphi) \frac{e^{i\kappa r}}{r} + O(r^{-2}) & \text{for } r \rightarrow \infty, \\ E_{s,\varphi}(r, \theta, \varphi) &= A_\varphi(\theta, \varphi) \frac{e^{i\kappa r}}{r} + O(r^{-2}) & \text{for } r \rightarrow \infty, \\ E_{s,r}(r, \theta, \varphi) &= O(r^{-2}) & \text{for } r \rightarrow \infty, \end{aligned} \quad (8)$$

where (r, θ, φ) is the spherical system of coordinates. That is, when going to infinity, radial component $E_{s,r}$ of a scattered field is decreasing much faster than its transverse component.

2.3 Transformation of the vector problem

Let us consider a widespread transformation of problem (4). Let us introduce the “vector” Laplace operator:

$$-\Delta \stackrel{\text{def}}{=} \nabla \times \nabla \times - \nabla \nabla \cdot \quad (9)$$

and rewrite (4) as follows:

$$\begin{aligned}
-\Delta \mathbf{E}_s - \kappa^2 \mathbf{E}_s &= 0 & \text{in } \mathbf{R}^3 \setminus \bar{\Omega}, \\
\mathbf{E}_s^\tau &= -\mathbf{E}_I^\tau & \text{on } \partial\Omega, \\
\nabla \cdot \mathbf{E}_s &= 0 & \text{on } \partial\Omega, \\
\frac{\partial \mathbf{E}_s}{\partial r} - i\kappa \mathbf{E}_s &= O(r^{-2}) & \text{for } r \rightarrow \infty.
\end{aligned} \tag{10}$$

Let us prove that problem (10) is equivalent to problem (4). Denoting $w = \nabla \cdot \mathbf{E}_s$ and applying operator $\nabla \cdot$ to the both sides of equation and Sommerfeld condition in (10) we get the following problem for w :

$$\begin{aligned}
-\Delta w - \kappa^2 w &= 0 & \text{in } \mathbf{R}^3 \setminus \bar{\Omega}, \\
w &= 0 & \text{on } \partial\Omega, \\
\frac{\partial w}{\partial r} - i\kappa w &= O(r^{-2}) & \text{for } r \rightarrow \infty.
\end{aligned} \tag{11}$$

It follows from results of [32] that $w \equiv 0$ in $\mathbf{R}^3 \setminus \bar{\Omega}$, thus, the statement has been proved.

Formulation (10) of the vector problem is often more convenient because of the following obvious property of “vector” Laplace operator (9):

$$\Delta \mathbf{E} = (\Delta E_1, \Delta E_2, \Delta E_3)^T, \tag{12}$$

where E_1, E_2, E_3 are Cartesian components of \mathbf{E} . Further we shall use it.

2.4 Approximation of the scalar problem

Let us construct in \mathbf{R}^m an infinite uniform Cartesian mesh:

$$x_i^j = hj, \quad j = \overline{-\infty, +\infty}, \quad i = \overline{1, m}, \tag{13}$$

where $h \ll \lambda$. Rectangles

$$x_i^j \leq x_i \leq x_i^{j+1}, \quad j = \overline{-\infty, +\infty}, \quad i = \overline{1, m},$$

will be referred to as “mesh cells”, their edges – as “mesh edges”, their sides (in 3D case) – as “mesh sides”, their vertices – as “mesh nodes”.

Basing on (13) let us construct a mesh $(\mathbf{R}^m \setminus \bar{\Omega})_h$ in the considered domain. It can be a mesh with the first-order (“staircase”) approximation of $\partial\Omega$ (see [10, 25]) or a mesh locally fitted to $\partial\Omega$ and approximating it with the second order of accuracy [8, 11, 12].

Having constructed a mesh $(\mathbf{R}^m \setminus \bar{\Omega})_h$ we approximate the problem by finite-difference or finite element (Galerkin) method. In the latter case we suppose that bilinear (trilinear in 3D) basic functions are used in rectangular cells. So we get a mesh problem

$$\mathcal{A}u^h = f^h \quad \text{on} \quad (\mathbf{R}^m \setminus \bar{\Omega})_h, \quad (14)$$

where \mathcal{A} is an infinite symmetric mesh operator, u^h is a mesh function defined at mesh nodes, f^h is a right-hand side generated by the right-hand side of the boundary condition in (1) only.

Note that no radiation condition is yet imposed on an approximate solution u^h . We shall do it further.

2.5 Finite-difference approximation of the vector problem

Now let us construct a finite difference approximation of vector problem (4). To do this, the approach described in [10, 29, 35] is used, i.e. approximation of differential operators $\nabla \times$, $\nabla \cdot$, ∇ on uniform staggered grids retaining interrelations of the operators.

Let us take mesh (13) for $m = 3$ and denote $(\mathbf{R}^3 \setminus \bar{\Omega})_h$ a set of mesh cells whose centers belong to $\mathbf{R}^3 \setminus \bar{\Omega}$. It gives a first-order “staircase” approximation of the domain under consideration. A mesh edge or side will be referred to as boundary one if it belongs to both a cell of $(\mathbf{R}^3 \setminus \bar{\Omega})_h$ and to a cell not belonging to $(\mathbf{R}^3 \setminus \bar{\Omega})_h$.

Let us introduce the following notations:

\mathcal{M}_i^E , $i = \overline{1, 3}$, – a set of middle points of all mesh edges of $(\mathbf{R}^3 \setminus \bar{\Omega})_h$ parallel to coordinate direction x_i ;

$$\mathcal{M}^E = \mathcal{M}_1^E \cup \mathcal{M}_2^E \cup \mathcal{M}_3^E;$$

$\overset{\circ}{\mathcal{M}}_i^E$, $i = \overline{1, 3}$, – a subset of points of \mathcal{M}_i^E not belonging to boundary edges of $(\mathbf{R}^3 \setminus \bar{\Omega})_h$;

$$\overset{\circ}{\mathcal{M}}^E = \overset{\circ}{\mathcal{M}}_1^E \cup \overset{\circ}{\mathcal{M}}_2^E \cup \overset{\circ}{\mathcal{M}}_3^E;$$

$$\partial \mathcal{M}_i^E = \mathcal{M}_i^E \setminus \overset{\circ}{\mathcal{M}}_i^E, \quad \partial \mathcal{M}^E = \mathcal{M}^E \setminus \overset{\circ}{\mathcal{M}}^E;$$

\mathcal{M}_i^H , $i = \overline{1, 3}$, – a set of centers of all mesh sides of $(\mathbf{R}^3 \setminus \bar{\Omega})_h$ normal to coordinate direction x_i ;

$$\mathcal{M}^H = \mathcal{M}_1^H \cup \mathcal{M}_2^H \cup \mathcal{M}_3^H;$$

\mathcal{M}_d – a set of all mesh nodes of $(\mathbf{R}^3 \setminus \bar{\Omega})_h$;

$\overset{\circ}{\mathcal{M}}_d$ – a subset of non-boundary nodes of \mathcal{M}_d ;

$\partial\mathcal{M}_d = \mathcal{M}_d \setminus \overset{\circ}{\mathcal{M}}_d$;

\mathcal{E} – a set of “vector” mesh functions \mathbf{E}^h whose i th Cartesian component E_i^h is defined at \mathcal{M}_i^E , $i = \overline{1,3}$;

$\overset{\circ}{\mathcal{E}}$ – a subset of functions of \mathcal{E} equaled to zero at $\partial\mathcal{M}^E$;

\mathcal{H} – a set of “vector” mesh functions \mathbf{H}^h whose i th Cartesian component H_i^h is defined at \mathcal{M}_i^H , $i = \overline{1,3}$;

\mathcal{D} – a set of scalar mesh functions w^h defined at \mathcal{M}_d ;

$\overset{\circ}{\mathcal{D}}$ – a subset of functions of \mathcal{D} equaled to zero at $\partial\mathcal{M}_d$.

Further, let us introduce in the usual way the finite-difference approximations of differential operators $\nabla \times$, $\nabla \cdot$, ∇ :

$$\begin{aligned} (\nabla \times)_E^h &: \mathcal{E} \rightarrow \mathcal{H}; \\ (\nabla \times)_H^h &: \mathcal{H} \rightarrow \overset{\circ}{\mathcal{E}}; \\ (\nabla \cdot)^h &: \mathcal{E} \rightarrow \overset{\circ}{\mathcal{D}}; \\ \nabla^h &: \mathcal{D} \rightarrow \mathcal{E}. \end{aligned} \tag{15}$$

Here defining the operator $(\nabla \cdot)^h$ we put

$$(\nabla \cdot)^h \mathbf{E}^h = 0 \quad \text{at} \quad \partial\mathcal{M}_d \tag{16}$$

according with the corresponding boundary condition in (10).

The following easy-to-check interrelation is valid for the introduced mesh operators:

$$(\nabla \cdot)^h (\nabla \times)_H^h \mathbf{H}^h \equiv 0 \quad \text{at} \quad \overset{\circ}{\mathcal{M}}^E \quad \text{for} \quad \forall \mathbf{H}^h \in \mathcal{H}. \tag{17}$$

It is a mesh analog of (6).

Thereafter we can approximate problem (4):

$$\begin{aligned} (\nabla \times)_H^h (\nabla \times)_E^h \mathbf{E}_s^h - \kappa^2 \mathbf{E}_s^h &= 0 & \text{at} & \quad \overset{\circ}{\mathcal{M}}^E, \\ E_{s,i}^h &= -E_{1,i} & \text{at} & \quad \partial\mathcal{M}_i^E, \quad i = \overline{1,3}. \end{aligned} \tag{18}$$

As in the scalar case, we have imposed no radiation condition yet.

It follows immediately from (17) that for $\kappa \neq 0$ a solution of (18) satisfies the condition

$$(\nabla \cdot)^h \mathbf{E}_s^h \equiv 0 \quad \text{at} \quad \overset{\circ}{\mathcal{D}}, \tag{19}$$

which is a mesh analog of (5).

Let us represent a solution of (18) in the form

$$\begin{aligned} \mathbf{E}_s^h &= \mathring{\mathbf{E}}_s^h - \mathring{\mathbf{E}}_I^h, \quad \text{where} \\ \mathring{E}_{I,i}^h &= \begin{cases} E_{I,i} & \text{at } \partial\mathcal{M}_i^E, \quad i = \overline{1,3}, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (20)$$

It is obvious that

$$\mathring{E}_{s,i}^h = 0 \quad \text{at } \partial\mathcal{M}_i^E, \quad i = \overline{1,3}.$$

Using (20), problem (18) can be rewritten as follows:

$$\begin{aligned} (\nabla \times)_H^h (\nabla \times)_E^h \mathring{\mathbf{E}}_s^h - \kappa^2 \mathring{\mathbf{E}}_s^h &= (\nabla \times)_H^h (\nabla \times)_E^h \mathring{\mathbf{E}}_I^h \quad \text{at } \mathring{\mathcal{M}}^E, \\ \mathring{E}_{s,i}^h &= 0 \quad \text{at } \partial\mathcal{M}_i^E, \quad i = \overline{1,3}. \end{aligned} \quad (21)$$

2.6 Transformation of the finite-difference problem

As in the differential case, let us transform problem (21) to a problem with the mesh “vector” Laplace operator

$$-\Delta_E^h \stackrel{\text{def}}{=} (\nabla \times)_H^h (\nabla \times)_E^h - \nabla^h (\nabla \cdot)^h \quad (22)$$

as follows:

$$\begin{aligned} -\Delta_E^h \mathbf{E}_s^h - \kappa^2 \mathbf{E}_s^h &= (\nabla \times)_H^h (\nabla \times)_E^h \mathring{\mathbf{E}}_I^h \quad \text{at } \mathring{\mathcal{M}}^E, \\ (\mathbf{E}_s^h)_i &= 0 \quad \text{at } \partial\mathcal{M}_i^E, \quad i = \overline{1,3}. \end{aligned} \quad (23)$$

Equivalency of problems (21) and (23) is proved in the same way as in the differential case. That is, denoting $w^h = (\nabla \cdot)^h \mathbf{E}_s^h$, applying operator $(\nabla \cdot)^h$ to the both sides of equation in (10) and taking into account (19) we get the following problem for w^h :

$$\begin{aligned} -(\nabla \cdot)^h \nabla^h w^h - \kappa^2 w^h &= 0 \quad \text{at } \mathring{\mathcal{M}}_d, \\ w^h &= 0 \quad \text{at } \partial\mathcal{M}_d. \end{aligned} \quad (24)$$

It is easy to see that the operator $(\nabla \cdot)^h \nabla^h$ is the seven-point finite-difference approximation of the “scalar” Laplace operator (2). Thus, problem (24) coincides with (14) where the latter one corresponds to the finite-difference approximation and the

zero Dirichlet boundary condition. Suppose that a mesh radiation condition is imposed on \mathbf{E}_s^h in such a way that problem (24) has a unique solution. Then $w^h \equiv 0$ at \mathcal{M}_d .

