



A Symmetric Trefftz-DG formulation based on a local boundary element method for the solution of the Helmholtz equation

Hélène Barucq, Abderrahmane Bendali, M'Barek Fares, Vanessa Mattesi,
Sébastien Tordeux

► To cite this version:

Hélène Barucq, Abderrahmane Bendali, M'Barek Fares, Vanessa Mattesi, Sébastien Tordeux. A Symmetric Trefftz-DG formulation based on a local boundary element method for the solution of the Helmholtz equation. *Journal of Computational Physics*, 2017, 330, pp.1069-1092. 10.1016/j.jcp.2016.09.062 . hal-01395861

HAL Id: hal-01395861

<https://hal.science/hal-01395861>

Submitted on 16 Jan 2018

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.

A Symmetric Trefftz-DG Formulation based on a Local Boundary Element Method for the Solution of the Helmholtz Equation.

H. Barucq^{b,a,*}, A. Bendali^{d,c}, M. Fares^c, V. Mattesi^{b,a,c}, S. Tordeux^{b,a,c}

^aINRIA Bordeaux Sud-Ouest Magique-3D Team, Pau (France)

^bUniversity of Pau (France)

^cCERFACS Algo Team, Toulouse (France)

^dUniversity of Toulouse, Mathematical Institute of Toulouse, INSA, Toulouse (France)

Abstract

A general symmetric Trefftz Discontinuous Galerkin method is built for solving the Helmholtz equation with piecewise constant coefficients. The construction of the corresponding local solutions to the Helmholtz equation is based on a boundary element method. A series of numerical experiments displays an excellent stability of the method relatively to the penalty parameters, and more importantly its outstanding ability to reduce the instabilities known as the “*pollution effect*” in the literature on numerical simulations of long-range wave propagation.

Keywords: Helmholtz equation, pollution effect, dispersion, Trefftz method, Discontinuous Galerkin method, Boundary Element Method

1. Introduction

When used for solving the Helmholtz equation over several hundreds of wavelengths, usual Finite Element Methods (FEM) are faced with the drawback generally called “*pollution effect*”. Roughly speaking, it is necessary to augment the density of nodes to maintain a given level of accuracy, when increasing the size of the computational domain. This in turn rapidly exceeds the capacities in storage and computing even in the framework of massively parallel computer platforms (cf., for example, [1, 2, 3] and the references therein).

Several approaches have been proposed to cure this flaw. At first, for such kinds of numerical solutions, it turns out that Discontinuous Galerkin (DG) methods are more efficient than standard FEMs, also called Continuous Galerkin (CG) methods. This efficiency seems to be due in part to the less strong inter-element continuity characterizing these methods (cf., for example, [4, 5]). This was confirmed in [6] where it is shown that

*Corresponding author

Email addresses: helene.barucq@inria.fr (H. Barucq), abendali@insa-toulouse.fr (A. Bendali), fares@cerfacs.fr (M. Fares), vanessa.mattesi@inria.fr (V. Mattesi), sebastien.tordeux@inria.fr (S. Tordeux)

Preprint submitted to Elsevier

October 4, 2016

it is possible to keep the efficiency of the DG methods by allowing discontinuities only at the interior of the elements in terms of bubble functions with penalized jumps.

Another advantage of DG methods lies in the opportunity to use shape functions, more adapted to the approximation of the solution to the interior Partial Differential Equations (PDE) of the problem, but, contrary to polynomials, with poor properties for enforcing the usual inter-element continuity conditions of the FEM. In this respect, Trefftz methods, or in other words methods for which the local shape functions are wave functions (cf., for example, [7, 8] and the references therein), were intensively used to alleviate the aforementioned “*pollution effect*”. The combination of a Trefftz method with a DG one therefore resulted in numerous approaches for solving wave equation problems called Trefftz DG method (TDG) (see, for example, [9, 10, 11, 7] and the references therein).

Actually, Trefftz methods without strong inter-element continuity were used for some time now in the context of the so called Ultra Weak Variational Formulation (UWVF) devised by Després [12, 13]. It was discovered later that this formulation can be recast in the context of a TDG method [14, 15, 9].

Some criticisms have been however addressed to the DG methods. They mainly concern the increase of the coupled degrees of freedom and a suboptimal convergence of their approximate fluxes. Hybridized versions of the DG (HDG) methods were proposed in response to these challenges [16]. However to our knowledge, HDG methods have not been used yet in the framework of a Trefftz method but only with usual local polynomial approximations [17], except in a recent paper [18], where these methods were combined in an elaborate way with geometrical optics at the element level to efficiently solve the Helmholtz equation in the high frequency regime. Since the local shape functions are only asymptotic solutions to the Helmholtz equation then, such a kind of method can be called quasi-Trefftz HDG. It is worth noting that this kind of methods cannot handle the low and the mid frequency regimes.

Instead of DG methods, some authors prefer to use a Lagrange multiplier or a least-square technique to enforce the continuity conditions (cf. [19, 20, 21]). These continuity conditions are also ensured from consistency and uniqueness arguments in the Variational Theory of Complex Ray (VTCR) method (cf. [22, 7]). We have not retained these approaches in this paper, mainly because we were not convinced on the capabilities of these methods, not based on a global handling of the problem equations, to efficiently face the “*pollution effect*”. However, further investigations are necessary to confirm this claim.

On the other hand, it is generally admitted that Boundary Integral Equations (BIE) lead to less “*pollution effect*” than FEMs even if to our knowledge, no formal study has confirmed such a property. Such a good behavior is probably due to the fact that BIEs can be seen as particular Trefftz methods, posed in one element, when such an interpretation is taken to the extreme. It is hence tempting to use the free space Green kernel in an approximation procedure for the interior Helmholtz equation to reduce the “*pollution effect*”. This way to proceed has been already considered in [23]. However it relies upon the construction of a uniform grid in a homogeneous medium of propagation, which is manageable with very special boundaries and boundary conditions only. It seems thus difficult to extend this approach to problems involving varying coefficients or realistic geometries and boundary conditions. The aim of this study is precisely to mix two approaches by combining DG methods with BIEs, to devise a TDG method which can

efficiently handle particular Helmholtz equations with varying coefficients. Specifically, either the coefficients of the Helmholtz equation are piecewise constant or they can be approximated by piecewise constant functions on a sufficiently refined decomposition of the computational domain, called *DG formulation mesh* in the rest of this paper.

The idea of building a FEM in which local shape functions are obtained on the basis of a BIE has been recently investigated in [24, 25]. Our work stands out from these approaches in using DG framework and an improved approximation of the Dirichlet-to-Neumann (DtN) operator. The latter is a key feature here for matching the local solutions at the interfaces of the mesh. The corresponding method can be viewed globally as a DG method at the level of the DG formulation mesh and locally as a BIE at the element level. Actually, BIEs are used only to compute the DtN operator within each element of the DG formulation mesh. As shown below, the quality of the overall solution strongly depends on the accuracy of the approximation of the DtN operator. Specific numerical procedures have therefore been developed to increase the accuracy of this approximation. Similar techniques were considered in [26, 27]. Our method is also formulated as a symmetric variational formulation of the corresponding boundary-value problem. Its derivation is inspired by [28] (see also [29, p. 122]) where Symmetric Interior Penalty (SIP) methods have been designed. The symmetry yields an important gain. The storage of the boundary integral operators involved in the formulation is indeed avoided, the contribution of the BIEs being element-wise only. It is also worth noting that all the degrees of freedom of the discrete problem to be solved are located on the “skeleton” of the mesh, that consists of the boundaries of the elements of this mesh. Such a feature moves towards HDG methods even if it keeps with unknowns on both sides of the interfaces. Note as well that our choice of the local approximating functions allows the approximation of both the propagating and evanescent waves in a natural way. It should be noted also that, even if the method, considered here, is of Trefftz type, the local approximations are done by means of a Boundary Element Method (BEM) (cf., for example, [30, 31]). As a result, these approximations are ultimately performed in terms of piecewise polynomial functions on a BEM mesh. In contrast then to usual Trefftz methods (cf., for example, [19, 11, 32, 9, 13] to cite a few), h or p refinements are as simple and efficient as in a standard FEM. In the following, this method is called the BEM Symmetric Trefftz DG method and more concisely denoted by BEM-STDG.

The paper is organized as follows. In Section 2, after stating the boundary-value problem, we first derive the variational formulation of the symmetric TDG method and show how it can be connected to previous DG formulations. Section 3 develops the BEM procedure used to define the Trefftz method. Section 4 is devoted to the numerical validation of the method in two dimensions, while Section 5 deals with the comparison of its performances with a standard Interior Penalty DG (IPDG) method based on element-wise polynomial approximations. A final brief section gives some concluding remarks and indicates further studies that can extend the current one.

2. The symmetric Trefftz DG method

After stating the wave propagation problem, we describe the most general DG formulation considered in this study.

2.1. The Helmholtz boundary-value problem

The DG variational formulations of the Helmholtz equation (cf., for example, [7, 9, 10]) are generally obtained by writing the wave equation in the form of a first-order PDE system. Most of the studies dedicated to the DG solution of this problem (in addition to the previous references, see, for example, [13, 15, 33, 3]) deal with the Helmholtz equation with constant coefficients. When focusing on acoustics, the fluctuations of pressure and velocity then correspond to the propagation of an acoustic wave in an ideal stagnant and uniform fluid (cf., for example, [34, Chap. 2]). We follow here a more general path and consider as in [35] that the propagation is related to an ideal stagnant fluid but not necessarily uniform. The acoustic system for such a configuration can be written as follows (cf., for example, [36, Eqs. (64.5) and (64.3)])

$$\begin{cases} \frac{1}{c^2 \varrho} \partial_t p + \nabla \cdot \mathbf{v} = 0, \\ \varrho \partial_t \mathbf{v} + \nabla p = 0, \end{cases} \quad (1)$$

where c and ϱ are the speed of sound and the density within the stagnant fluid and p and \mathbf{v} are respectively the acoustic fluctuations of the pressure and the velocity. Hereafter data c and ϱ are assumed to be piecewise positive constants. As this will be clear below, the handling of the related discontinuities is an important feature of the DG formulation.