A reason for constructing problem (23) is the same as in the differential case, i.e. it is more convenient in practice than (21). It is easy to check that the mesh analog of relation (12)

$$\Delta_E^h \mathbf{E}^h = (\Delta_1^h E_1^h, \Delta_2^h E_2^h, \Delta_3^h E_3^h)^T \quad (25)$$

is valid everywhere in \mathcal{M}^E except for some special points (see below). Here each of Δ_i^h , $i = 1, 3$, is the usual seven-point finite-difference approximation of the “scalar” Laplace operator (2) defined on \mathbf{E}_i^h :

$$\begin{aligned} -(\Delta_1^h E_1^h)_{j+\frac{1}{2}, k, l} = h^{-2} (6E_{1, j+\frac{1}{2}, k, l}^h - E_{1, j-\frac{1}{2}, k, l}^h - E_{1, j+\frac{3}{2}, k, l}^h - \\ E_{1, j+\frac{1}{2}, k-1, l}^h - E_{1, j+\frac{1}{2}, k+1, l}^h - E_{1, j+\frac{1}{2}, k, l-1}^h - E_{1, j+\frac{1}{2}, k, l+1}^h), \end{aligned} \quad (26)$$

analogously for Δ_2^h, Δ_3^h . For clarity the half-integer numeration is used here.

Relation (26) for $(\Delta_E^h \mathbf{E}^h)_1$ is valid if all the points $(j - \frac{1}{2}, k, l)$, $(j + \frac{1}{2}, k, l)$, $(j + \frac{3}{2}, k, l)$ belong to \mathcal{M}_1^E . Otherwise the expression for $(\Delta_E^h \mathbf{E}^h)_1$ has such a form as if the Cartesian component E_1^h satisfies the mesh Neumann homogeneous boundary condition on mesh sides normal to the direction x_1 . Let for definiteness $(j + \frac{3}{2}, k, l) \notin \mathcal{M}_1^E$, then the expression for $\Delta_E^h \mathbf{E}^h$ at the point $(j + \frac{1}{2}, k, l)$ corresponds to setting

$$E_{1, j+\frac{1}{2}, k, l}^h = E_{1, j+\frac{3}{2}, k, l}^h$$

and using it for elimination $E_{1, j+\frac{3}{2}, k, l}^h$ from (26).

Of special interest is a case when a mesh node (j, k, l) belongs to the boundary (i.e. to $\partial \mathcal{M}_d$) and there are two non-boundary mesh edges of $(\mathbf{R}^3 \setminus \bar{\Omega})_h$ with this vertex not parallel to each other. Let for definiteness them be $(j + \frac{1}{2}, k, l)$ and $(j, k + \frac{1}{2}, l)$. The expression for $(\Delta_E^h \mathbf{E}^h)_{1, j+\frac{1}{2}, k, l}$ is the following one here:

$$\begin{aligned} -(\Delta_E^h \mathbf{E}^h)_{1, j+\frac{1}{2}, k, l} = h^{-2} (5E_{1, j+\frac{1}{2}, k, l}^h - E_{2, j, k+\frac{1}{2}, l}^h - E_{1, j+\frac{3}{2}, k, l}^h - \\ E_{1, j+\frac{1}{2}, k-1, l}^h - E_{1, j+\frac{1}{2}, k+1, l}^h - E_{1, j+\frac{1}{2}, k, l-1}^h - E_{1, j+\frac{1}{2}, k, l+1}^h) \end{aligned} \quad (27)$$

(supposing that all the points involved into (27) belong to \mathcal{M}^E). The expression for $(\Delta_E^h \mathbf{E}^h)_{2, j, k+\frac{1}{2}, l}$ is analogous.

A sign in front of the distinctive term $(E_{2, j, k+\frac{1}{2}, l}^h$ in (27)) depends on orientations (with respect to the coordinate system) of two rays starting with the node (j, k, l)

along the two considered edges. If both of them have positive or negative direction then the sign is minus, otherwise it is plus.

It is worthy of note that the above-considered special case is the only one where relation (25) is not valid, i.e. where there is a connection between different Cartesian components of \mathbf{E}^h in the mesh operator Δ_E^h .

2.7 Finite-element approximation of the vector problem

To construct a second-order discretization of problem (4) for an arbitrary-shaped obstacle Ω , one can use finite-element Galerkin method. At first glance it can be done in the ordinary way using the piecewise-linear and/or trilinear continuous basic functions for approximation of Cartesian components of the searched vector function \mathbf{E}_s . But it is well known that such approximation possess considerable intrinsic drawbacks. First of them relates to an approximation of the boundary condition on $\partial\Omega$ from (4). Namely, approximating it in the straightforward way we get (generally speaking) the additional parasite condition for the normal component:

$$E_s^n = -E_I^n \quad (28)$$

(i.e. $\mathbf{E}_s = -\mathbf{E}_I$ as a whole). The second drawback: condition (5) is not valid for the mesh solution so it is unclear how to derive correctly a mesh analogue of (23). Thus, it can be said this kind of approximation is “unnatural” for the problem.

To obtain a good finite-element approximation the approach described in [7, 14, 32] can be used. It is based on the use of special piecewise-polynomial vector basic functions whose normal component on boundaries of finite elements is (generally speaking) discontinuous but tangential one is always continuous. It is easy to check that $\nabla \times \Psi^h$ is defined in the generalized sense for such functions Ψ^h , and $\nabla \times \Psi^h$ is locally integrable together with its square. Thus, it is possible to build the conformal finite-element Galerkin approximation of problem (4) using these basic functions.

Basic functions of this (Nédélec) approximation are associated with mesh edges, not nodes. Their coefficients (i.e. mesh unknowns) correspond to the orthogonal projections of electric field onto the edges. It is easy to see that the boundary condition on $\partial\Omega$ being approximated in the obvious straightforward way does not lead to parasite condition (28). It has been also shown in [7, 14, 32] that a mesh solution satisfies a mesh weak analogue of (5); basing on it, the mesh problem corresponding to (4) can be exactly transformed to a mesh analogue of (23) [7].

Suppose that we are going to apply this way of approximation using a rectangular locally fitted mesh (mentioned in Subsection 2.4). In doing so it is natural

and convenient to build up the finite element mesh from tetrahedric cells in near-boundary layer and from regular cubic cells beyond it taking linear Nédélec vector basic functions in tetrahedrons and bilinear-constant ones² [7] in cubes.

In doing so, the problem of tetrahedron-cube matching arises. It can be solved as follows. Let us take any side $ABCD$ of a regular (cubic) cell adjacent to the tetrahedron layer and assume that the trace of the tetrahedric mesh on this side consists of two triangles ABC and ACD . For every lateral edge AB, BC, CD, DA there is the independent coefficient corresponding to the orthogonal projection of electric field onto it (assuming this projection to be constant at the approximation). To get conformality, let us associate with diagonal AC two linear edge basic functions (in tetrahedrons only!) and choose their coefficients not independent but connected with the coefficients of the lateral sides by linear relations providing coincidence of traces of any spanned finite-element function from the “cube” side of $ABCD$ and from its “tetrahedric” side. It is easy to check out that it is truly possible.

Remark. Suppose that we have approximated problem (4) on a non-fitted “staircase” mesh (13) using piecewise bilinear-constant basic functions [7] for the Cartesian components of the searched electric field. Suppose also that in doing this mass lumping was applied. It has been shown in [7] that the resulting algebraic problem coincides with finite-difference problem (18).

3 Fictitious domain method

In this Section we shall present fictitious domain method in the treatment of [1, 8, 10, 11, 12, 13, 23, 24, 25, 28, 30] for solution of problems (14), (23). To simplify the presentation, we are carrying it out mainly for scalar problem (14). Generalization of the approach to solving vector problem (23) is quite obvious.

3.1 General idea

Let us consider the following problem:

$$-\Delta u - \kappa^2 u = 0 \quad \text{in } \mathbf{R}^m \quad (29)$$

and approximate it in the same way on the same grid \mathbf{R}_h^m from (13). We get mesh operator \mathcal{B} on \mathbf{R}_h^m . Suppose that we introduce also an inverse operator \mathcal{B}^{-1} defined

²The corresponding 2D basic functions are linear-constant and look like the simplest “house of cards”.

for finite mesh functions ϕ^h so that $\mathcal{B}^{-1}\phi^h$ satisfies a mesh approximation of the radiation condition (this approximation will be introduced in Section 4). Hereafter a finite mesh function is a function not equaled to zero at finite number of mesh nodes.

Let us enlarge (formally) problem (14) to be solved:

$$\begin{aligned} \hat{\mathcal{A}}\hat{u}^h &= \hat{f}^h \quad \text{on } \mathbf{R}_h^m, \\ \text{where } \hat{\mathcal{A}} &= \begin{bmatrix} \mathcal{A} & \mathcal{A}_{\text{rf}} \\ 0 & \mathcal{A}_{\text{ff}} \end{bmatrix}, \quad \hat{f}^h = \begin{bmatrix} f^h \\ 0 \end{bmatrix}, \quad \hat{u}^h = \begin{bmatrix} u_{\text{r}}^h \\ u_{\text{f}}^h \end{bmatrix}. \end{aligned} \quad (30)$$

It is easy to see that problems (14) and (30) are equivalent to each other if $\mathcal{A}_{\text{rf}} = 0$ or matrix \mathcal{A}_{ff} is non-degenerate³.

For solving system (30) the iterative process is introduced:

$$\begin{aligned} \hat{u}_k^h - \hat{u}_{k-1}^h &= -\tau \mathcal{B}^{-1}(\hat{\mathcal{A}}\hat{u}_{k-1}^h - \hat{f}^h), \quad k = 1, \dots, k_{\text{max}}, \\ \text{where } \hat{u}_0^h &= \mathcal{B}^{-1}\hat{f}^h, \quad \tau = \text{const}, \end{aligned} \quad (31)$$

i.e. the operator \mathcal{B}^{-1} is used as a preconditioner.

Enlargement (30) can be done in many ways (being focussed on upgrading of iterative convergence). To be concrete, we exemplify three of them:

1. $\mathcal{A}_{\text{rf}} = \mathcal{B}_{\text{rf}}, \mathcal{A}_{\text{ff}} = \mathcal{B}_{\text{ff}};$
2. $\mathcal{A}_{\text{rf}} = \mathcal{B}_{\text{rf}}, \mathcal{A}_{\text{ff}}$ corresponds to approximation of the problem:

$$\begin{aligned} -\Delta v - \kappa^2 v &= 0 \quad \text{in } \Omega, \\ \frac{\partial v}{\partial \mathbf{n}} - i\kappa v &= 0 \quad \text{on } \partial\Omega; \end{aligned} \quad (32)$$

3. $\mathcal{A}_{\text{rf}} = 0, \mathcal{A}_{\text{ff}} = 0$ (the zeroth enlargement [23, 24, 30]).

Here the blocks $\mathcal{B}_{\text{rf}}, \mathcal{B}_{\text{ff}}$ are taken from the 2×2 -block partition of \mathcal{B} :

$$\mathcal{B} = \begin{bmatrix} \mathcal{B}_{\text{rr}} & \mathcal{B}_{\text{rf}} \\ \mathcal{B}_{\text{fr}} & \mathcal{B}_{\text{ff}} \end{bmatrix}, \quad (33)$$

which is the same as in (30).

³Note that anyway the condition $u_{\text{f}}^h = 0$ (with computational precision) required for the equivalence when $\mathcal{A}_{\text{rf}} \neq 0$ can be checked out a posteriori, i.e. after solution of (30).

Note that way 2 being applied to the scalar finite-difference Dirichlet problems has given very good results in numerical experiments [26] (in respect to convergence of the iterative process). This can be explained as follows. Let $\mathcal{A} = \mathcal{B}_{\text{rr}}$; it is easy to see that the choice $\mathcal{A}_{\text{rf}} = \mathcal{B}_{\text{rf}}$ and $\mathcal{A}_{\text{ff}} = \mathcal{B}_{\text{ff}} - \mathcal{B}_{\text{fr}}\mathcal{B}_{\text{rr}}^{-1}\mathcal{B}_{\text{rf}}$ makes process (31) to be a direct solver (for $\tau = 1$). This choice being formally the best one has obviously no practical significance, but it points out the way of the enlargement optimization. Namely, the operator $\mathcal{B}_{\text{ff}}^S = \mathcal{B}_{\text{ff}} - \mathcal{B}_{\text{fr}}\mathcal{B}_{\text{rr}}^{-1}\mathcal{B}_{\text{rf}}$ corresponds to the mesh Helmholtz problem inside Ω with the exact absorbing boundary condition on $\partial\Omega$; thus, operator \mathcal{A}_{ff} has to be chosen as an explicitly known sparse matrix as close to $\mathcal{B}_{\text{ff}}^S$ as possible. The mesh operator corresponding to problem (32) satisfies this requirement because it can be treated as the simplest rough approximation of $\mathcal{B}_{\text{ff}}^S$.

3.2 Implementation in the subspace

Let us introduce the mesh operator \mathcal{C} :

$$\mathcal{C} \stackrel{\text{def}}{=} \mathcal{B} - \hat{\mathcal{A}}$$

and rewrite the iterative process as follows:

$$\begin{aligned} \xi_k^h &= (1 - \tau)\xi_{k-1}^h + \tau\mathcal{C}\mathcal{B}^{-1}\xi_{k-1}^h, \quad k = 1, \dots, k_{\max}, \\ \text{where } \xi_k^h &\stackrel{\text{def}}{=} \hat{\mathcal{A}}\hat{u}_k^h - \hat{f}^h - \text{the residual}, \quad \xi_0^h = -\mathcal{C}\mathcal{B}^{-1}\hat{f}^h, \\ \chi_k^h &= \chi_{k-1}^h - \tau\xi_{k-1}^h, \\ \text{where } \chi_k^h &\stackrel{\text{def}}{=} \mathcal{B}\hat{u}_k^h, \quad \chi_0^h = \hat{f}^h. \end{aligned} \tag{34}$$

Having completed this process it is possible to calculate required components of an approximate solution $\hat{u}_{k_{\max}}^h$ as follows:

$$\hat{u}_{k_{\max}}^h = \mathcal{B}^{-1}\chi_{k_{\max}}^h \tag{35}$$

(by definition of χ^h).

It follows from the definitions of $\hat{\mathcal{A}}$ (anyone from Subsection 3.1) and of \mathcal{B} that for any mesh function w^h a value of $\mathcal{C}w^h$ depends only on values of w^h at boundary and near-boundary nodes, and $\mathcal{C}w^h$ can be non-zeroth only at the same nodes (maybe, not at all of them). Let us denote S a set of all such nodes. Thus, it is easy to see that the iterative process can be implemented in the subspace of mesh functions which can be non-zeroth only at S .

Remark. Number of nodes in S is of course finite, so, among other things, we have proved correctness of the above-stated iterative process (indeed, operator \mathcal{B}^{-1} was supposed to be defined on finite mesh functions only).

In 2D case denote $i_{\min}, i_{\max}, j_{\min}, j_{\max}$ such values of indices that S is a subset of mesh rectangle $\Pi = [i_{\min}, i_{\max}] \times [j_{\min}, j_{\max}]$, and denote $N = (i_{\max} - i_{\min})(j_{\max} - j_{\min})$. In 3D case denote $i_{\min}, i_{\max}, j_{\min}, j_{\max}, k_{\min}, k_{\max}$ such values of indices that S is a subset of mesh parallelepiped $\Pi = [i_{\min}, i_{\max}] \times [j_{\min}, j_{\max}] \times [k_{\min}, k_{\max}]$, and denote $N = (i_{\max} - i_{\min})(j_{\max} - j_{\min})(k_{\max} - k_{\min})$. It is easy to see that $\dim S = O(N^{1/2})$ in 2D case and $\dim S = O(N^{2/3})$ in 3D case.

3.3 GMRES

In practice, the preconditioned generalized residual method (GMRES) is used for accelerating iterative process (34):

$$\xi_k^h = \xi_{k-1}^h - \sum_{t=1}^p \tau_k^t (\hat{\mathcal{A}}\mathcal{B}^{-1})^t \xi_{k-1}^h, \quad (36)$$

where τ_k^t , $t = 1, \dots, p$, are chosen to minimize $\|\xi_k^h\|_2$. We remind that

$$\hat{\mathcal{A}}\mathcal{B}^{-1} = (\mathcal{I} - \hat{\mathcal{C}}\mathcal{B}^{-1}).$$

As it has been shown, all the vectors $\xi_{k-1}^h, \chi_{k-1}^h$ are very sparse, so it is possible (in practice) to take a big value of p , up to $p = k$. In the last case we have the full orthogonalization version of GMRES:

$$\begin{aligned} \xi_0^h &= -\mathcal{C}\mathcal{B}^{-1}\hat{f}^h, \tilde{g}_1^h = \hat{\mathcal{A}}\mathcal{B}^{-1}\xi_0^h, \\ \tilde{g}_k^h &= \hat{\mathcal{A}}\mathcal{B}^{-1}\xi_{k-1}^h - \sum_{t=1}^{k-1} \rho_k^t g_t^h, \quad k = 2, \dots, k_{\max}, \\ \text{where } \rho_k^t &= (\hat{\mathcal{A}}\mathcal{B}^{-1}\xi_{k-1}^h, g_t^h), \\ g_k^h &= \frac{\tilde{g}_k^h}{\|\tilde{g}_k^h\|_2}, \\ \xi_k^h &= \xi_{k-1}^h - \beta_k g_k^h, \quad k = 1, \dots, k_{\max}, \\ \text{where } \beta_k &= (\xi_{k-1}^h, g_k^h), \\ \chi_0^h &= \hat{f}^h, \quad \tilde{\tau}_1^h = \xi_0^h, \end{aligned} \quad (37)$$

$$\begin{aligned}\tilde{r}_k^h &= \xi_{k-1}^h - \sum_{t=1}^{k-1} \rho_k^t r_t^h, \quad k = 2, \dots, k_{\max}, \\ r_k^h &= \frac{\tilde{r}_k^h}{\|\tilde{g}_k^h\|_2}, \\ \chi_k^h &= \chi_{k-1}^h - \beta_k r_k^h, \quad k = 1, \dots, k_{\max}.\end{aligned}$$

3.4 Algebraic Optimization of the Enlargement and Additional Preconditioning

In this Subsection we propose some advanced variants of the fictitious domain method based on ideas of algebraic optimization of the enlargement and introducing of an additional preconditioner. We assume here that elements of \mathcal{B}^{-1} are known explicitly because it is necessary for practical implementation of the presented ideas⁴.

First, let us consider a case when operators \mathcal{A} and \mathcal{B}_{rr} coincide or they are in some sense close to each other. It has been shown in Subsection 3.1 that the best choice is then $\mathcal{A}_{\text{rf}} = \mathcal{B}_{\text{rf}}$ and $\mathcal{A}_{\text{ff}} = \mathcal{B}_{\text{ff}}^S$. Thus, an actual matrix \mathcal{A}_{ff} has to be chosen as close to $\mathcal{B}_{\text{ff}}^S$ as possible. Let us try it in the form $\mathcal{A}_{\text{ff}} = \mathcal{B}_{\text{ff}} - \mathcal{C}_{\text{ff}}$ where \mathcal{C}_{ff} is a matrix with some prescribed sparsity pattern, where the latter one is a subset of a sparsity pattern of $\mathcal{B}_{\text{fr}}\mathcal{B}_{\text{rr}}^{-1}\mathcal{B}_{\text{rf}}$. The choice when $(\mathcal{B}_{\text{ff}} - \mathcal{C}_{\text{ff}})(\mathcal{B}_{\text{ff}}^S)^{-1} = I_{\text{f}}$ would be the best one, so let us search the elements c_{ij} of \mathcal{C}_{ff} allowed to be non-zero from the criterion of coincidence of the matrix $(\mathcal{B}_{\text{ff}} - \mathcal{C}_{\text{ff}})(\mathcal{B}_{\text{ff}}^S)^{-1}$ and the identity matrix I_{f} at all pattern positions⁵. It is known that $(\mathcal{B}_{\text{ff}}^S)^{-1} = (\mathcal{B}^{-1})_{\text{ff}}$, so knowing \mathcal{B}^{-1} explicitly we are able to solve this optimization problem constructively. To do this, it is necessary to build up for each row i of \mathcal{C}_{ff} the corresponding p_i -dimensional linear problem for unknown values c_{ij} , $j = 1, \dots, p_i$, and to solve it.

Another proposed approach does not lean upon the suggestion $\mathcal{A} = \mathcal{B}_{\text{rr}}$. It can be briefly called “the zeroth enlargement + additional preconditioning”.

Let us take way 3 of the enlargement from Subsection 3.1 and represent operators \mathcal{B} and $\hat{\mathcal{A}}$ in the block 3×3 -form:

$$\mathcal{B} = \begin{bmatrix} \mathcal{B}_{11} & \mathcal{B}_{12} & 0 \\ \mathcal{B}_{21} & \mathcal{B}_{22} & \mathcal{B}_{2\text{f}} \\ 0 & \mathcal{B}_{\text{f}2} & \mathcal{B}_{\text{ff}} \end{bmatrix}, \quad \hat{\mathcal{A}} = \begin{bmatrix} \mathcal{B}_{11} & \mathcal{B}_{12} & 0 \\ \mathcal{B}_{21} & \mathcal{A}_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (38)$$

⁴Calculation of elements of \mathcal{B}^{-1} is considered in Sections 4 – 7.

⁵It is easy to see that this approach has some treats in common with the incomplete factorization approaches.

Here the third f-group consists of all fictitious unknowns and the second one consists of all real unknowns whose rows in \mathcal{B} and in $\hat{\mathcal{A}}$ differ from each other (which is possible at near-boundary layer only).

Then let us introduce an additional preconditioner H replacing everywhere \mathcal{B}^{-1} by $\mathcal{B}^{-1}H$ where (38)-form of H is:

$$H = \begin{bmatrix} I_1 & 0 & 0 \\ 0 & H_{22} & 0 \\ 0 & 0 & I_f \end{bmatrix} \quad (39)$$

with yet unknown non-degenerate matrix H_{22} . Thus, iterative process (31) takes the form:

$$\begin{aligned} \hat{u}_k^h - \hat{u}_{k-1}^h &= -\tau \mathcal{B}^{-1}H(\hat{\mathcal{A}}\hat{u}_{k-1}^h - \hat{f}^h), \quad k = 1, \dots, k_{\max}, \\ \text{where } \hat{u}_0^h &= \mathcal{B}^{-1}\hat{f}^h, \quad \tau = \text{const.} \end{aligned} \quad (40)$$

It is easy to check that the choice of H_{22} :

$$H_{22} = (\mathcal{B}_{21}(\mathcal{B}^{-1})_{12} + \mathcal{A}_{22}(\mathcal{B}^{-1})_{22})^{-1}$$

makes iterative process (40) to be a direct solver (for $\tau = 1$). Thus, in practice we can prescribe some sparsity pattern of H_{22} and then calculate corresponding coefficients in the same way as it was proposed for \mathcal{A}_f in the first approach.

The arithmetical expenses and the required computer memory for both proposed approaches depend on the number of pattern positions and on a used method of solving the arising problems with dense $p_i \times p_i$ -matrices. If $\max p_i$ is a constant then anyway in 3D case these expenses and the required memory are of the order of $N^{2/3}$.

3.5 Application to the vector problem

For solving “vector” mesh problem (23) just the same approach can be applied. For construction of a preconditioner we consider the problem

$$-\Delta \mathbf{E} - \kappa^2 \mathbf{E} = 0 \quad \text{in } \mathbf{R}^3 \quad (41)$$

and discretize it in the way described in Subsections 2.5-2.6. As a result we get the following mesh operator \mathcal{B} (see (12)):

$$\mathcal{B} = \begin{bmatrix} \mathcal{B} & 0 & 0 \\ 0 & \mathcal{B} & 0 \\ 0 & 0 & \mathcal{B} \end{bmatrix}, \quad (42)$$

where the operator \mathcal{B} here is the same as for the scalar case above. Unknowns of the i th group in (42), $i = \overline{1, 3}$, correspond to values of the Cartesian component E_i^h of the searched vector function.

We see that

$$\mathcal{B}^{-1} = \begin{bmatrix} \mathcal{B}^{-1} & 0 & 0 \\ 0 & \mathcal{B}^{-1} & 0 \\ 0 & 0 & \mathcal{B}^{-1} \end{bmatrix} \quad (43)$$

and each Cartesian component of a solution of the vector problem satisfies the same condition at infinity as a solution of the scalar one (see (4)). Hence, construction of an algorithm for multiplying \mathcal{B}^{-1} by a finite “vector” mesh function ϕ^h reduces to construction of an algorithm for multiplying \mathcal{B}^{-1} by a finite “scalar” mesh function ϕ^h .

Thus, there is no difference between the vector and the scalar cases in application of the fictitious domain method when solving finite-difference systems of equations or equations of the finite-element approximation with the nodal Lagrange basic functions. But the method cannot be applied directly to the finite-element approximation of the vector problem when the edge basic functions are used (see Subsection 2.7) because there is no topological equivalence between edges of mesh \mathbf{R}_h^m and of a locally fitted mesh $(\mathbf{R}^m \setminus \bar{\Omega})_h$ in the near-boundary layer.

To generalize the method to this case, let us modify it as follows:

1. Considered problem (14) is represented in the form

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix} \begin{bmatrix} u_1^h \\ u_2^h \end{bmatrix} = \begin{bmatrix} f_1^h \\ f_2^h \end{bmatrix}, \quad (44)$$

where the second group corresponds to all mesh unknowns on $(\mathbf{R}^m \setminus \bar{\Omega})_h$ having no topological counterparts on \mathbf{R}_h^m .