To be consistent with the notation used in previous works [7, 9, 10, 33, 3], we denote the phasors of the pressure and the velocity by a different symbol: u for p and $\boldsymbol{\sigma}$ for \mathbf{v} defined according to the following identities and characterizations

$$p(x, t) = \Re(e^{-i\omega t} u(x)), \quad \mathbf{v}(x, t) = \Re(e^{-i\omega t} \boldsymbol{\sigma}(x)). \quad (2)$$

In the above definitions, $\Re z$ is the real part of the complex number z , and $\omega > 0$ is the angular frequency. The solution of (1) is hence reduced to

$$\begin{cases} -\frac{i\omega}{c^2 \varrho} u + \nabla \cdot \boldsymbol{\sigma} = 0, \\ -i\omega \varrho \boldsymbol{\sigma} + \nabla u = 0. \end{cases} \quad (3)$$

We now assume that the equations are set in a bounded polygonal/polyhedral domain $\Omega \subset \mathbb{R}^d$ ($d = 2, 3$) and denote by $\partial\Omega$ its boundary. Using the piecewise constant wave number $\kappa = \omega/c$, and considering a non overlapping decomposition $\partial\Omega_D$, $\partial\Omega_N$, and $\partial\Omega_R$ of $\partial\Omega$, we recast the above system as the following Helmholtz equation with varying coefficients supplemented with typical boundary conditions

$$\begin{cases} \nabla \cdot \frac{1}{\varrho} \nabla u + \frac{\kappa^2}{\varrho} u = 0 & \text{in } \Omega, \\ u = g_D & \text{on } \partial\Omega_D, \\ \frac{1}{\varrho} \nabla u \cdot \mathbf{n} = g_N & \text{on } \partial\Omega_N, \\ \frac{1}{\varrho} \nabla u \cdot \mathbf{n} - iY u = g_R & \text{on } \partial\Omega_R. \end{cases} \quad (4)$$

The third boundary condition is expressed in terms of a function Y yielding the surface compliance of $\partial\Omega_R$ up to a multiplicative constant, assumed to be also piecewise constant. The sources producing the wave are embodied in the right-hand sides g_D , g_N and g_R . We have denoted by \mathbf{n} the unit normal on $\partial\Omega$ outwardly directed to Ω (see FIG. 1).

If the right-hand sides g_D , g_N and g_R are respectively in $H^{1/2}(\partial\Omega_D)$, in $L^2(\partial\Omega_N)$ and in $L^2(\partial\Omega_R)$ and assuming furthermore that $\Re Y \geq \nu > 0$ on a part of $\partial\Omega_R$ with a

non vanishing length/area, it is well-known that problem (4) admits one and only one solution u in $H^1(\Omega)$ (cf., for example, [37, 38]).

The Helmholtz equation with variable coefficients in system (4) is exactly the wave equation considered in [35], where the boundary condition is written as

$$\frac{1}{\varrho} \partial_{\mathbf{n}} u - i\eta u = Q \left(-\frac{1}{\varrho} \partial_{\mathbf{n}} u - i\eta u \right) + g \quad (5)$$

with $\eta = \kappa/\varrho$ and Q and g given here by

$$\begin{cases} Q = -1, & g = -2i\eta g_D, & \text{on } \partial\Omega_D, \\ Q = +1, & g = 2g_N, & \text{on } \partial\Omega_N, \\ Q = (1 - Y/\eta) / (1 + Y/\eta), & g = (1 + Q)g_R, & \text{on } \partial\Omega_R, \end{cases} \quad (6)$$

thus expressing the three boundary conditions in (4) in terms of a single one. This is more than another way of writing the boundary conditions. It makes it possible to express the “incoming wave” $(1/\varrho) \partial_{\mathbf{n}} u - i\eta u$ in terms of a reflection of the “outgoing wave” $Q(-(1/\varrho) \partial_{\mathbf{n}} u - i\eta u)$ and a source term g .

It is worth noting however that Helmholtz equation is involved in other kinds of wave propagation problems. An important example of these is related to seismic waves where attenuation effects must be accounted for in addition to the propagation features. The Helmholtz equation governing this kind of waves is then in the following form [39]

$$\Delta u + \frac{\varrho \omega^2}{E} u = 0 \quad (7)$$

where ϱ and ω are the density and the angular frequency and E is the complex modulus of the medium. Clearly this equation can be put in the above setting by substituting $1/c^2$ for ϱ/E and ϱ for 1 in system (3). This thus leads to a complex wave number κ . Since we are mainly interested on accurately accounting for long-range propagation, we limit ourselves below to real coefficients.

2.2. The variational formulation

2.2.1. The DG formulation mesh

At first, we consider a non overlapping decomposition \mathcal{T} of Ω in polyhedral/polygonal subdomains of the computational domain Ω , called the DG formulation mesh.

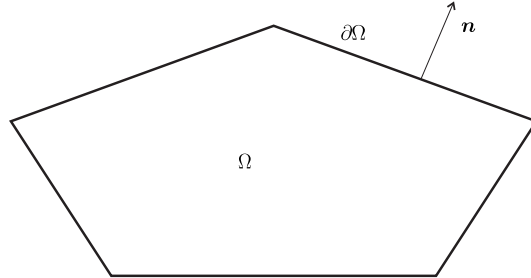


Figure 1: Schematic view of the computational domain.

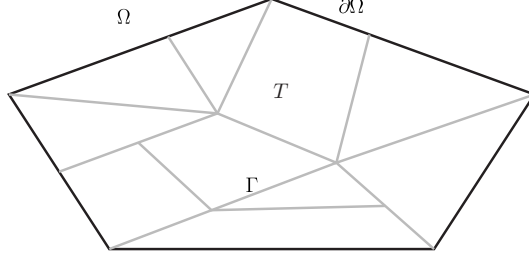


Figure 2: The DG formulation mesh in 2D.

Considering that the elements $T \in \mathcal{T}$ are open sets of \mathbb{R}^d , we therefore assume that

$$\bar{\Omega} = \bigcup_{T \in \mathcal{T}} \bar{T}, \quad T \cap K = \emptyset \quad \text{if } T \neq K.$$

It is worth recalling that the DG formulation mesh can be quite arbitrary. An example of a suitable mesh in the two-dimensional case is depicted in FIG. 2.

We always assume that the coefficients ϱ and κ of the Helmholtz equation are real positive constants within each $T \in \mathcal{T}$ and denoted there by ϱ_T and κ_T respectively.

We move on defining the interior and boundary edges/faces which play an important role in the DG methods. An interior edge/face F is a special part of the boundary ∂T of $T \in \mathcal{T}$ shared by the boundary ∂K of another $K \in \mathcal{T}$. It is defined as follows

$$F = \partial T \cap \partial K \text{ when the length/area of } F \text{ is } > 0. \quad (8)$$

Some other definitions characterize F by requiring that it contains at least d points constituting a non degenerated simplex (segment and triangle in the two- and the three-dimensional case respectively) [28]. Boundary edges/faces F are defined similarly by replacing K with the exterior of Ω . We use a set notation \mathcal{F}_I and \mathcal{F}_∂ to refer to the collections of interior and boundary edges/faces respectively. We are hence led to define the skeleton

$$\mathcal{F} = \mathcal{F}_I \cup \mathcal{F}_\partial \quad (9)$$

of the DG formulation mesh \mathcal{T} , which as for some other TDG methods (see, for example, [7]) play an important role in the construction of the method proposed in this study.

Clearly the interior edges/faces F constitute a non-overlapping decomposition of the following curve/surface

$$\Gamma = \bigcup_{F \in \mathcal{F}_I} F.$$

In the same way, the boundary edges/faces yield a non-overlapping decomposition of $\partial\Omega$

$$\partial\Omega = \bigcup_{F \in \mathcal{F}_\partial} F.$$

In FIG. 2, Γ is depicted in grey while $\partial\Omega$ is in black.

2.2.2. Basic principle of the variational formulation

For a test function v in $H^1(\Omega)$, since the solution u to the boundary-value problem (4) is sought in $H^1(\Omega)$ and satisfies $\nabla \cdot 1/\varrho \nabla u = -(\kappa^2/\varrho)u$ in Ω , yielding that $\nabla \cdot 1/\varrho \nabla u$ is in $L^2(\Omega)$, we can use the following Green formula to get

$$\begin{aligned} \sum_{T \in \mathcal{T}} \int_T \left(\frac{1}{\varrho_T} \nabla u \cdot \nabla v - \frac{\kappa_T^2}{\varrho_T} uv \right) dx \\ = \sum_{T \in \mathcal{T}} \int_T \left(\frac{1}{\varrho_T} \nabla u \cdot \nabla v + v \nabla \cdot \frac{1}{\varrho_T} \nabla u \right) dx \\ = \sum_{T \in \mathcal{T}} \left\langle \left(\frac{1}{\varrho} \nabla u \right) |_{\partial T} \cdot \mathbf{n}_T, v_T \right\rangle_{\partial T} \end{aligned} \quad (10)$$

with

$$v_T = v|_{\partial T} \quad (11)$$

the trace being taken from the values of v inside T , and $\langle \cdot, \cdot \rangle_{\partial T}$ the bilinear form underlying the duality product $H^{-1/2}(\partial T) \times H^{1/2}(\partial T)$ which reduces to the integral

$$\left\langle \left(\frac{1}{\varrho} \nabla u \right) |_{\partial T} \cdot \mathbf{n}_T, v_T \right\rangle_{\partial T} = \int_{\partial T} \left(\frac{1}{\varrho} \nabla u \right) |_{\partial T} \cdot \mathbf{n}_T v_T ds \quad (12)$$

when $\left(\frac{1}{\varrho} \nabla u \right) |_{\partial T} \cdot \mathbf{n}_T$ is in $L^2(\partial T)$ (cf., for example, [40]). Vector \mathbf{n}_T is the unit normal to ∂T directed outwards T .

In the context of DG methods, the $L^2(\partial T)$ regularity in formula (12) is of fundamental importance. Actually, this regularity is ensured by either of the two following equivalent assumptions: $u_T \in H^1(\partial T)$ or $u \in H^{3/2}(T)$ (cf., for example, [40, Th. 4.24 and p. 209]). Thus, when setting a DG method, two interpretations are equally possible: assume that the variational solution u owns the extra regularity $u \in H^{3/2}(T)$ for all $T \in \mathcal{T}$ or that u is approximated in each T by a function of an *a priori* assumed smoothness. We choose the first alternative in the rest of this paper: u is sought in the following space

$$X_{\mathcal{T}} = \left\{ v \in L^2(\Omega); \Delta v + \kappa_T^2 v = 0 \text{ in } T, v|_T \in H^{3/2}(T), \forall T \in \mathcal{T} \right\}$$

simultaneously ensuring that $u_T \in H^1(\partial T)$ and $(1/\varrho_T \nabla u) |_{\partial T} \cdot \mathbf{n}_T \in L^2(\partial T)$.

Actually below, we can take advantage of the fact that the normal component of $1/\varrho \nabla u$ is continuous across F to write the sum of the integrals on each ∂T in the following manner

$$\sum_{T \in \mathcal{T}} \int_{\partial T} \left(\frac{1}{\varrho} \nabla u \right) |_{\partial T} \cdot \mathbf{n}_T v_T ds = \sum_{F \in \mathcal{F}_{\mathcal{T}}} \int_F \frac{1}{\varrho} \nabla u \cdot \llbracket v \rrbracket ds + \sum_{F \in \mathcal{F}_{\partial}} \int_F \frac{1}{\varrho} \nabla u \cdot \mathbf{n} v ds \quad (13)$$

using the widespread notation (cf., for example, [41]) for the jump of v across F

$$\llbracket v \rrbracket = \mathbf{n}_T v_T + \mathbf{n}_K v_K, \quad (14)$$

T and K being the two elements of the mesh sharing edge/face F .

It is part of the derivation of the variational formulation of the DG method to express the continuity of the normal component of $1/\varrho \nabla u$ across any edge/face F from the mean of its traces on both sides of F

$$\llbracket \frac{1}{\varrho} \nabla u \rrbracket = \frac{1}{2} \left(\left(\frac{1}{\varrho} \nabla u \right) |_{\partial T} + \left(\frac{1}{\varrho} \nabla u \right) |_{\partial K} \right) \quad (15)$$

thus arriving to

$$\sum_{T \in \mathcal{T}} \int_T \frac{1}{\varrho T} (\nabla u \cdot \nabla v - \kappa_T^2 uv) dx = \int_{\Gamma} \llbracket \frac{1}{\varrho} \nabla u \rrbracket \cdot \llbracket v \rrbracket ds + \int_{\partial \Omega} \frac{1}{\varrho} \nabla u \cdot \mathbf{n} v ds. \quad (16)$$

2.2.3. General variational formulation of the symmetric Trefftz-DG method

In the same way, assuming now that test function v is also an element of $X_{\mathcal{T}}$ and based this once on the fact that the unknown u is continuous across the interfaces F , we can write

$$\sum_{T \in \mathcal{T}} \int_T \frac{1}{\varrho T} (\nabla u \cdot \nabla v - \kappa_T^2 uv) dx = \int_{\Gamma} u \llbracket \frac{1}{\varrho} \nabla v \rrbracket ds + \int_{\partial \Omega} u \frac{1}{\varrho} \nabla v \cdot \mathbf{n} ds \quad (17)$$

where the jump $\llbracket 1/\varrho \nabla v \rrbracket$ is defined by

$$\llbracket \frac{1}{\varrho} \nabla v \rrbracket = \left(\frac{1}{\varrho} \nabla v \right) |_{\partial T} \cdot \mathbf{n}_T + \left(\frac{1}{\varrho} \nabla v \right) |_{\partial K} \cdot \mathbf{n}_K. \quad (18)$$

Remark. We have adopted the notation usually used in the DG discretizations to denote by the same symbol the jumps of scalar- and vector-valued functions. The context clearly indicates what kind of jump is considered (cf., for example, [41, 29]).

In the same way as for $1/\varrho \nabla u$, we substitute the mean value

$$\llbracket u \rrbracket = \frac{1}{2} (u_T + u_K), \quad (19)$$

for the trace of u and use (16) to obtain the following variational equation set on the edges/faces of the DG formulation mesh \mathcal{T}

$$\int_{\Gamma} \left(\llbracket u \rrbracket \llbracket \frac{1}{\varrho} \nabla v \rrbracket - \llbracket \frac{1}{\varrho} \nabla u \rrbracket \cdot \llbracket v \rrbracket \right) ds + \int_{\partial \Omega} \left(u \frac{1}{\varrho} \nabla v \cdot \mathbf{n} - \frac{1}{\varrho} \nabla u \cdot \mathbf{n} v \right) ds = 0. \quad (20)$$

To design a symmetric formulation, we proceed as in [28, 2] (see also [29, p. 122]). We make use of the following jump values

$$\llbracket u \rrbracket = 0 \quad \text{and} \quad \llbracket \frac{1}{\varrho} \nabla u \rrbracket = 0 \quad (21)$$

and the boundary conditions to add some consistent terms to Eq. (20) thus arriving to

$$\begin{aligned} & \int_{\Gamma} \left(\llbracket u \rrbracket \llbracket \frac{1}{\varrho} \nabla v \rrbracket + \llbracket \frac{1}{\varrho} \nabla u \rrbracket \llbracket v \rrbracket - \llbracket \frac{1}{\varrho} \nabla u \rrbracket \cdot \llbracket v \rrbracket - \llbracket u \rrbracket \llbracket \frac{1}{\varrho} \nabla v \rrbracket \right) ds \\ & - \int_{\partial \Omega_D} \left(u \frac{1}{\varrho} \nabla v \cdot \mathbf{n} + \frac{1}{\varrho} \nabla u \cdot \mathbf{n} v \right) ds \\ & + \int_{\partial \Omega_N \cup \partial \Omega_R} \left(u \frac{1}{\varrho} \nabla v \cdot \mathbf{n} + \frac{1}{\varrho} \nabla u \cdot \mathbf{n} v \right) ds + 2 \int_{\partial \Omega_R} (-iY) uv ds \\ & = -2 \int_{\partial \Omega_D} g_D \frac{1}{\varrho} \nabla v \cdot \mathbf{n} ds + 2 \int_{\partial \Omega_N} g_N v ds + 2 \int_{\partial \Omega_R} g_R v ds. \end{aligned} \quad (22)$$

To stabilize the formulation, in view of already known DG methods [41, 9, 2], we finally add consistent penalty terms expressed by means of given functions α , β , and δ defined on Γ and $\partial\Omega$. In this way, we arrive to the following most general variational formulation on which are based the TDG methods considered in this paper

$$\begin{cases} u \in X_{\mathcal{T}}, \forall v \in X_{\mathcal{T}}, \\ a(u, v) = Lv, \end{cases} \quad (23)$$

where a is the following symmetric bilinear form

$$\begin{aligned} a(u, v) = & \int_{\Gamma} \left(\llbracket u \rrbracket \llbracket \frac{1}{\varrho} \nabla v \rrbracket + \llbracket \frac{1}{\varrho} \nabla u \rrbracket \llbracket v \rrbracket - \llbracket \frac{1}{\varrho} \nabla u \rrbracket \cdot \llbracket v \rrbracket - \llbracket u \rrbracket \cdot \llbracket \frac{1}{\varrho} \nabla v \rrbracket \right) ds \\ & + \int_{\Gamma} (\alpha \llbracket u \rrbracket \llbracket v \rrbracket + \beta \llbracket \nabla_{\top} u \rrbracket \cdot \llbracket \nabla_{\top} v \rrbracket) ds \\ & - \int_{\partial\Omega_D} \left(u \frac{1}{\varrho} \nabla v \cdot \mathbf{n} + \frac{1}{\varrho} \nabla u \cdot \mathbf{n} v \right) ds \\ & + \int_{\partial\Omega_N \cup \partial\Omega_R} \left(u \frac{1}{\varrho} \nabla v \cdot \mathbf{n} + \frac{1}{\varrho} \nabla u \cdot \mathbf{n} v \right) ds - 2 \int_{\partial\Omega_R} iYuv ds \\ & + \int_{\partial\Omega_D} (\alpha uv + \beta \nabla_{\top} u \nabla_{\top} v) ds + \int_{\partial\Omega_N} \delta \frac{1}{\varrho} \nabla u \cdot \mathbf{n} \frac{1}{\varrho} \nabla v \cdot \mathbf{n} ds \\ & + \int_{\partial\Omega_R} \delta \left(\frac{i}{Y} \frac{1}{\varrho} \nabla u \cdot \mathbf{n} \frac{1}{\varrho} \nabla v \cdot \mathbf{n} + \frac{1}{\varrho} \nabla u \cdot \mathbf{n} v + u \frac{1}{\varrho} \nabla v \cdot \mathbf{n} - iYuv \right) ds, \end{aligned} \quad (24)$$

and where the right-hand side is defined by

$$\begin{aligned} Lv = & \int_{\partial\Omega_D} \left(-2g_D \frac{1}{\varrho} \nabla v \cdot \mathbf{n} + \alpha g_D v + \beta \nabla_{\top} g_D \nabla_{\top} v \right) ds \\ & + \int_{\partial\Omega_N} \left(2g_N v + \delta g_N \frac{1}{\varrho} \nabla v \cdot \mathbf{n} \right) ds + \int_{\partial\Omega_R} 2g_R v + \delta g_R \left(\frac{i}{Y} \frac{1}{\varrho} \nabla v \cdot \mathbf{n} + v \right) ds. \end{aligned} \quad (25)$$

In the above expressions, ∇_{\top} is the tangential gradient and

$$\begin{aligned} \llbracket \nabla_{\top} u \rrbracket \cdot \llbracket \nabla_{\top} v \rrbracket &= \nabla_{\top} (u_T - u_K) \cdot \nabla_{\top} (v_T - v_K) \\ &= \nabla_{\top} (u_K - u_T) \cdot \nabla_{\top} (v_K - v_T) \end{aligned}$$

on any interior edge/face F shared by elements T and K .