2. The equivalent form of (44) is constructed:

$$(\mathcal{A}_{11} - \mathcal{A}_{12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21})u_1^h = f_1^h - \mathcal{A}_{12}\mathcal{A}_{22}^{-1}f_2^h, \quad (45)$$

$$u_2^h = \mathcal{A}_{22}^{-1}(f_2^h - \mathcal{A}_{21}u_1^h). \quad (46)$$

3. The iterative method described in Subsections 3.1-3.3 is applied to the problem in its form (45) instead of (14). It is possible by the construction of (44).

Of course, for the implementation of this variant it is necessary among other things to multiply matrix $\mathcal{A}_{12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21}$ by a vector ξ_1^h at each iterative step, i.e. to solve problem with matrix \mathcal{A}_{22} . This can be done by GMRES preconditioned by the incomplete factorization approach. The corresponding arithmetical expenses are comparatively small because \mathcal{A}_{22} is an $O(N^{2/3})$ -dimensional sparse matrix with the strong diagonal dominance at almost each row.

4 Approximation of the radiation condition

In this Section we shall set forth a way of an approximation of the radiation condition for constructing an algorithm of multiplying \mathcal{B}^{-1} by a finite mesh function ϕ^h . For this purpose we shall use the Fourier analysis.

4.1 The Fourier form for the exact radiation condition

Let us first establish some properties of an exact solution u of differential problem (1). Let Ω belong to the layer $x_1^{\min} < x_1 < x_1^{\max}$. For $x_1 \leq x_1^{\min}$ and $x_1 \geq x_1^{\max}$ let us represent an exact solution u of the 2D differential problem in the form:

$$u(x_1, x_2) = \int_{-\infty}^{+\infty} U(x_1, \gamma_2) e^{i\gamma_2 x_2} d\gamma_2 \quad (47)$$

Substituting it into the equation we get the easy-to-solve ordinary differential equation for each value of γ_2 and $x_1 \leq x_1^{\min}$, $x_1 \geq x_1^{\max}$:

$$U''_{x_1 x_1} + (\kappa^2 - \gamma_2^2)U = 0, \quad (48)$$

so

$$U(x_1, \gamma_2) = U_1(\gamma_2) e^{i\sqrt{\kappa^2 - \gamma_2^2} x_1} + U_2(\gamma_2) e^{-i\sqrt{\kappa^2 - \gamma_2^2} x_1} \quad (49)$$

with some $U_1(\gamma_2)$, $U_2(\gamma_2)$. Hereafter for negative a we suppose $\sqrt{a} = i\sqrt{|a|}$.

Anyway, $U(x_1, \gamma_2)$ does not increase exponentially for $|x_1| \rightarrow \infty$, so we have for $|\gamma_2| > \kappa$:

$$U_1(\gamma_2) = 0 \text{ for } x_1 \leq x_1^{\min}, \quad U_2(\gamma_2) = 0 \text{ for } x_1 \geq x_1^{\max}. \quad (50)$$

But the same is valid for $|\gamma_2| < \kappa$ if u is a divergent wave. In other words, function $U(x_1, \gamma_2)$ for each $|\gamma_2| < \kappa$ is “the left-hand spiral” on left semiaxis $x_1 \leq x_1^{\min}$ and “the right-hand spiral” in right semiaxis $x_1 \geq x_1^{\max}$. It is even more general description of a radiated wave than the Sommerfeld condition.

Restriction (50) on the function from (49) is equivalent to the following boundary conditions:

$$\begin{aligned} U'_{x_1}(x_1, \gamma_2) + i\sqrt{\kappa^2 - \gamma_2^2} U(x_1, \gamma_2) &= 0 \text{ for } x_1 = x_1^{\min}, \\ U'_{x_1}(x_1, \gamma_2) - i\sqrt{\kappa^2 - \gamma_2^2} U(x_1, \gamma_2) &= 0 \text{ for } x_1 = x_1^{\max}. \end{aligned} \quad (51)$$

The same consideration is carried out in 3D case. Solution u of problem (1) is represented in the form:

$$u(x_1, x_2, x_3) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} U(x_1, \gamma_2, \gamma_3) e^{i\gamma_2 x_2} e^{i\gamma_3 x_3} d\gamma_2 d\gamma_3, \quad (52)$$

substituting gives us the ordinary differential equation:

$$U''_{x_1 x_1} + (\kappa^2 - (\gamma_2^2 + \gamma_3^2))U = 0, \quad (53)$$

so

$$U(x_1, \gamma_2, \gamma_3) = U_1(\gamma_2, \gamma_3) e^{i\sqrt{\kappa^2 - (\gamma_2^2 + \gamma_3^2)} x_1} + U_2(\gamma_2, \gamma_3) e^{-i\sqrt{\kappa^2 - (\gamma_2^2 + \gamma_3^2)} x_1}, \quad (54)$$

for any γ_2, γ_3 we have:

$$U_1(\gamma_2, \gamma_3) = 0 \text{ for } x_1 \leq x_1^{\min}, \quad U_2(\gamma_2, \gamma_3) = 0 \text{ for } x_1 \geq x_1^{\max}, \quad (55)$$

and equivalent boundary conditions are:

$$\begin{aligned} U'_{x_1}(x_1, \gamma_2, \gamma_3) + i\sqrt{\kappa^2 - (\gamma_2^2 + \gamma_3^2)} U(x_1, \gamma_2, \gamma_3) &= 0 \text{ for } x_1 = x_1^{\min}, \\ U'_{x_1}(x_1, \gamma_2, \gamma_3) - i\sqrt{\kappa^2 - (\gamma_2^2 + \gamma_3^2)} U(x_1, \gamma_2, \gamma_3) &= 0 \text{ for } x_1 = x_1^{\max}. \end{aligned} \quad (56)$$

4.2 An approximate radiation condition in the Fourier form

The main idea of an approximation of the radiation condition is to repeat all the consideration of the previous Subsection at mesh level.

For definiteness, let \mathcal{B} be the finite-difference 2D operator, i.e.

$$(\mathcal{B}v^h)_{ij} = h^{-2}(4v_{ij}^h - v_{i-1,j}^h - v_{i+1,j}^h - v_{i,j-1}^h - v_{i,j+1}^h) - \kappa^2 v_{ij}^h.$$

Let $h = 1$ for simplicity (without loss of generality).

So, we should solve the equation

$$\mathcal{B}v^h = \phi^h \quad (57)$$

for a finite mesh function ϕ^h in such a way that v^h satisfies an approximate radiation condition. We remind that a function ϕ^h can be non-zero only at points of S , and solution v^h should be found at points of S only; the problem in this form is called “the partial solution problem” [8, 11, 12, 23, 24, 27, 30].

Let us represent v^h in the well-known form:

$$v_{ij}^h = \int_{-\pi}^{\pi} V_i(\gamma_2) e^{ij\gamma_2} d\gamma_2. \quad (58)$$

Substituting it into equation (57) we get the infinite system of three-point mesh equations for each value of γ_2 :

$$-V_{i-1}(\gamma_2) + (2 - \kappa^2 + 4 \sin^2 \frac{\gamma_2}{2}) V_i(\gamma_2) - V_{i+1}(\gamma_2) = \Phi_i(\gamma_2), \quad i = \overline{-\infty, \infty}, \quad (59)$$

where

$$\Phi_i(\gamma_2) = \frac{1}{2\pi} \sum_{j=-\infty}^{+\infty} \phi_{ij}^h e^{-ij\gamma_2} d\gamma_2. \quad (60)$$

Obviously, $\Phi_i(\gamma_2) = 0$ for $i < i_{\min}$ and $i > i_{\max}$. Thus, for $i \leq i_{\min}$ and $i \geq i_{\max}$ we can search a solution of (59) in the form:

$$V_i(\gamma_2) = q^i(\gamma_2),$$

getting the quadratic characteristic equation for q :

$$q^2(\gamma_2) - t(\gamma_2)q(\gamma_2) + 1 = 0, \quad (61)$$

where

$$t(\gamma_2) \stackrel{\text{def}}{=} 2 - \kappa^2 + 4 \sin^2 \frac{\gamma_2}{2}.$$

Solving (61) we get:

$$q_1(\gamma_2) = \frac{t(\gamma_2) + i\sqrt{4 - t^2(\gamma_2)}}{2} = q(\gamma_2), \quad q_2(\gamma_2) = q^{-1}(\gamma_2),$$

so we have:

$$V_i(\gamma_2) = C_1(\gamma_2)q^i(\gamma_2) + C_2(\gamma_2)q^{-i}(\gamma_2).$$

The mesh radiation condition are analogous to condition (50):

$$C_1(\gamma_2) = 0 \quad \text{for } i \leq i_{\min}, \quad C_2(\gamma_2) = 0 \quad \text{for } i \geq i_{\max}.$$

for any value of $\gamma_2 \in [-\pi, \pi]$.

The equivalent boundary conditions are:

$$V_{i_{\min}-1}(\gamma_2) = q(\gamma_2)V_{i_{\min}}(\gamma_2), \quad V_{i_{\max}+1}(\gamma_2) = q(\gamma_2)V_{i_{\max}}(\gamma_2). \quad (62)$$

The same derivation is done in 3D case using the two-fold Fourier transformation instead of the one-fold one.

Remark. We note that such approximation of the radiation condition can be carried out in the same way when the considered problem has been approximated by means of finite-element method with bilinear (trilinear in 3D case) basic functions.

5 Algorithm for multiplying \mathcal{B}^{-1} by a finite function

In principle, the following method for solving (57) can be proposed. We calculate a solution v^h by means of relation (58) using some approximate integration formula for calculating the integral there. Suppose that this formula uses values of $V_i(\gamma_2)$ at points γ_2^t , $t = 1, \dots, n_\gamma$. We calculate values of $V_i(\gamma_2^t)$, $i = i_{\min}, \dots, i_{\max}$, for each γ_2^t by solving equation (59) with boundary conditions (62), where right-hand side $\Phi_i(\gamma_2^t)$ is calculated by formula (60) (it is easy to see that summation in (60) is in fact finite).

But estimation of number of arithmetical operations needed for implementation of this algorithm shows that it is too expensive. This is because the integrated function in (58) is singular and highly-oscillating, so a grid γ_2^t should be extremely fine everywhere in $[\pi, \pi]$, otherwise fast Fourier algorithms cannot be applied.

Now we propose another algorithm. Let us introduce the mesh Green function G^h :

$$(\mathcal{B}G^h)_{ij(k)} = \begin{cases} 1 & \text{if } i = 0, \quad j = 0, \quad (k = 0,) \\ 0 & \text{otherwise,} \end{cases}$$

and G^h satisfies mesh radiation condition (62) for $i_{\min} = i_{\max} = 0$.

It is easy to derive from equations (58), (59), (60) and (62) that in 2D case

$$G_{ij}^h = \frac{i}{2\pi} \int_{-\pi}^{\pi} \frac{q^{|i|}(\gamma_2) e^{ij\gamma_2}}{\sqrt{4 - t^2(\gamma_2)}} d\gamma_2. \quad (63)$$

It is seen that value $t(\gamma_2) = 2$ (i.e. $\kappa^2 = 4 \sin^2 \frac{\gamma_2}{2}$) is a singular point of the integrated function in (63), but it is easy to check out that this singularity is integrable.

The analogous formula in 3D case built up with the two-fold Fourier transformation is:

$$G_{ijk}^h = \frac{i}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{q^{|i|}(\gamma_2, \gamma_3) e^{ij\gamma_2} e^{ik\gamma_3}}{\sqrt{4 - t^2(\gamma_2, \gamma_3)}} d\gamma_2 d\gamma_3, \quad (64)$$

where

$$t(\gamma_2, \gamma_3) \stackrel{\text{def}}{=} 2 - \kappa^2 + 4 \sin^2 \frac{\gamma_2}{2} + 4 \sin^2 \frac{\gamma_3}{2}.$$

An efficient algorithm for calculation of the integral in (63) has been worked out by A.Padiy. He used results of the theory of analytical functions of complex-valued argument. The main idea consists in a choice of another path of integration in the complex plane (passing far from the singularity) and then in the use of a special formula for approximate integration of highly-oscillating functions [5]. This algorithm can also be used for the first integration in (64).

We have the obvious formulas for solution v^h of problem (57):

$$\begin{aligned} v_{i,j}^h &= \sum_{l,m} \phi_{lm}^h G_{i-l,j-m}^h \quad \text{in 2D,} \\ v_{i,j,k}^h &= \sum_{l,m,n} \phi_{lmn}^h G_{i-l,j-m,k-n}^h \quad \text{in 3D.} \end{aligned} \quad (65)$$

These formulas are valid because all nodes of an infinite uniform Cartesian grid are equivalent to each other.

Remark. Note that the Green function G^h is nothing but the row of \mathcal{B}^{-1} . Thus, having calculated needed components of G^h we are able to apply the optimization approaches from Subsection 3.4.

Let us estimate arithmetical expenses of partial solution of problem (57) by means of summation (65). We remind that $\dim S = O(N^{1/2})$ in 2D case and $\dim S = O(N^{2/3})$ in 3D case, thus, a number of arithmetical operations needed for the implementation of (65) is $O(N)$ in 2D case and $O(N^{4/3})$ in 3D case. We see that it is too expensive in the latter case.

However, a fast algorithm can be proposed for implementation of (65). It has been taken from the theory of fast multiplication of so-called circulant matrices by a vector. To diminish cumbersomeness, it is described for 2D case, but its generalization for 3D case is obvious.

Let

$$G^{lm} = \sum_{i=-(i_{\max}-i_{\min})}^{i_{\max}-i_{\min}} \sum_{j=-(j_{\max}-j_{\min})}^{j_{\max}-j_{\min}} G_{ij}^h e^{\frac{-i\pi il}{2(i_{\max}-i_{\min})+1}} e^{\frac{-i\pi jm}{2(j_{\max}-j_{\min})+1}} \quad (66)$$

$$l = -(i_{\max} - i_{\min}), \dots, i_{\max} - i_{\min}, \quad m = -(j_{\max} - j_{\min}), \dots, j_{\max} - j_{\min},$$

and

$$\phi^{lm} = \frac{1}{(2(i_{\max} - i_{\min}) + 1)(2(j_{\max} - j_{\min}) + 1)} \cdot \sum_{i=2i_{\min}-i_{\max}}^{i_{\max}} \sum_{j=2j_{\min}-j_{\max}}^{j_{\max}} \phi_{ij}^h e^{\frac{-i\pi il}{2(i_{\max}-i_{\min})+1}} e^{\frac{-i\pi jm}{2(j_{\max}-j_{\min})+1}}, \quad (67)$$

$$l = -(i_{\max} - i_{\min}), \dots, i_{\max} - i_{\min}, \quad m = -(j_{\max} - j_{\min}), \dots, j_{\max} - j_{\min},$$

then it can be easily shown that

$$v_{ij}^h = \sum_{m=-(j_{\max}-j_{\min})}^{j_{\max}-j_{\min}} \sum_{l=-(i_{\max}-i_{\min})}^{i_{\max}-i_{\min}} G^{lm} \phi^{lm} e^{\frac{i\pi il}{2(i_{\max}-i_{\min})+1}} e^{\frac{i\pi jm}{2(j_{\max}-j_{\min})+1}}, \quad (i, j) \in \Pi. \quad (68)$$

Thus, the following algorithm for summation (65) is proposed. First, we once calculate by (66) and store coefficients G^{lm} . Then we calculate coefficients ϕ^{lm} by (67) and next apply formula (68). Of course, fast Fourier transformation algorithms can be used here, so the arithmetical expenses are $O(N \log N)$ both in 2D and 3D cases. Adaptation of the algorithm for the partial solution is constructed in the ordinary way [6, 8, 11, 12, 23, 24, 27, 30], i.e. with the use of the trivial algorithm for the summations over j in (67) and over m in (68), and dealing there with nodes of S only.

Remark. Formulas (66)-(68) for summation (65) inside Π are valid for an arbitrary function G^h . But it is known that the Green function is even with respect to indices i, j . Besides that, $\phi_{ij}^h = 0$ for $i < i_{\min}, j < j_{\min}$. This can be used for a considerable decrease of number of arithmetic operations in practical implementation of this approach.

The proposed algorithm has an obvious drawback: calculation of the mesh Green function in the whole mesh parallelepiped Π is too expensive, and its storage requires $O(N)$ computer memory. To remove this drawback, a special trick is proposed in Section 7 for 3D case.

6 The mesh Green function of $(\nabla \times)_H^h (\nabla \times)_E^h - \kappa^2$

The first way of optimization presented in Subsection 3.4 is the most effective when the matrix \mathcal{A} coincides with \mathcal{B}_{rr} or is close to it. It is really so for problem (21) but up to now the fictitious domain method was considered as being applied to problem (23), for which it is not so. That's why it would be of the practical importance to apply the method to problem (21) itself.

For this purpose it is sufficient to replace in (65) the “scalar” Green function G^h of the operator $-\Delta^h - \kappa^2$ by the “vector” one \mathbf{G}^h of the operator $(\nabla \times)_H^h (\nabla \times)_E^h - \kappa^2$. To construct \mathbf{G}^h let us consider the problem:

$$\begin{aligned} -\Delta_E^h \hat{\mathbf{G}}^h - \kappa^2 \hat{\mathbf{G}}^h &= 0 \quad \text{at } (\mathbf{R}^3)_h^E \setminus (1/2, 0, 0), \\ \hat{\mathbf{G}}_{1,1/2,0,0}^h &= 1 \end{aligned} \quad (69)$$

(we remind that node $(1/2, 0, 0)$ of $(\mathbf{R}^3)_h^E$ is the middle point of mesh edge $[(0, 0, 0), (1, 0, 0)]$ and corresponds to the Cartesian component E_1). It can be easily seen from Section 2 that this problem is a particular case of problem (23) and its solution $\hat{\mathbf{G}}^h$ provides us the searched function \mathbf{G}^h (related with node $(1/2, 0, 0)$):

$$\mathbf{G}^h = \hat{\mathbf{G}}^h / \alpha_0 \quad \text{where} \quad \alpha_0 = ((\nabla \times)_H^h (\nabla \times)_E^h \hat{\mathbf{G}}^h - \kappa^2 \hat{\mathbf{G}}^h)_{1,1/2,0,0}. \quad (70)$$

Let us solve (69) by the fictitious domain method. In this case $\dim S = 15$ (see Subsection 3.2), and $\hat{\mathbf{G}}^h$ is represented by means of the “scalar” Green functions G^h of the fifteen neighboring nodes as follows:

$$\begin{aligned} \hat{\mathbf{G}}_{1,i+1/2,j,k}^h &= \alpha_1 G_{i,j,k}^h + \alpha_2 (G_{i-1,j,k}^h + G_{i+1,j,k}^h) + \alpha_3 (G_{i,j-1,k}^h + G_{i,j+1,k}^h \\ &\quad + G_{i,j,k-1}^h + G_{i,j,k+1}^h) + \delta_{i,j,k}^{0,0,0}, \\ \hat{\mathbf{G}}_{2,i,j+1/2,k}^h &= \alpha_4 (G_{i,j,k}^h - G_{i,j+1,k}^h - G_{i-1,j,k}^h + G_{i-1,j+1,k}^h), \\ \hat{\mathbf{G}}_{3,i,j,k+1/2}^h &= \alpha_4 (G_{i,j,k}^h - G_{i,j,k+1}^h - G_{i-1,j,k}^h + G_{i-1,j,k+1}^h) \end{aligned} \quad (71)$$

($\delta_{i,j,k}^{0,0,0}$ is the Kronecker symbol). Here there are only 4 different coefficients instead of 15 ones because of the symmetry and $h = 1$ condition. Substituting (71) into (69) we get:

$$\alpha_4 = \frac{6(G_{1,0,0}^h)^2 / G_{0,0,0}^h - G_{0,0,0}^h - G_{2,0,0}^h - 4G_{1,1,0}^h + 1}{-2(G_{1,0,0}^h)^2 / G_{0,0,0}^h + 5G_{0,0,0}^h + G_{2,0,0}^h + 4G_{1,1,0}^h - 1 - 8G_{1,0,0}^h},$$

$$\begin{aligned}
\alpha_3 &= 1, \\
\alpha_2 &= \alpha_4 + 1, \\
\alpha_1 &= -\frac{(2\alpha_2 + 4)G_{1,0,0}^h}{G_{0,0,0}^h}.
\end{aligned} \tag{72}$$

From the physical viewpoint, \mathbf{G}^h is the finite-difference approximation of the wave radiated by the point unit dipole placed at $(1/2, 0, 0)$, directed along x_1 and harmonically oscillating with the frequency $\omega = c\kappa$.

Remark. Of course, similar formulas can be analogously derived for the finite-element bilinear-constant approximation.

7 The trick

It can be expected that the mesh Green function G^h for $r \neq 0$ has to converge locally to the Green function $\frac{1}{4\pi} \frac{e^{i\kappa r}}{r}$ of the differential Helmholtz operator for $\lambda/h \rightarrow \infty$. This has been validated in numerical experiments. As an example, see Figure 1 where the exact and mesh Green functions are compared for $r < 16h$ in the case of finite-difference approximation with $h = \lambda/10$. The first plot in this Figure is the real and imaginary parts of $\frac{1}{4\pi} \frac{e^{i\kappa r}}{r}$, the other three plots are the same for the mesh Green function obtained by (64), in directions $(1, 0, 0)$, $(1, 1, 0)$, $(1, 1, 1)$ correspondingly.

This reasoning leads us to the idea to substitute everywhere the exact G^h for the function \tilde{G}^h :

$$\tilde{G}_{ijk}^h = \begin{cases} G_{ijk}^h & \text{if } r = \sqrt{i^2 + j^2 + k^2} < R_{\text{sw}}, \\ \frac{1}{4\pi} \frac{e^{i\kappa r}}{r} & \text{otherwise.} \end{cases} \tag{73}$$

Of course, this substitution produces some error, but it is of the same order as the approximation error.

Thus, now we need to calculate and to store values G_{ijk}^h inside the sphere $r \leq R_{\text{sw}}$ only, where a value of R_{sw} may depend on the parameter λ/h and needed accuracy. So, the computer memory required for implementation of the method is $O(N^{2/3})$ in 3D case.

Numerical experiments for the function \tilde{G}^h have shown that modulus of the local residual $|((-\Delta^h - \kappa^2)\tilde{G}^h)_{i,j,k} - \delta_{i,j,k}^{0,0,0}|$ at nodes (i, j, k) neighboring to the surface $r = R_{\text{sw}}$ is 10-20 times more than at other points with $r > R_{\text{sw}}$, depending on R_{sw} and λ/h . To avoid this detrimental phenomenon, the least squares smoothing can

be used, i.e. we can take as \tilde{G}^h for $r < R_{\text{sw}}$ a mesh function minimizing the norm $\|(-\Delta^h - \kappa^2)\tilde{G}^h - \delta^{0,0,0}\|_2$ calculated throughout all mesh nodes in which the residual depends on values of \tilde{G}^h in sphere $r < R_{\text{sw}}$.

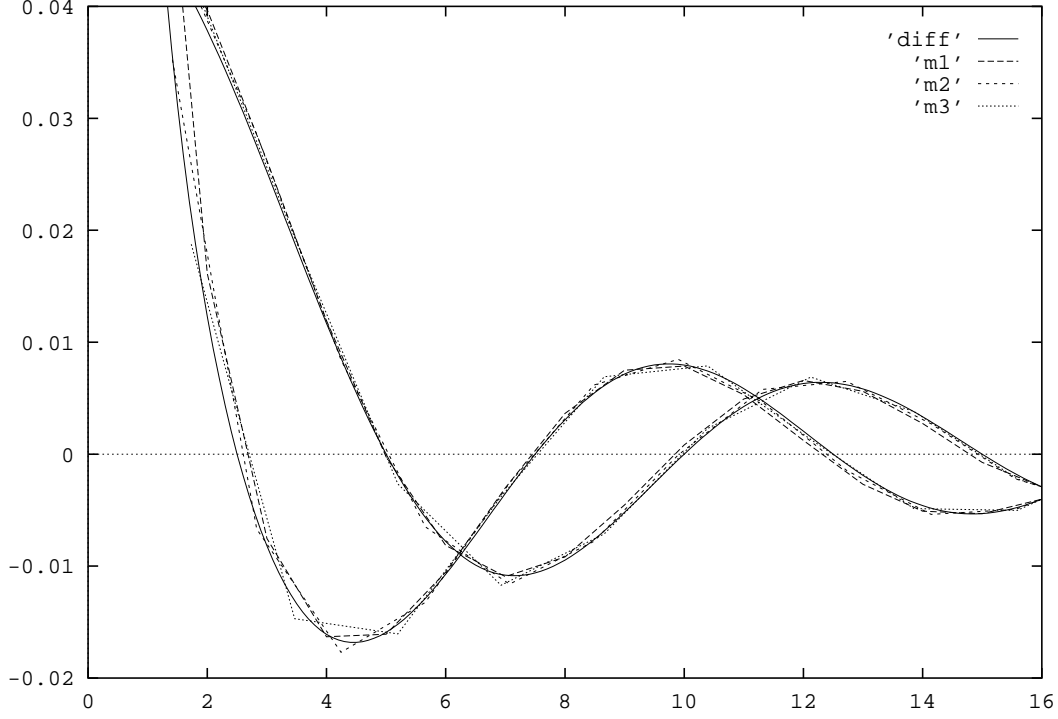


Figure 1. Comparison of the differential and mesh Green functions

Remark. Note also that the least squares smoothing allows to avoid practical calculation of G^h by (63) or (64). But these formulas remain significant from the theoretical viewpoint because they provide the exact mesh Green functions.

8 Calculation of the asymptotic function

It is well known that calculation of the functions A from (3), (8) is of great practical importance. In this Section an approximate formula is derived for it in 3D case.