Remark. Both of the following penalization terms involve only the Dirichlet traces of the solution and of the test function

$$\int_{\Gamma} \alpha \llbracket u \rrbracket \llbracket v \rrbracket ds, \quad \int_{\Gamma} \beta \llbracket \nabla_{\top} u \rrbracket \cdot \llbracket \nabla_{\top} v \rrbracket ds \quad (26)$$

It is worth noting that the integrals involving the tangential gradient ∇_{\top} are well-defined for $u, v \in X_{\mathcal{T}}$. However, these penalization terms are not really comparable. When α is large enough, the first one ensures the invertibility of the resulting linear system even if $\beta = 0$. For $\beta \neq 0$, α can be set independently from the mesh-size. When $\beta = 0$, α must be chosen very large, actually proportional to $1/h^2$, h being the characteristic mesh-size. This point is confirmed by the numerical experiments given below. The following term

$$\int_{\Gamma} \gamma \llbracket \frac{1}{\varrho} \nabla u \rrbracket \llbracket \frac{1}{\varrho} \nabla v \rrbracket ds \quad (27)$$

could have been also considered to cover most of the usual DG methods. However, several numerical experiments, conducted all along this study, have always shown that the choice of parameter γ is critical. Its modulus must be small enough (of order 10^{-2} of the wavelength) to avoid high instabilities which otherwise seriously damage the results. However, even for small γ , no real improvement was observed as compared to the case $\gamma = 0$. This is very important since considering (27) severely complicates the assembly process as this will be shown below.

At this point, we have set up the general variational framework of the formulation at the continuous level. Two additional steps are still pending.

- Define a Galerkin approximation of $X_{\mathcal{T}}$. It is precisely a first novelty of this study to propose a BEM for carrying out this construction.
- Compute the bilinear and linear forms involved in the formulation. The main difficulty relies upon the need to compute as accurately as possible an approximation of the DtN map related to the Helmholtz equation within each $T \in \mathcal{T}$. The second novelty of the method proposed in this paper is to perform this approximation through the introduction of a refinement of the mesh used for the BEM.

Before addressing these issues, we conclude this section by situating this formulation in the context of some DG methods formerly proposed to solve the above wave propagation problem.

2.3. Some previous Trefftz-DG formulations

A thorough review of Trefftz methods for solving the Helmholtz equation has been recently performed in [7]. We limit ourselves here to a comparison with methods of DG type. A clear definition of such a kind of methods is given in [7]: “*DG*” [...] [are] *methods that arrive at local variational formulation by applying integration by parts to the PDE to be approximated*. In this respect, our formulation cannot be compared to the VTCR method which is a weighted residual approach (cf. [22, 7]), moreover written in terms of a nonsymmetric variational system.

2.3.1. Interior Penalty DG Methods

Interior Penalty DG (IPDG) methods are mostly introduced as above by integration by parts at the element level and adding consistent penalty terms (see for instance [28, 2, 29] and the references therein).

Actually adapting the IPDG introduced in [28] for the Helmholtz equation in (4) and considering that $\partial\Omega_D = \partial\Omega$ as there, we obtain

$$\begin{aligned} & \sum_{T \in \mathcal{T}} \int_T \frac{1}{\epsilon_T} (\nabla u \cdot \nabla v - \kappa_T^2 uv) \, dx \\ & - \int_{\Gamma} \left(\left\{ \frac{1}{\epsilon} \nabla u \right\} \cdot [v] + [u] \left\{ \frac{1}{\epsilon} \nabla v \right\} \right) ds - \int_{\partial\Omega_D} \left(u \frac{1}{\epsilon} \nabla v \cdot \mathbf{n} + \frac{1}{\epsilon} \nabla u \cdot \mathbf{n} v \right) ds \\ & + \int_{\Gamma} \alpha [u] \cdot [v] \, ds + \int_{\partial\Omega_D} \alpha uv \, ds = \int_{\partial\Omega_D} \left(-g_D \frac{1}{\epsilon} \nabla v \cdot \mathbf{n} + \alpha g_D v \right) ds. \end{aligned}$$

Using the fact that v is also a solution to the Helmholtz equation in T and integrating by parts once again, we get

$$\begin{aligned} & \int_{\Gamma} \left(u \llbracket \frac{1}{\rho} \nabla v \rrbracket - \llbracket \frac{1}{\rho} \nabla u \rrbracket \cdot \llbracket v \rrbracket - \llbracket u \rrbracket \llbracket \frac{1}{\rho} \nabla v \rrbracket \right) ds \\ & + \int_{\partial\Omega_D} u \frac{1}{\rho} \nabla v \cdot \mathbf{n} ds - \int_{\partial\Omega_D} \left(u \frac{1}{\rho} \nabla v \cdot \mathbf{n} + \frac{1}{\rho} \nabla u \cdot \mathbf{n} v \right) ds + \int_{\partial\Omega_D} \alpha u v ds \\ & + \int_{\Gamma} \alpha \llbracket u \rrbracket \cdot \llbracket v \rrbracket ds = \int_{\partial\Omega_D} \left(-g_D \frac{1}{\rho} \nabla v \cdot \mathbf{n} + \alpha g_D v \right) ds. \end{aligned}$$

Using the equivalent expressions

$$\int_{\Gamma} u \llbracket \frac{1}{\rho} \nabla v \rrbracket ds = \int_{\Gamma} \llbracket u \rrbracket \llbracket \frac{1}{\rho} \nabla v \rrbracket ds \text{ and } \int_{\Gamma} \llbracket \frac{1}{\rho} \nabla u \rrbracket \llbracket v \rrbracket ds = 0$$

and substituting g_D for u in the first integral on $\partial\Omega_D$, we directly arrive to formulation (23) with $\beta = \delta = 0$.

Proceeding in the same way for the IPDG method considered in [2], we find again formulation (23) with $Y = -\kappa$, $g_D = 0$, $\delta = 0$, $\partial\Omega_N = \emptyset$.

It is clear from the above examples that, up to some consistent terms, any IPDG method can be put in the form of variational formulation (23) with suitable values for the penalty parameters α , β , δ .

2.3.2. DG methods based on numerical fluxes

For the Helmholtz equation, DG methods based on numerical fluxes can be divided in two broad classes: those which are a simple reformulation of the above IPDG methods and those which can be linked to an upwinding numerical scheme. Actually, in the context of the solution of the Helmholtz equation, the upwinding techniques are intimately related to the UWVF as this was brought out in [14]. However, upwinding is seemingly stated in the literature in a clear manner only for the Helmholtz equation with constant coefficients. We found it useful to recall some features about these techniques to more clearly set out the difference between a real upwind scheme and a simple enforcement of the continuity conditions when the PDE coefficients are discontinuous.

The starting point is the use of either of the following techniques performed in every element T of the DG formulation mesh:

- **the primal method**, as it is called in [9], which consists in integrating by parts the Helmholtz equation with the additional feature that v is a solution to the local Helmholtz equation,
- **the mixed method** [10], where the integration by parts is carried out on a first-order system, which is an equivalent formulation of the Helmholtz equation with as test function a pairing $(v, \boldsymbol{\tau})$ solution to the complex conjugate system (this can also be done without reference to the scalar equation, directly on system (3), in [14]).

Both of these approaches give rise to the following variational equation

$$\int_{\partial T} (\hat{\boldsymbol{\sigma}} \cdot \mathbf{n}_T v_T + \hat{u} \mathbf{n}_T \cdot \boldsymbol{\tau}) ds = 0 \quad (28)$$

where $\hat{\sigma} = \sigma_T$ and $\hat{u} = u_T$, without further steps being taken, (see [33] also).

In a series of papers (cf. [7] and the references therein), Hiptmair, Moiola, Perugia, and their co-authors obtained variational formulation (23) without the consistent terms added to the above IPDG methods to get a symmetric variational formulation. It is worth mentioning that the variational formulation used in these studies is not symmetric. It can lead however to a symmetric linear system if the involved edge/face integrals are calculated exactly. However, it is often not possible to benefit from such a favorable possibility.

An obvious advantage of a symmetric formulation is to reduce by a factor 2 the memory storage and the computational time. Primal FEM formulations of a symmetric problem are symmetric by nature. This is not the case for other formulations like DG methods or BEMs. Even if it is well-established for problems leading to a linear system with a hermitian positive definite matrix only, several studies, specially on BIE formulations of transmission problems or on BEM-FEM coupling, insist on the fact that the variational formulations of symmetric problems must preferably be symmetric (cf., for example, [42, 43]). Relatively more recently, a theoretical study in 1D [44] showed that symmetric DG methods own strong stability properties so that they do not need to be stabilized by penalty terms.