It follows from formulas (35), (65) and the used trick that for nodes (i, j, k) which are far from the obstacle we have (for $h = 1$):

$$u_{ijk}^h = \frac{1}{4\pi} \sum_{lmn} \chi_{lmn}^h \frac{e^{i\kappa \sqrt{(i-l)^2 + (j-m)^2 + (k-n)^2}}}{\sqrt{(i-l)^2 + (j-m)^2 + (k-n)^2}}. \quad (74)$$

Making the Taylor expansion in (74) for $r = \sqrt{i^2 + j^2 + k^2} \rightarrow \infty$ we get:

$$\begin{aligned} \frac{e^{i\kappa \sqrt{(i-l)^2 + (j-m)^2 + (k-n)^2}}}{\sqrt{(i-l)^2 + (j-m)^2 + (k-n)^2}} &= \frac{e^{i\kappa r(1 - \frac{l i + m j + k n}{r^2})}}{r} + O(r^{-2}) = \\ &= e^{-i\kappa(\beta_1 l + \beta_2 m + \beta_3 n)} \frac{e^{i\kappa r}}{r} + O(r^{-2}), \end{aligned} \quad (75)$$

where $\beta_1 = i/r$, $\beta_2 = j/r$, $\beta_3 = k/r$. The unit vector $(\beta_1, \beta_2, \beta_3)$ defines a direction of observation, its components are connected with the spherical variables θ, φ as follows:

$$\beta_1 = \cos \theta \cos \varphi, \quad \beta_2 = \cos \theta \sin \varphi, \quad \beta_3 = \sin \theta. \quad (76)$$

Comparing (3) and (74)-(75) we have:

$$A^h(\theta, \varphi) = \frac{1}{4\pi} \sum_{lmn} \chi_{lmn}^h e^{-i\kappa(\beta_1 l + \beta_2 m + \beta_3 n)}. \quad (77)$$

The same formula is valid for the asymptotic form of the Cartesian components of a scattered electric field \mathbf{E}_s^h .

9 Results of numerical experiments

In this Section we present some results of numerical experiments carried out for 3D scalar and vector (Maxwell) problems with the above-described approach for its verification and testifying. In all of them the author used the full GMRES iterative method (37), with the criterion $\frac{\|\xi\|_2}{\|f\|_2} = 10^{-6}$ for stopping, and the trick described in Section 7 (where $R_{sw} = \lambda$ was taken) with the least squares smoothing.

First, a set of experiments had been carried out for the 3D acoustic problem with spherical obstacles, from $r = \lambda/2$ to $r = 4\lambda$ and from $h = \lambda/10$ to $h = \lambda/40$, using way 2 of the enlargement from Subsection 3.1. Each calculated RCS plot (i.e. $|A^h(\theta, \varphi)|$ from (77)) was compared with the exact one obtained by means of the

exact semi-explicit formula from [34]. This comparison has demonstrated the satisfactory precision of the calculations (taking into account the rough approximation of spheres). For $h = \lambda/40$ these two plots have almost coincided.

The number of iterations in these experiments depended almost not at all on h for fixed κ and obstacle, and it was nearly a linear growing function of κ for a fixed obstacle. The calculation for the sphere $r = \lambda$ with $h = \lambda/10$ required 18 iterations and 2 min of HP9000/735 CPU time (without an expense for postprocessing (77)).

Calculations were also carried out for several other obstacles, for example for a semi-open cavity of circular section (see Fig. 2). There results were also quite satisfactory. The calculation for the Oxford-95 Workshop acoustic test case C3D_3 with $h = \lambda/30$ and the minimal embracing parallelepiped $38 \times 65 \times 38$ steps required 31 iterations and 20 min of HP9000/735 CPU time. The calculated RCS plot was coincide very well with an independently obtained result of Dr. M.-O. Bristeau (INRIA-Rocquencourt, France).

Second, the same set of experiments with spherical obstacles had been carried out for the 3D finite-difference Maxwell problem (23) using way 2 of the enlargement from Subsection 3.1 again (the corresponding computer code is called ElPack1-1.0, i.e. Electromagnetic Package, the 1st order approximation, version 1.0). The calculated RCS plots for plane-polarized incident waves were also compared with the exact ones obtained by means of the exact semi-explicit formula [34]. The results were also satisfactory.

But it should be noted that this variant of the method has exhibited rather poor convergence, much worse than in the acoustic case. For example, the calculation for the sphere $r = \lambda/2$ with $h = \lambda/20$ required 107 iterations and with $h = \lambda/30$ required 192 iterations. The number of iterations was growing with h^{-1} faster than linearly, and it was again nearly a linear growing function of κ for fixed obstacle and λ/h -parameter.

Third, the approach has been implemented for solving 3D finite-difference Maxwell problem (21) using the enlargement optimization from Subsection 3.4 and the approach described in Section 6 (the corresponding computer code is called ElPack1-3.1). In the optimization, an element c_{lm} of \mathcal{C}_{ff} was allowed to be non-zero iff

$$\max\{|i_l - i_m|, |j_l - j_m|, |k_l - k_m|\} \leq d, \quad (78)$$

where in practice the value $d = 2$ was used. The dense linear systems arising during the optimization were solved by the Gauss elimination method. Experiments have shown that for the same problem ElPack1-3.1 required 3-4 times less iterations than ElPack1-1.0, but the above-stated dependencies of their number on h and κ remained nearly the same.

By means of ElPack1-3.1 all the 3D electromagnetic stationary test cases for uncoated obstacles of the Oxford-95 Workshop [9] have been successfully solved. For example, test case C3D_1 corresponded to scattering on semi-open cavity of inner depth 2λ and of circular cross section of diameter λ (see Fig. 2-8), for various directions and polarizations of the incident wave. The calculations were carried out with $h = \lambda/20$ and the minimal embracing parallelepiped $24 \times 42 \times 24$ steps. They required 130-140 iterations and 3 hours of HP9000/735 CPU time, of which 2 hours 10 min was the overhead for the optimization from Section 3.4.

Another test case, O3D_2, corresponded to an ogive of length 2λ and of elliptic cross-section $0.4\lambda \times 0.8\lambda$ (see Fig. 9-13). The calculations were carried out with $h = \lambda/40$ and the minimal embracing parallelepiped $32 \times 80 \times 16$ steps. They required around 100 iterations and 1 hour 30 min of HP9000/735 CPU time, of which 40 min was the overhead for the optimization.

Some results of the calculations (components of the total near-field \mathbf{E}) are presented in Fig. 2-13. We remind that the frame of each of these Figures is not an artificial infinity but a graphical frame only. It can be seen that the plotted fields do not satisfy the boundary condition near entering angles of obstacles; but this phenomenon has arisen during the graphical postprocessing only, because being plotted the functions discontinuous at that points were approximated by continuous piecewise-linear functions.

The calculated RCS plots were compared during the Workshop with results of other researches obtained mainly by the boundary integral method. They coincided with a satisfactory precision (taking into account the rough staircase approximation of obstacle boundaries).

Fourth, the finite-element version of the method (see Subsection 2.7) has been implemented in a computer code called ElPack2-1.1. It is based on the zeroth enlargement with the additional preconditioner. The latter one is calculated by the method described in Subsection 3.4. Transformation (45)-(46) is also used. For generation of 3D locally fitted Cartesian meshes the code developed by Drs. S. Finogenov and A. Supalov (Institute of Numerical Mathematics of RAS, Moscow) is used.

As it had been expected, for the same test cases and with the same basic meshes (13) the finite-element code gave more precise results than the finite-difference one (being compared with analytical results for spheres and with results of other researches for the Oxford-95 test cases). For instance, the calculation was carried out for the sphere $r = 0.748\lambda$ using the mesh $h = \lambda/16$. L_2 -norm of the error of the bistatic logarithmic RCS plot calculated by ElPack2 was about 3 times less than for ElPack1.

But it must be noted that the finite-element code is far more memory-consuming and arithmetically expensive. For example, Elpack2-1.1 requires about 3 times more iterations than Elpack1-3.1 for the same value of the parameter d from (78).

In all the experiments, during postprocessing (77) the mesh values of

$$\left(\lim_{d(\Gamma) \rightarrow \infty} \frac{\int_{\Gamma} |E_s^r|^2}{\int_{\Gamma} |E_s|^2} \right)^{1/2} \quad (79)$$

were calculated, where E_s^r was the radial component of the scattered wave E_s , Γ was a circle centering at the origin, and the spherical components of the vector field were calculated from its Cartesian components. Values of (79) should be zeroth (see (8)) but it was not exactly so in calculations because of approximation and truncation errors. However, for $\lambda/h = 10$ they were not greater than 1% and were decreasing as $(\lambda/h)^{-2}$ when the parameter λ/h increased. These results seem to be satisfactory.

Acknowledgments. The author is very grateful to Profs. Yu. Kuznetsov and O. Pironneau for the scientific supervising, to Dr. K. Lipnikov for auxiliary computer codes, and to all of them for fruitful discussions. He is also very grateful to Drs. S. Finogenov and A. Supalov for putting their mesh generator into his disposal.

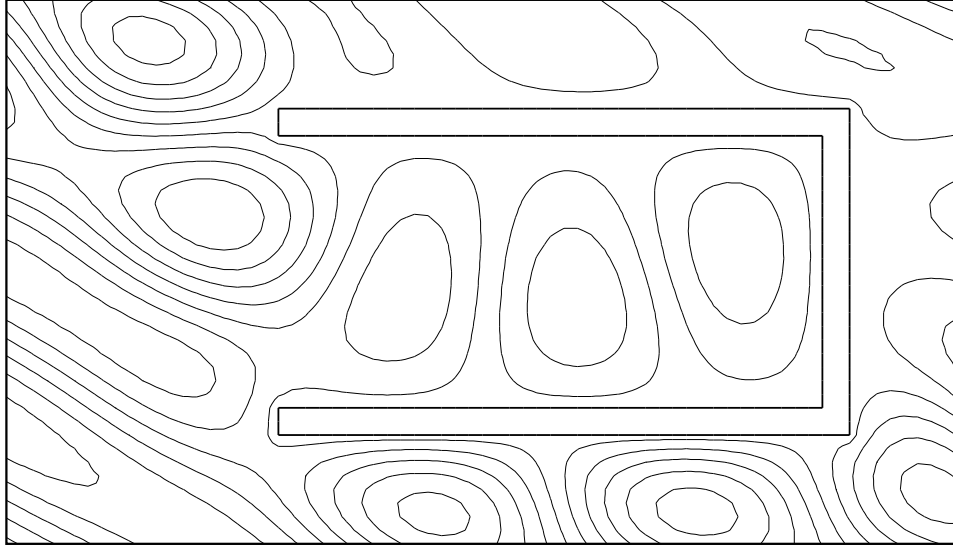


Figure 2. Test Case C3D_1a
The real part of the total field E_3 in plane (X_1, X_2) (10 isolines)

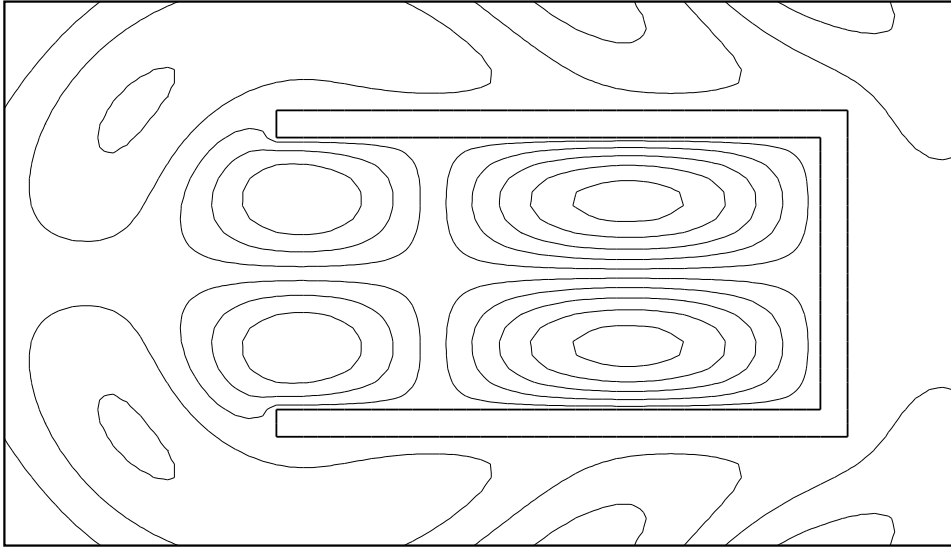


Figure 3. Test Case C3D.1a
The imaginary part of the total field E_1 in plane (X_2, X_3) (10 isolines)