2.3.3. The upwinding scheme

The UWVF formulation for the Helmholtz equation can actually be seen as a upwinding scheme derived at the level of the second-order scalar PDE. As mentioned above in the introduction, it is shown in [15, 9] that, for the Helmholtz equation with constant coefficients, the UWVF can be recast in the framework of the above TDG method for particular values of α , β , δ and also parameter γ when considering augmented formulations involving the jump of the normal trace. Formulation (23) can hence be viewed as a symmetric variational extension of the UWVF method if some specific mathematical properties of the UWVF are not pursued, namely, the fact that the UWVF leads to a formulation set in terms of the perturbation of the identity by a norm diminishing operator [13]. However, one must be aware that then our formulation can no longer be considered as an upwinding scheme. In the same way, the extension given in [35] for boundary-value problem (4) for piecewise constant coefficients, cannot be considered as an upwinding scheme. Actually, this extension can be interpreted as a centered method for designing a local homogeneous propagation environment first and using a upwinding scheme then. A similar handling of discontinuous coefficients is standard in the numerical solution of time domain hyperbolic systems. A nice presentation of this technique is given in [45]. It is indeed shown in [46] that the propagation medium can be set arbitrarily before performing the upwinding scheme while keeping the general properties of the UWVF. In the case of homogeneous propagation medium, [14] clearly shows the connection of the UWVF with an upwinding scheme. It turns out that matching conditions (21) can be rewritten as a balance sheet of the incoming and outgoing waves crossing an edge/face F .

3. The BEM symmetric Trefftz DG method

For $T \in \mathcal{T}$, the DtN operator associates the “dual variable”

$$q_T = \left(\frac{1}{\varrho_T} \nabla v \right) |_{\partial T} \cdot \mathbf{n}_T \quad (29)$$

with the trace v_T on ∂T of $v|_T$, hereafter called the “primal variable”, for $v \in X_{\mathcal{T}}$. We first show how $X_{\mathcal{T}}$ and the bilinear and linear forms involved in the DG formulation can be handled by expressing the DtN operator by means of a BIE approach. The BEM-STDG method can then be fully derived from a BEM approximation of q_T for each $T \in \mathcal{T}$.

3.1. The boundary integral equation within each element of the DG formulation mesh

From the well-known integral representations of the solutions to the Helmholtz equation with constant coefficients, the restriction $v|_T$ to each $T \in \mathcal{T}$ of every $v \in X_{\mathcal{T}}$ can be expressed explicitly in terms of the above primal variable v_T and dual variable q_T

$$v|_T = \varrho_T V_T q_T + N_T v_T$$

where the single-layer and the double-layer boundary integral operators are defined by

$$V_T q_T(x) = \int_{\partial T} G_T(x, y) q_T(y) ds_y \quad (x \in \partial T) \quad (30)$$

and

$$N_T v_T(x) = - \int_{\partial T} \partial_{\mathbf{n}_T(y)} G_T(x, y) v_T(y) ds_y \quad (x \in \partial T). \quad (31)$$

respectively. The kernel $G_T(x, y)$ involved in the above formulas is that corresponding to the outgoing solutions to the Helmholtz equation with wavenumber κ_T . For the moment, we assume that the interior Dirichlet problem is well-posed within every $T \in \mathcal{T}$. A geometrical criterion ensuring this property is given below. As a result, the single-layer boundary integral operator is invertible (see, for example, [40, p. 226]). It then follows that q_T can be obtained from v_T by solving the following BIE

$$V_T q_T = \frac{1}{\varrho_T} \left(\frac{1}{2} - N_T \right) v_T. \quad (32)$$

For all the properties related to the solution of Helmholtz equation by BIEs, we refer for instance to [37, 47, 31].

As a consequence, every $v \in X_{\mathcal{T}}$ can be parametrized by the set of traces $\{v_T\}_{T \in \mathcal{T}}$. Space $X_{\mathcal{T}}$ can thus be identified with the product space

$$\prod_{T \in \mathcal{T}} H^1(\partial T). \quad (33)$$

From now on, we use this identification to refer to $X_{\mathcal{T}}$.

We now turn our attention to the above-mentioned geometric criterion ensuring the well-posedness of the Dirichlet problem. It is stated as follows.

Geometric criterion. Assume that there exists a unit vector \mathbf{v} such that

$$(x - y) \cdot \mathbf{v} \leq \lambda_T/2 \text{ for all } x \text{ and } y \text{ in } T, \quad (34)$$

where $\lambda_T = 2\pi/\kappa_T$ is the wavelength within T . Then, the boundary-value problem for the Helmholtz equation with Dirichlet boundary condition and wavenumber κ_T is well posed in T .

Set $\ell = \sup (x - y) \cdot \mathbf{v}$. With no loss of generality, we can assume that $T \subset]0, \ell[\times \prod_{i=2,\dots,d}]0, \ell_i[$. From the minmax principle, it can be argued that the first eigenvalue χ^2 of the Laplace operator with a Dirichlet boundary condition satisfies $\chi^2 \geq \pi^2/(\ell^2 + \ell_2^2 + \dots + \ell_d^2)$, thus establishing the criterion since $\ell \leq \lambda_T/2$, and therefore $\kappa_T \leq \pi/\ell < \chi$.

It is well worth insisting on the fact that the geometric criterion is not used to make sure that p_T is uniquely determined by BIE (32) but is rather intended to ensure that the solutions to the Helmholtz equation in the interior of any T can be uniquely defined in terms of their Dirichlet trace on ∂T . The invertibility of the single-layer potential V_T is a “by-product” of this condition on DG formulation mesh. Such a requirement should not be understood as a restriction but rather as a way to prevent the size of the local problems to increase at a scale that the assembly process becomes unmanageable.

3.2. The BEM symmetric Trefftz DG method

3.2.1. Two additional meshes

The set-up of the BEM within each $T \in \mathcal{T}$ requires a mesh of its boundary ∂T . All of these local meshes are actually obtained by meshing the elements F of the skeleton \mathcal{F} of the DG formulation mesh \mathcal{T} introduced in (9). This is the second mesh involved in the formulation. It is referred to below as the BEM mesh and denoted by \mathcal{F}^h , h being as usual the related mesh-size. To be consistent with the functional framework (33), which, as discussed above, requires v_T to be in $H^1(\partial T)$, the BEM mesh \mathcal{F}^h must induce a boundary mesh $\mathcal{F}_{\partial T}^h$ on ∂T compatible with the construction of a continuous FEM on ∂T . This is not at all restrictive in the two-dimensional case. In three dimensions, it is necessary to mesh each face F according to the usual matching conditions of a continuous FEM within the boundary ∂T of T (cf., for example, [48, 49]).

As already mentioned, the method requires a third mesh for accurately approximating the DtN map: the refined BEM mesh $\mathcal{F}_{N_{\text{add}}}^h$, which is obtained by subdividing each element ω of \mathcal{F}^h in N_{add} sub-elements. In the two-dimensional case, the refined BEM mesh $\mathcal{F}_{N_{\text{add}}}^h$ is obtained just by subdividing each segment of the BEM mesh \mathcal{F}^h in N_{add} segments. In the same way than for $\mathcal{F}_{\partial T}^h$, the refined BEM mesh $\mathcal{F}_{N_{\text{add}}}^h$ induces a refined boundary mesh $\mathcal{F}_{N_{\text{add}}, \partial T}^h$ on ∂T .

A schematic view of these three meshes is displayed in FIG. 3.

3.2.2. Approximation of the primal variable

Each component v_T , $T \in \mathcal{T}$, of the primal variable v is sought as a polynomial function of degree m , that is in \mathbb{P}_m , within each element of the boundary mesh $\mathcal{F}_{\partial T}^h$, continuous on ∂T . The primal variable v is multi-valued on $\Gamma \cup \partial\Omega$. This is in contrast with the BEM-based FEM considered in [25, 24] where the approximation of the primal

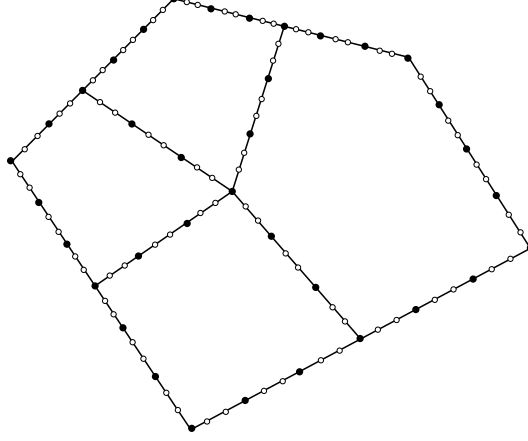


Figure 3: Schematic view of the three kinds of meshes involved in the BEM-STDG method. The 4 polygonals constitute the DG formulation mesh \mathcal{T} . The vertices of the BEM mesh \mathcal{F}^h and the refined BEM mesh \mathcal{F}_3^h are marked by large dots and small circles respectively. Refinement parameter N_{add} is taken here equal to 3. The boundary ∂T of every polygonal $T \in \mathcal{T}$ is endowed with a boundary mesh $\mathcal{F}_{\partial T}^h$ induced by \mathcal{F}^h and a refined boundary mesh $\mathcal{F}_{3,\partial T}^h$ induced by \mathcal{F}_3^h .

variable is single-valued on $\Gamma \cup \partial\Omega$. To be more specific, v_T is assumed to be continuous everywhere on ∂T even at cross-points but has no connection with any other v_K even if ∂T and ∂K are sharing common points or a common part. An idea on the continuity conditions that are imposed on the considered element-wise BEM is given in FIG. 4.