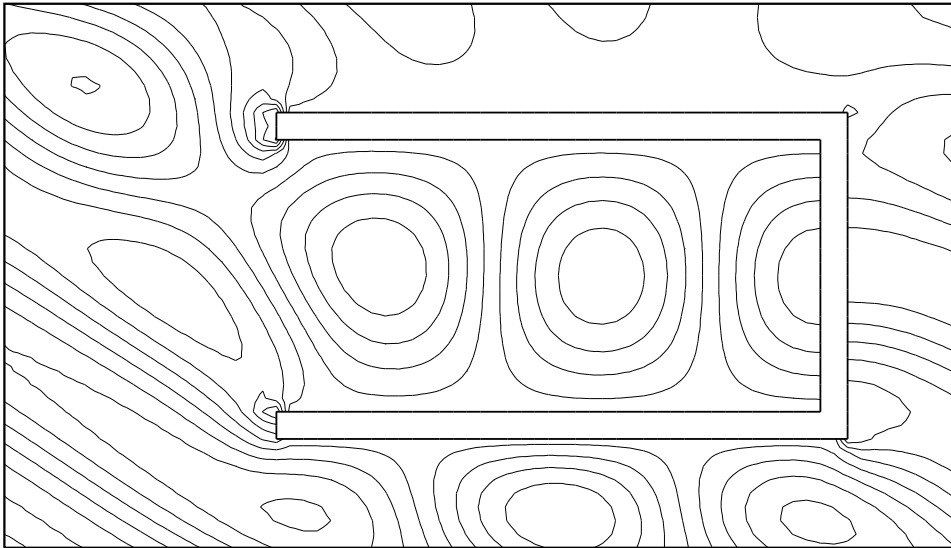


Figure 4. Test Case C3D.1b
The imaginary part of the total field E_1 in plane (X_1, X_2) (10 isolines)

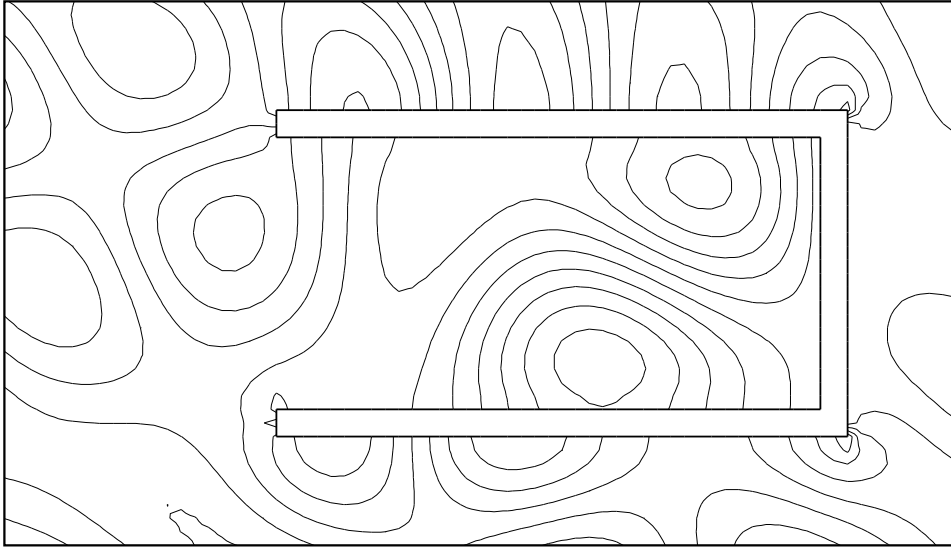


Figure 5. Test Case C3D_1b
The real part of the total field E_2 in plane (X_1, X_2) (10 isolines)

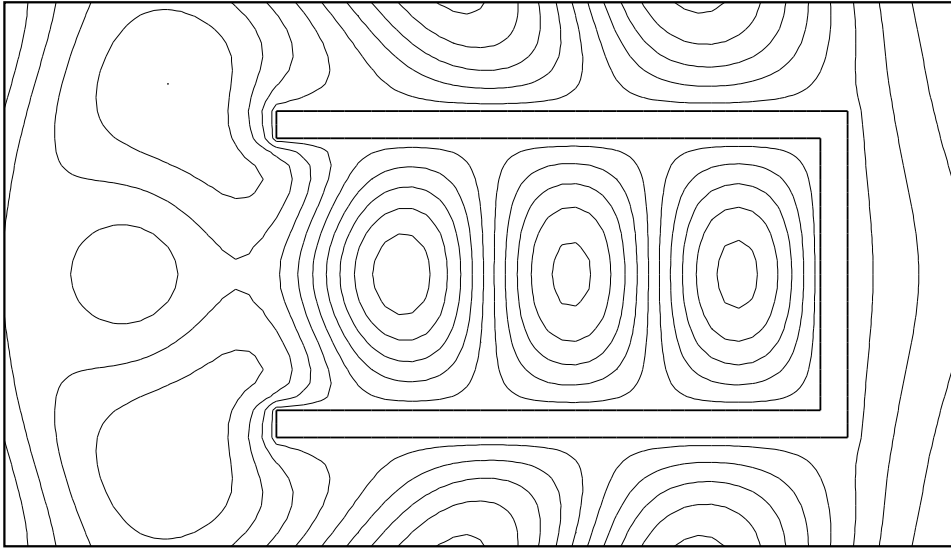


Figure 6. Test Case C3D_1b
The imaginary part of the total field E_1 in plane (X_2, X_3) (10 isolines)

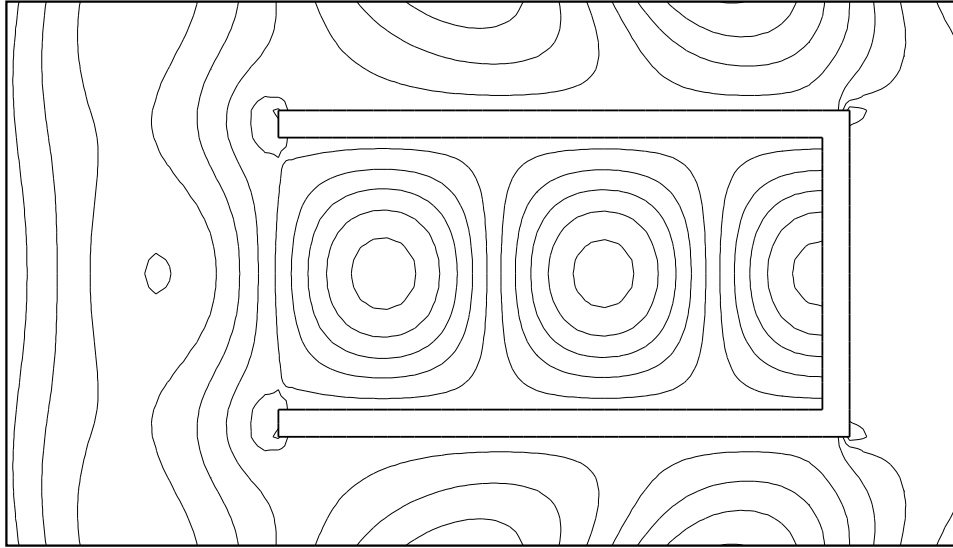


Figure 7. Test Case C3D.1b
The imaginary part of the total field E_2 in plane (X_2, X_3) (10 isolines)

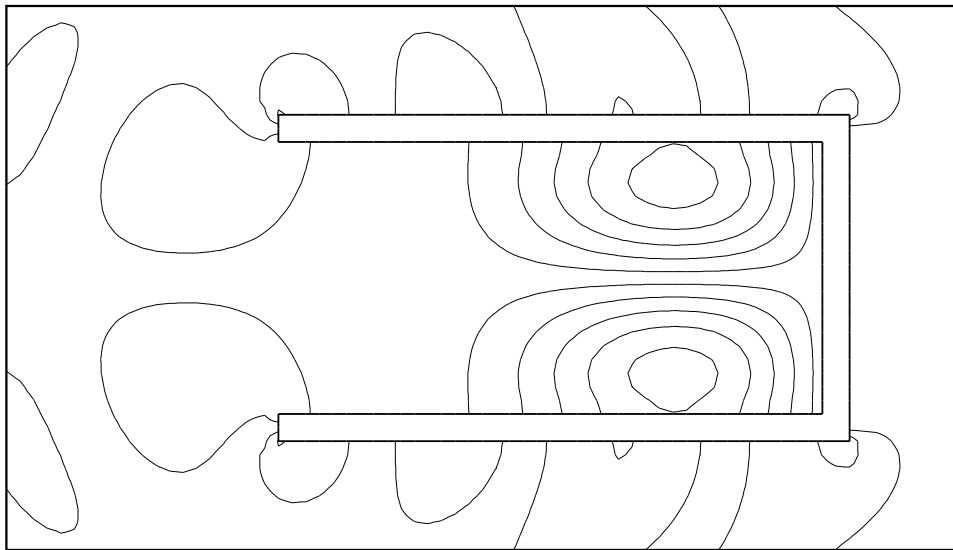


Figure 8. Test Case C3D.1b
The imaginary part of the total field E_3 in plane (X_2, X_3) (10 isolines)

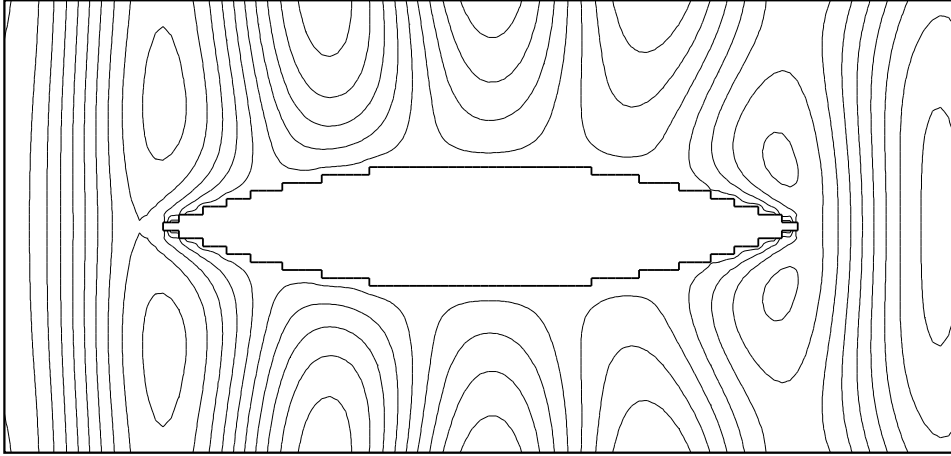


Figure 9. Test Case O3D_2b
The real part of the total field E_1 in plane (X_2, X_3) (10 isolines)

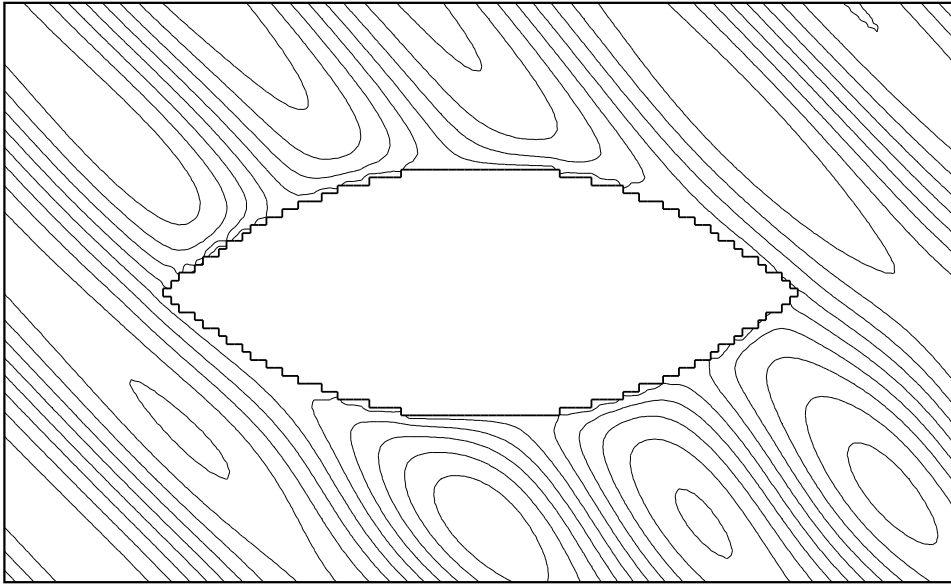


Figure 10. Test Case O3D_2c
The real part of the total field E_3 in plane (X_1, X_2) (10 isolines)

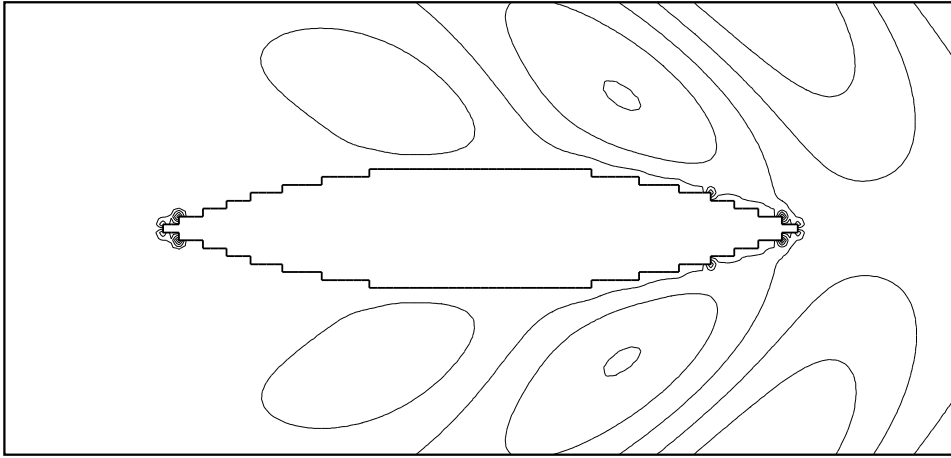


Figure 11. Test Case O3D_2c

The imaginary part of the total field E_1 in plane (X_2, X_3) (10 isolines)

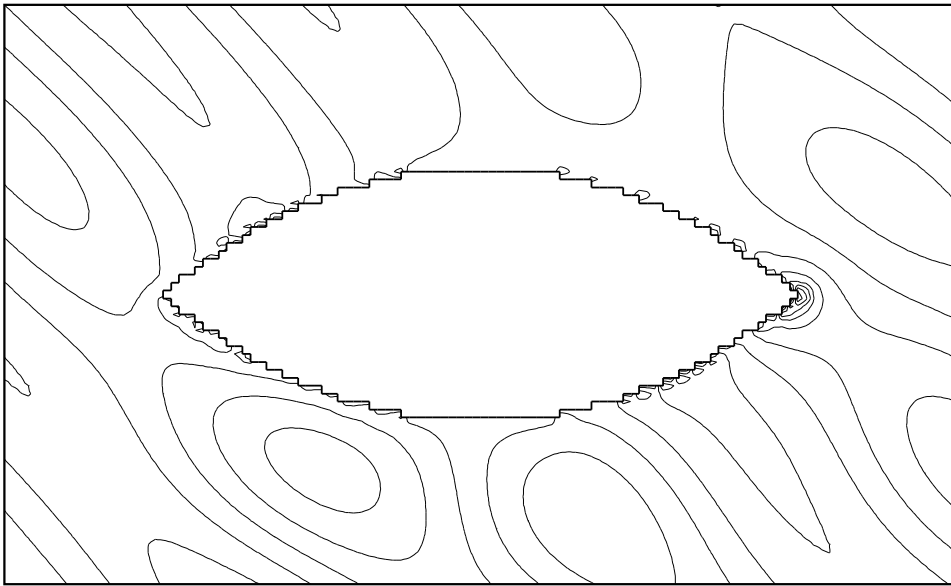


Figure 12. Test Case O3D_2d

The real part of the total field E_1 in plane (X_1, X_2) (10 isolines)

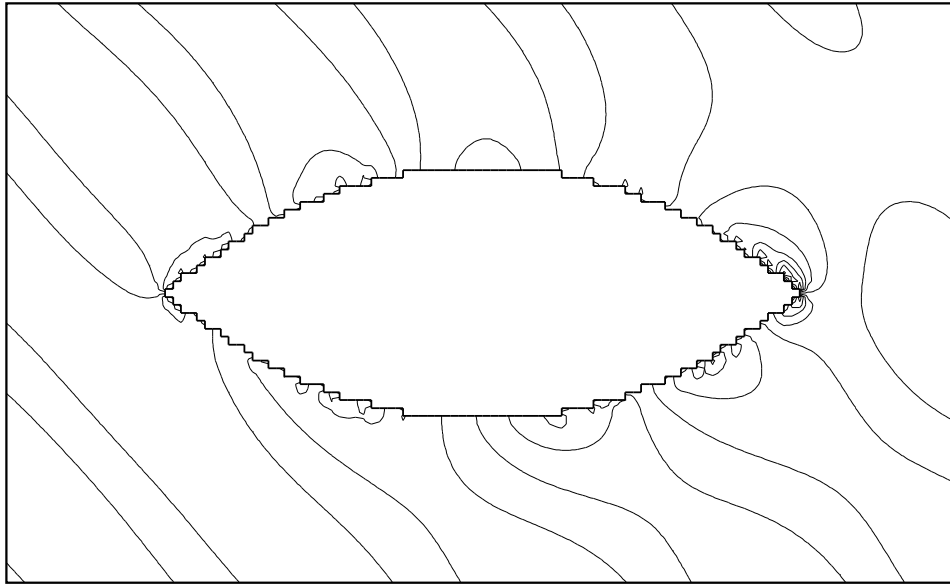


Figure 13. Test Case O3D_2d

The real part of the total field E_2 in plane (X_1, X_2) (10 isolines)

References

- [1] G.P. Astrakhantsev, “Methods of fictitious domain for a second order elliptic equation with natural boundary conditions”, *USSR Computational Math. and Math. Phys.* (1978) **18**, pp. 114-121
- [2] C. Atamian, J. Périaux and R. Glowinski, “Solving the Helmholtz equation on unbounded 3D regions by control/fictitious domain methods”, in: *Second World Congress on Computational Mechanics*, Stuttgart, 1990
- [3] C. Atamian, Q. V. Dinh, R. Glowinski, Jiwen He and J. Périaux, “Control approach to fictitious domain methods. Application to Fluid Dynamics and Electro-Magnetics”, in: *Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations*, Eds. R. Glowinski, Yu. Kuznetsov, G. Meurant, J. Périaux and O. B. Widlund, SIAM, Philadelphia, 1991, pp. 275-309

-
- [4] C. Atamian, Q. V. Dinh, R. Glowinski, Jiwen He and J. Périaux, “On some imbedding methods applied to fluid dynamics and electro-magnetics”, *Comp. Methods in Appl. Mech. and Eng.* (1991) **91**, pp. 1271-1299
 - [5] N. S. Bakhvalov. Numerical Methods, V. 1. Moscow, Nauka, 1973 (in Russian)
 - [6] A. Banegas, “Fast Poisson solvers for problems with sparsity”, *Math. Comp.*, (1978) **3**, pp. 441-446
 - [7] A. Besspalov, “Finite element method for the eigenmode problem of a RF cavity resonator”, *Sov. J. Numer. Anal. and Math. Modelling* (1988) **3**, pp. 163-178
 - [8] A. Besspalov. Application of fictitious domain method to solving the Helmholtz equation in unbounded domain. *Preprint of INRIA No. 1797*, 1992
 - [9] A. Besspalov, “Solution of the Workshop Test Cases for Steady Diffraction Problems by Fictitious Domain Method”, *Proceedings of the 3rd Workshop on Approx. and Num. Methods for the Solution of the Maxwell Eq., Oxford, UK, March 23-24, 1995*, Oxford University Computing Laboratory, 1995, 10 p.
 - [10] A. Besspalov and Yu. Kuznetsov, “RF cavity computations using the domain decomposition and fictitious component methods”, *Sov. J. of Numer. Anal. and Math. Modelling* (1987) **2**, pp. 259-278
 - [11] A. Besspalov, Yu. Kuznetsov, O. Pironneau and M.-G. Vallet, “Fictitious domains with separable preconditioners versus unstructured adapted meshes”, *IMPACT of Computing in Science and Engineering* (1992) **4**, pp. 217-249
 - [12] A. Besspalov, Yu. Kuznetsov, “Fictitious Domain Method for Solving the Helmholtz Equation”, *Proc. of the 2nd Int. Workshop on Approx. and Num. Methods for the Solution of the Maxwell Eq.* Washington, DC, October 28-29, 1993, *to appear*
 - [13] C. Börgers and O. Widlund. Finite element capacitance matrix methods. *Tech. Rep. 261*, New York University, Dept. of Computer Science, 1986
 - [14] A. Bossavit, “Simplicial finite elements for scattering problems in electromagnetism”, *Comp. Meth. Appl. Mech. Eng.* (1989) **76** pp. 299-316
 - [15] M. O. Bristeau, R. Glowinski and J. Périaux, “Exact controllability methods for the Helmholtz equations”, *Proceedings of the 4th International Conference on Hyperbolic Problems*, Taormina, Italy, April 1992

- [16] J.E.Bussoletti, F.T. Johnson, S.S. Samant, D.P. Young and R.H. Burkhart, "EM-TRANAIR: steps toward solution of general 3D Maxwell's equations", *Proc. of the 10th International Conference on Computing Methods in Applied Sciences and Engineering*, 11-14 February of 1992, Paris (Le Vésinet), France, Ed. R. Glowinski, INRIA - Nova Science Publishers, Inc., New York, 1992
- [17] F. El Dabaghi and J. Tuomela, "2D finite element solution of the wave equation by absorbing boundary conditions", in: *Hydraulic engineering software applications*, Eds. W.R. Blain and D. Ouazar, Computational Mechanics Publications, 1990
- [18] B. Despres. Une méthode de décomposition de domaine pour le problème de Helmholtz. *Preprint of INRIA No. 1524*, 1991 (in French)
- [19] Q. V. Dinh, R. Glowinski, Jiwen He, V. Kwok, T. W. Pan and J. Périaux, "Lagrange multiplier approach to fictitious domain methods: application to Fluid Dynamics and Electro-Magnetics", *Proc. of the 5th International Symposium on Domain Decomposition Methods for Partial Differential Equations*, Eds. R. Glowinski, Yu. Kuznetsov, G. Meurant, J. Périaux and O.B. Widlund. SIAM, Philadelphia, 1992
- [20] L. W. Ehrlich and D. M. Silver, "Iterative approaches in solving a 3D complex vector Helmholtz equation", *Proc. of Copper Mountain Conference on Iterative Methods* **2**, Copper Mountain, Colorado, 1992
- [21] B. Engquist and A. Majda, "Absorbing boundary conditions for the numerical simulation of waves", *Math. Comp.* (1977) **31**, pp. 629-651
- [22] B. Engquist and A. Majda, "Radiation boundary conditions for acoustic and elastic wave calculations", *Comm. Pure Appl. Math.* (1979) **32**, pp. 313-357
- [23] Yu. A. Kuznetsov, "Matrix computational processes in subspaces", *Comp. Math. in Appl. Sci. and Eng.* (1984) pp. 15-31
- [24] Yu. A. Kuznetsov, "Matrix iterative methods in subspaces", *Proc. Int. Congress Math.* (1984), pp. 1509-1521
- [25] Yu. A. Kuznetsov and K. N. Lipnikov, "Fictitious domain method for solving the Helmholtz wave equation for unbounded domain", *Numerical Methods and Mathematical Modelling*, pp. 56-69, Institute of Numerical Mathematics of Russian Academy of Sciences, Moscow, 1992 (in Russian)

-
- [26] Yu. A. Kuznetsov and K. N. Lipnikov, "Fictitious Domain Methods for 3D Acoustic Problem", *Proc. of the 2nd Int. Conf. on Approx. and Num. Methods for the Solution of the Maxwell Eq.* Washington, DC, October 25-27, 1993, *to appear*
 - [27] Yu. A. Kuznetsov and A. Matsokin, "On partial solution of systems of linear algebraic equations", in: *Computational Methods of Linear Algebra*, Ed. G. I. Marchuk, Novosibirsk Comp. Centre of the USSR Acad. Sci., Novosibirsk, 1980, pp. 62-89 (in Russian) English translation in: *Sov. J. Numer. Anal. and Math. Modelling* (1989) **4**, pp. 453-468
 - [28] D. P. O'Leary and O. Widlund, "Capacitance matrix methods for Helmholtz equation on general three-dimensional regions", *Math. Comp.* (1979) **3**, pp. 849-879
 - [29] V. I. Lebedev, "Finite-difference analogs of orthogonal expansions, basic differential operators and some boundary problems of mathematical physics", *J. Vych. Matem. i Matem. Fiz.* (1964) **3**, pp. 449-465, **4**, pp. 649-659 (in Russian)
 - [30] G. I. Marchuk, Yu. A. Kuznetsov and A. M. Matsokin, "Fictitious domain and domain decomposition methods", *Sov. J. Numer. Anal. Math. Modelling* (1986) **1**, pp. 3-35
 - [31] J. C. Nédélec, "Approximation des equations integrales en mecanique et en physique", *Cours de l'ecole d'ete d'analyse numerique*, EDF-CEA-INRIA, 1977 (in French)
 - [32] J. C. Nédélec, "Mixed finite elements in \mathbf{R}^3 ", *Numerische Mathematik*, (1980) **35**, pp. 315-341
 - [33] W. Proskurowski and O. B. Widlund, "On the numerical solution of Helmholtz's equation by the capacitance matrix method", *Math. Comp.* (1976) **30**, pp. 433-468
 - [34] G. T. Ruck, D. E. Barrick, W. D. Stuart, C. K. Krichbaum. Radar Cross Section Handbook. Vol. 1. Plenum Press, New York – London, 1970.
 - [35] T. Weiland, "On the numerical solution of Maxwell's equations and application in the fields of accelerator physics", *Particle Accelerators* (1984) **15**, pp. 245-292



Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irista, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399