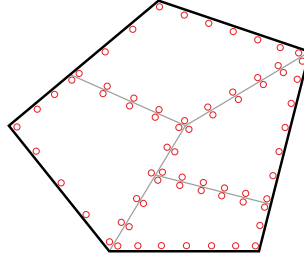


Figure 4: BEM mesh and nodes used in the 2D case.

Actually, this construction results in a finite dimensional Galerkin approximation $X_{\mathcal{T}}^{m,h}$ of $X_{\mathcal{T}}$, which can be identified to the product space $\prod_{T \in \mathcal{T}} X_{\partial T}^{m,h}$, where $X_{\partial T}^{m,h}$ is a \mathcal{C}^0 -BEM approximation of degree m of $H^1(\partial T)$ on the boundary mesh $\mathcal{F}_{\partial T}^h$

$$X_{\partial T}^{m,h} = \{v_T \in \mathcal{C}^0(\partial T); v_T|_{\omega} \in \mathbb{P}_m, \forall \omega \in \mathcal{F}_{\partial T}^h\}.$$

3.2.3. Approximation of the dual variables

The involvement of the BEM at the level of the TDG method is completely embodied in the approximation of the DtN operator expressing the dual variable q_T in terms of v_T

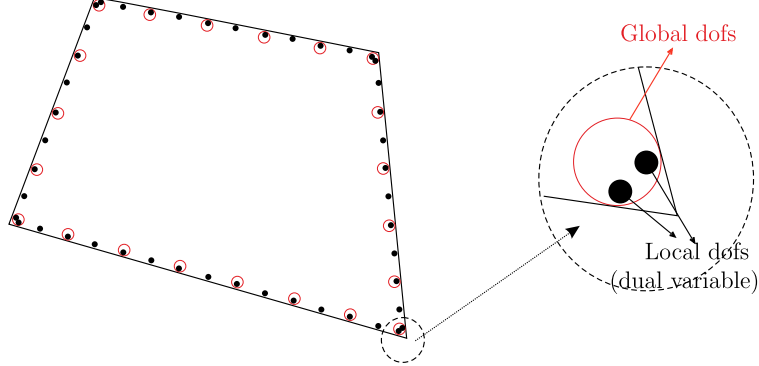


Figure 5: Nodes for v_T and q_T within an element T of the DG formulation mesh. Nodes for v_T are represented by large circles whereas nodes for q_T are depicted by small points.

by solving Eq. (32). The accuracy of this approximation is crucial for the reduction of the “pollution effect”. To enhance the sharpness of this procedure, we have adopted the following strategy:

- q_T is approximated on the boundary mesh $\mathcal{F}_{N_{\text{add}}, \partial T}^h$;
- Contrary to v_T , q_T is continuous within each edge/face F of \mathcal{F} only, but not at the junctions of the edges/faces; in other words, q_T can be multi-valued at the “cross-points” lying on ∂T . The schematic drawing depicted in FIG. 5 provides an illustration of the continuity conditions imposed to the primal variable v_T and to the dual variable q_T . It corresponds to the following values for the discretization parameters

- polynomial degree of the local BEM on each ∂T : $m = 1$,
- each element T of the BEM mesh is divided in 2 sub-elements: $N_{\text{add}} = 2$.

Actually, the above approximation of the dual variable q_T corresponds to a Galerkin approximation $M_{\mathcal{T}}^{m,h} = \prod_{T \in \mathcal{T}} M_{\partial T}^{m,h}$ of the product space $M_{\mathcal{T}} = \prod_{T \in \mathcal{T}} L^2(\partial T)$ built as follows. Denoting by \mathcal{F}_T the set of edges/faces F , which form ∂T , we have

$$M_{\partial T}^{m,h} = \{q_T \in L^2(\partial T); q_T \in \mathcal{C}^0(F), \forall F \in \mathcal{F}_T, q_T|_F \in \mathbb{P}_m, \forall \omega \in \mathcal{F}_{N_{\text{add}}, \partial T}^h\}.$$

The following property $X_{\mathcal{T}}^{m,h} \subset M_{\mathcal{T}}^{m,h}$, expressing of course that

$$X_{\partial T}^{m,h} \subset M_{\partial T}^{m,h}, \forall T \in \mathcal{T} \quad (35)$$

deserves a particular attention. It results from the construction of the refined BEM mesh $\mathcal{F}_{N_{\text{add}}}^h$ and plays an important role below in the construction of the elemental matrices involved in the BEM.

Remark. It is important to note that only the nodal values of the primal variable are global. Those of the dual variables are used only at the level of element $T \in \mathcal{T}$ in the assembly process.

3.2.4. The elemental matrices of the local BEM

Let us denote by $[u_T]$, $[v_T]$ the column-wise vector whose components are the nodal values of u_T , v_T as elements of $X_{\partial T}^{m,h}$, and $[p_T^\#]$, $[q_T^\#]$ those of p_T , q_T as elements of $M_{\partial T}^{m,h}$. Symbol $\#$ distinguishes the two sets of nodes and the related matrices. In view of (35), u_T and v_T can be also singled out by the column-vectors $[u_T^\#]$ and $[v_T^\#]$ of their nodal values as elements of $M_{\partial T}^{m,h}$. The respective nodal values $[u_T^\#]$ and $[v_T^\#]$ of u_T and v_T on the augmented set of nodes are obtained by interpolating u_T and v_T respectively on the refined mesh and by duplicating nodes where p_T or q_T are not continuous. In this way, $[u_T^\#]$, $[v_T^\#]$, $[p_T^\#]$, and $[q_T^\#]$ are all of the same length and have components all referring to the same nodes. The components of $[u_T^\#]$ are expressed in terms of those of $[u_T]$ by means of an explicit matrix $[P_T]$

$$[u_T^\#] = [P_T] [u_T]. \quad (36)$$

Thus, assuming that all the involved functions are in $M_{\partial T}^{m,h}$, we can define the matrices $[M_T^\#]$, $[V_T^\#]$, and $[N_T^\#]$ through the following identifications

$$\begin{cases} [q_T^\#]^\top [M_T^\#] [p_T^\#] = \int_{\partial T} p_T q_T ds, \\ [q_T^\#]^\top [V_T^\#] [p_T^\#] = \int_{\partial T} (V_T p_T) q_T ds, \\ [q_T^\#]^\top [N_T^\#] [p_T^\#] = \int_{\partial T} (N_T p_T) q_T ds, \end{cases}$$

p_T and q_T playing the role of generic functions in $M_{\partial T}^{m,h}$ without referring specifically to dual variables exceptionally in the above and the following relations.

Any of the above matrices is obtained by an assembly process at the level of the above mesh $\mathcal{F}_{\text{Nadd},\partial T}^h$ from the following elemental matrices

$$\begin{cases} [q_\omega^\#]^\top [M_\omega^\#] [p_\omega^\#] = \int_\omega p_T q_T ds, \\ [q_v^\#]^\top [V_{v\omega}^\#] [p_\omega^\#] = \int_v \int_\omega G_T(x, y) p_T(y) q_T(x) ds_y ds_x, \\ [q_v^\#]^\top [N_{v\omega}^\#] [p_\omega^\#] = \int_v \int_\omega -\partial_{\mathbf{n}_y} G_T(x, y) p_T(y) q_T(x) ds_y ds_x \end{cases} \quad (37)$$

where v and ω are two generic elements of the refined boundary mesh $\mathcal{F}_{\text{Nadd},\partial T}^h$.

Equation (32) then yields that nodal values $[p_T^\#]$ and $[q_T^\#]$ of the dual variables p_T and q_T are expressed at the level of element $T \in \mathcal{T}$ in terms of the extended nodal values $[u_T^\#]$ and $[v_T^\#]$ of the primal variables u_T and v_T by

$$\begin{cases} [p_T^\#] = [D_T^\#] [u_T^\#], & [q_T^\#] = [D_T^\#] [v_T^\#], \\ [D_T^\#] = \frac{1}{\varrho_T} [V_T^\#]^{-1} \left(\frac{1}{2} [M_T^\#] - [N_T^\#] \right). \end{cases} \quad (38)$$

It is at this level that the well-posedness of the interior Dirichlet problem for the laplacian enters into the picture. It ensures the invertibility of matrix $\begin{bmatrix} V_T^\# \end{bmatrix}$.

Using (36), we thus get the numerical approximation of the DtN operator

$$\begin{bmatrix} p_T^\# \end{bmatrix} = \begin{bmatrix} D_T^\# \end{bmatrix} [P_T] [u_T] \quad (39)$$

in terms of the nodal values $[u_T]$ of u_T and $\begin{bmatrix} p_T^\# \end{bmatrix}$ of p_T .

3.2.5. The BEM-STDG method

Collecting the vectors $[u_T]$ and $[v_T]$ for $T \in \mathcal{T}$ in column-wise vectors $[u]$ and $[v]$ respectively, and expressing $\begin{bmatrix} p_T^\# \end{bmatrix}$ and $\begin{bmatrix} q_T^\# \end{bmatrix}$ from (39), we form by means of an assembly process, detailed below, the square matrix $[A]$ and column-wise vector $[b]$ through the following identifications

$$[v]^\top [A] [u] = a(u, v), \quad [v]^\top [b] = Lv, \quad \text{for } u, v \in X_{\mathcal{T}}^{m,h}.$$

We are hence led to solve the symmetric linear system

$$[A] [u] = [b].$$

Clearly, $[A]$ is also a sparse matrix in the meaning that any couple of degrees of freedom, belonging to two elements of the DG formulation mesh \mathcal{T} , and not sharing a common edge/face, are not connected.

3.2.6. The assembly process

As usual the assembly process is based on a decomposition of the above bilinear and linear forms as a sum of local forms related to each element $T \in \mathcal{T}$

$$a(u, v) = \sum_{T \in \mathcal{T}} \sum_{F \in \mathcal{F}_T} a_{F,T}(u, v), \quad Lv = \sum_{T \in \mathcal{T}} \sum_{F \in \mathcal{F}_T} L_{F,T} v. \quad (40)$$

However, some additional notation and observations are required before the explicit expressions of these local forms can be obtained.

When F is an interior edge/face shared by T and K , defining in Eq. (29) by p_K and q_K the dual variables related to K , the integrals on F involved in $a(u, v)$ can be written in a simpler form

$$\begin{aligned} \int_F (\{u\} [a \nabla v] + [a \nabla u] \{v\} - \{a \nabla u\} \cdot [v] - [u] \cdot \{a \nabla v\}) ds \\ = \int_F (u_T q_K + u_K q_T + p_T v_K + p_K v_T) ds, \end{aligned} \quad (41)$$

$$\begin{aligned} \int_F (\alpha [u] [v] + \beta [\nabla_\top u] \cdot [\nabla_\top v]) ds = \\ \int_F \alpha (u_T - u_K) v_T + \beta \nabla_\top (u_T - u_K) \cdot \nabla_\top v_T ds + \\ \int_F \alpha (u_K - u_T) v_K ds + \beta \nabla_\top (u_K - u_T) \cdot \nabla_\top v_K ds, \end{aligned} \quad (42)$$

In this way, the related contribution $a_{F,T}(u, v)$ to the global bilinear form $a(u, v)$ reads

$$\begin{aligned} a_{F,T}(u, v) = & \int_F (p_T v_K + u_K q_T) ds \\ & + \int_F (\alpha u_T (v_T - v_K) + \beta \nabla_{\top} u_T \cdot \nabla_{\top} (v_T - v_K)) ds. \end{aligned} \quad (43)$$

The expressions of $a_{F,T}(u, v)$ and $L_{F,T}v$ for $F \in \mathcal{F}_{\partial}$ are obtained in a straightforward way by using the appropriate integral according to the involved part of $\partial\Omega$ and substituting p_T and q_T for respectively $\left(\frac{1}{\varrho_T} \nabla u\right)|_T \cdot \mathbf{n}_T$ and $\left(\frac{1}{\varrho_T} \nabla v\right)|_T \cdot \mathbf{n}_T$.

Remark. It is very important to point out that the variational formulation involves no penalty terms for matching the dual variables. Only p_T and q_T occur in the expressions of the local forms but neither p_K nor q_K related to an adjacent element K . Boundary element matrix $\left[D_T^{\#}\right]$ can therefore be computed only at the level of the assembly of element T and has not to be stored. Otherwise, to consider the term

$$\int_F \gamma [a \nabla u] [a \nabla v] ds = \int_F \gamma (p_T + p_K) q_T ds + \int_K \gamma (p_K + p_T) q_K ds \quad (44)$$

in the formulation, we must either store matrix $\left[D_K^{\#}\right]$ expressing p_K from u_K or compute it twice. As mentioned above, the consideration of this term would generate difficulties at the level of both the stability of the method and its effective implementation with no gain.

3.2.7. Calculation of the elemental integrals

To maintain the accuracy of the global procedure, and specifically the error corresponding to the “pollution effect” at a minimum level, a particular attention must be paid to the accuracy of the integrals involved in the BEM. Here, we limit ourselves to the 2D case which is the framework of the numerical tests given in the subsequent text.

In view of (37), we are led to calculate integrals in the following form

$$\int_v \int_{\omega} \mu(x) \lambda(y) G(x, y) ds_y ds_x \quad \text{and} \quad \int_v \int_{\omega} \mu(x) \lambda(y) \partial_{\mathbf{n}_{\omega}} G(x, y) ds_y ds_x$$

with λ and μ in \mathbb{P}_m . The following notation and properties are used to describe how the above integrals have been dealt with:

- ω and v are respectively the segments $[a, b]$ and $[c, d]$;
- $y_s = a + s\boldsymbol{\tau}_{\omega}$, $x_t = c + t\boldsymbol{\tau}_v$; $h = |b - a|$, $\ell = |d - c|$; $\boldsymbol{\tau}_{\omega} = (b - a) / \ell$, $\boldsymbol{\tau}_v = (d - c) / h$; $y_{h/2} = y_{\omega}$, the mid-point of ω ; $x_{\ell/2} = x_v$, that of v ;
- $\lambda(y_s)$ and $\mu(x_t)$ are in \mathbb{P}_m as functions of $s \in [0, h]$ and $t \in [0, \ell]$ respectively;
- $G(x, y) = (i/4)H_0^{(1)}(\kappa|x - y|)$; $H_0^{(1)}$ is the Hankel function of the first kind of order 0 and κ the local wavenumber;

- \mathbf{n}_ω is the unit normal to segment ω obtained by rotating $\boldsymbol{\tau}_\omega$ by $\pi/2$ in the clockwise direction, assumed to be compatible with the unit normal to the boundary ∂T outwardly directed to the element $T \in \mathcal{T}$.

Clearly the calculation reduces to the evaluation of the following two types of integrals

$$\int_{t=0}^h \int_{s=0}^\ell t^k s^l G(x_t, y_s) ds dt \quad \text{and} \quad \int_{t=0}^h \int_{s=0}^\ell t^k s^l \partial_{\mathbf{n}_\omega} G(x_t, y_s) ds dt \quad (45)$$

with $0 \leq k, l \leq m$.

When $|x_v - y_\omega| > 1.5 \max(h, \ell)$, the above integrals are calculated by a Gauss formula integrating polynomials of degree $3m$ exactly. Otherwise, the singular parts of the integrals are extracted from the following expansions of the Hankel functions

$$\begin{aligned} \frac{i}{4} H_0^{(1)}(\kappa |x - y|) &= -\frac{1}{2\pi} \ln |x - y| + R_0^{(1)}(\kappa, |x - y|), \\ -\frac{i}{4\kappa} H_1^{(1)}(\kappa |x - y|) &= -\frac{1}{2\pi} \frac{1}{|x - y|} + R_1^{(1)}(\kappa, |x - y|). \end{aligned} \quad (46)$$

The functions $R_0^{(1)}(\kappa, |x - y|)$ and $R_1^{(1)}(\kappa, |x - y|)$ are obtained from an adaptation of the codes classically used to evaluate the Bessel-Neumann functions $Y_0(\kappa |x - y|)$ and $Y_1(\kappa |x - y|)$ which mainly avoids to lose significant digits which may result from floating point cancellation. The parts of the integrals related to $R_0^{(1)}(\kappa, |x - y|)$ and $R_1^{(1)}(\kappa, |x - y|)$ are grouped with the other integrals involving a bounded singularity coming from the procedure described right below and integrated using the CUBPACK library [50].

One is thus led to the evaluation of the two integrals

$$\int_{t=0}^h \int_{s=0}^\ell t^k s^l \ln |x_t - y_s| ds dt \quad \text{and} \quad \int_{t=0}^h \int_{s=0}^\ell t^k s^l \partial_{\mathbf{n}_\omega} \ln |x_t - y_s| ds dt. \quad (47)$$

The following decomposition

$$s^k = s^{k-1} ((y_s - x_t) \cdot \boldsymbol{\tau}_\omega - (c - x_t) \cdot \boldsymbol{\tau}_\omega) \quad (48)$$

and an induction procedure reduce the evaluation of the above singular integrals to the following ones

$$\begin{aligned} \int_{t=0}^h \int_{s=0}^\ell t^k s^l f(|x_t - y_s|) ds dt = \\ \int_{t=0}^h \int_{s=0}^\ell t^k \left(\sum_{p=0}^{l-1} s^{l-1-p} (-(c - x_t) \cdot \boldsymbol{\tau}_\omega)^p \right) (y_s - x_t) \cdot \boldsymbol{\tau}_\omega f(|x_t - y_s|) ds dt + \\ \int_{t=0}^h t^k (-(c - x_t) \cdot \boldsymbol{\tau}_\omega)^l \int_{s=0}^\ell f(|x_t - y_s|) ds dt \end{aligned} \quad (49)$$

with $f(|x_t - y_s|) = \ln |x_t - y_s|$ or $f(|x_t - y_s|) = \partial_{\mathbf{n}_\omega} \ln |x_t - y_s|$. The first term in the right-hand side is an integral on a rectangle with only bounded singularities: it is inte-

grated as above with the CUBPACK library. For the second term, elementary calculations, but too long to be reported here, yield

$$\int_{s=0}^{\ell} f(|x_t - y_s|) ds = \begin{cases} (d - x_t) \cdot \boldsymbol{\tau}_\omega \ln |d - x_t| - (c - x_t) \cdot \boldsymbol{\tau}_\omega \ln |c - x_t| - \ell + d_{x_t} \theta_{x_t}, \\ \quad \text{for } f(|x_t - y_s|) = \ln |x_t - y_s|, \\ \text{sgn}((c - x_t) \cdot \boldsymbol{n}_\omega) \theta_{x_t}, \\ \quad \text{for } f(|x_t - y_s|) = \partial_{\boldsymbol{n}_\omega} \ln |x_t - y_s|, \end{cases}$$

where d_{x_t} and $0 \leq \theta_{x_t} \leq \pi$ are respectively the distance of x_t to the line supporting segment v and the angle under which it is seen from x_t , and sgn is the function sign. Therefore, the evaluation of the second term in the right-hand side of (49) is reduced to that of a 1D integral with only bounded singularities, similarly computed with the CUBPACK library.

The integrals evaluated by the CUBPACK library are reduced to integrals on the square $]0, 1[\times]0, 1[$ for double integrals and on the segment $]0, 1[$ for simple integrals respectively. The parameters used for these evaluations are the following (see the documentation of the CUI routine (CUBPACK USER INTERFACE)): KEY = 7, JOB = 1, TUNED = 1, EPSABS = 1.D-8, EPSREL = 1.D-10, NEVAL = 10,000,000. The program is stopped if parameter IFAIL is returned with a value not equal to 0, that is, when the adaptive integration fails to converge.

4. Validation of the numerical method

We begin with the statement of a problem, which involves long-range wave propagation in a typical way. This problem will provide us with a good guideline for measuring the level of “*pollution effect*” occurring in any numerical solution of the problem. We will hence be able to compare the performances of the BEM-STDG method with those of the usual polynomial IPDG one. Prior to that, we first give in this section some numerical results validating the numerical approach.

4.1. The boundary-value problem

We consider the following example inspired from the wave propagation in a duct with rigid walls as presented in [51]

$$\begin{cases} \Delta u + \kappa^2 n^2 u = 0 \text{ in } \Omega, \\ u(0, y) = 1, \quad \partial_x u(2L, y) - i\kappa u(2L, y) = 0, \quad 0 < y < H, \\ \partial_y u(x, 0) = \partial_y u(x, H) = 0, \quad 0 < x < 2L, \end{cases} \quad (50)$$

set in

$$\Omega = \{(x, y) \in \mathbb{R}^2; 0 < x < 2L, 0 < y < H\}, \quad (51)$$

(see FIG. 6) where κ is constant and n is the piecewise constant function given by

$$n = \begin{cases} 1 & \text{for } |x - L| > D, \\ n_0 & \text{for } |x - L| < D. \end{cases} \quad (52)$$

Comparatively with the problem considered in [51], we added a Dirichlet boundary condition on the inlet boundary. In this way, we deal with the three kinds of boundary conditions since we additionally have Neumann and Fourier-Robin boundary conditions on respectively the rigid walls and the outlet boundary. Moreover here, it is possible to consider a non homogeneous duct by choosing n_0 constant but $\neq 1$.

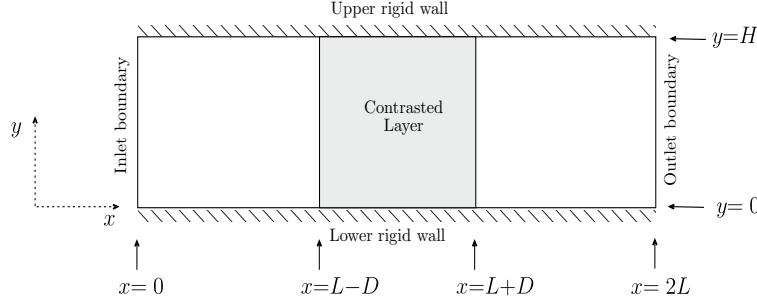


Figure 6: Geometry of the inhomogeneous duct with rigid walls.

Indeed, the solution to this problem is independent of y and can be expressed in terms of four parameters: R , T , R_D , and T_D as follows

$$u(x, y) = \begin{cases} T_D e^{i\kappa n(L-D)x} + R_D e^{-i\kappa n(L-D)x}, & \text{for } |x - L| < D, \\ (1 - R) e^{i\kappa x} + R e^{-i\kappa x}, & \text{for } x < L - D, \\ T e^{i\kappa x}, & \text{for } x > L + D. \end{cases} \quad (53)$$

Parameters R , T , and R_D can be expressed in terms of T_D through

$$\begin{cases} e^{-i\kappa n_0 D} R_D = \frac{n_0 - 1}{n_0 + 1} T_D e^{i\kappa n_0 D}, & e^{i\kappa L} T = \frac{2n_0}{n_0 + 1} e^{i\kappa(n_0 - 1)D} T_D, \\ e^{-i\kappa L} R = -\frac{n_0 - 1}{2} e^{-i\kappa(n_0 + 1)D} (1 - e^{4i\kappa n_0 D}) T_D, \end{cases} \quad (54)$$

which itself is given by

$$T_D = \frac{2e^{i\kappa n_0 D}}{(n_0 + 1) e^{-i\kappa(L-D)} \left(1 - e^{4i\kappa n_0 D} \frac{(n_0 - 1)^2}{(n_0 + 1)^2}\right) - (n_0 - 1) e^{i\kappa(L-D)} (1 - e^{4i\kappa n_0 D})}. \quad (55)$$

We use the following errors for characterizing the accuracy of the numerical results:

- Maximum global error

$$\text{Err}_\infty = 100 \frac{\max_{(x_m, y_m)} |u(x_m, y_m) - u_m|}{\max |u(x, y)|} \quad (56)$$

where u_m is the nodal value at node (x_m, y_m) of the solution delivered by the BEM-STDG method;

- Error on the transmitted wave

$$\text{Err}_T = 100 |T - T_{\text{comp}}| \quad (57)$$

where T is the coefficient, given above, characterizing the solution for $x > L - D$, and T_{comp} is its approximate value obtained from the numerical simulation;

- Error on the reflected wave

$$\text{Err}_R = 100 |R - R_{\text{comp}}| \quad (58)$$

obtained similarly to Err_T .

In this part, the BEM mesh \mathcal{F}^h is obtained by dividing each edge F of the skeleton \mathcal{F} of \mathcal{T} in N_h segments.

Parameters $\alpha = \beta = 1.0 \cdot 10^2$, and $\delta = 0$ have been specified empirically. Parameter β was proposed in [2] to improve the stability of the DG method. At first sight, this term may be interpreted as a way to automatically increase the penalty parameter when considering a more refined BEM mesh. For instance, for the case of an evanescent wave, as considered below, it was necessary to take $\alpha = 1.0 \cdot 10^6$ for $\beta = 0$ to reach the accuracy obtained with the previous choice for α and β . However, at least in domain decomposition methods, it was observed for a long time that writing the matching conditions at the interfaces of the subdomains in this way greatly improves the convergence of the related iterative process [52]. Actually, the method has a low sensitivity relatively to these parameters as soon as α and β are taken sufficiently large, greater than $1.0 \cdot 10^2$ and less than $1.0 \cdot 10^7$.

4.2. Validation of the BEM-STDG method

We first validate the BEM-STDG method on two problems of small size. The first one concerns the duct problem considered above and the second one is related to the approximation of an evanescent wave.

4.2.1. A duct problem of small size

We consider the above duct problem with the following data:

- $\kappa = \pi$,
- length of the duct: $2L = 10$ half-wavelengths, width of the duct: $H = 2$ half-wavelengths,
- thickness of the contrasted layer: 4 half-wavelengths ($D = 2$) and its refractive index relatively to the rest of the duct: $n_0 = 2$.

The DG formulation mesh of the duct is depicted in FIG. 7. The two vertical straight lines define the boundary of the contrasted layer.

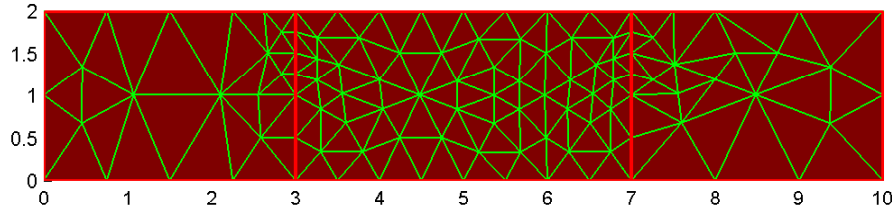


Figure 7: The DG formulation mesh used for solving the small size duct problem.

The parameters used for the BEM-STDG method are the following:

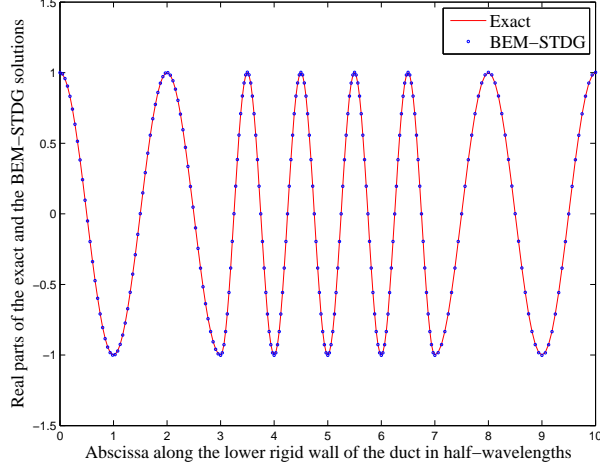


Figure 8: Real parts of the exact and the BEM-STDG solutions for the considered example of the duct problem.

- Mesh-size of the DG formulation mesh \mathcal{T} outside the contrasted layer: $h_{\max} = 1$,
- Mesh-size of \mathcal{T} inside the contrasted layer: $h_{\text{layer}} = 0.5$,
- Number of segments per edge of the DG formulation mesh to get the BEM mesh: $N_h = 16$,
- Number of added segments for the approximation of the DtN operator: $N_{\text{add}} = 4$,
- Polynomial degree used in the BEM: $m = 1$.

The plots in FIG. 8 depict the real parts of the exact and computed solutions on the nodes located on the lower rigid wall $\{y = 0\}$ of the duct. The two curves cannot be distinguished.

The following errors, which are all less than 1 %, validate the BEM-STDG method:

- Maximum error: $\text{Err}_{\infty} = 0.4 \%$;
- Transmitted wave: $\text{Err}_T = 0.06 \%$;
- Reflected wave: $\text{Err}_R = 0.3 \%$.

4.2.2. Approximation of an evanescent mode

Now, we test the ability of the BEM-STDG method to correctly approximate evanescent waves. For this case too, we adapt the conditions leading to an evanescent mode in [51]. We thus consider the same duct geometry than for the previous example with the same wave number $\kappa = \pi$ but we now assume that the duct is homogeneous, that is, $n_0 = 1$, and take

$$u(0, y) = \cos(2\pi y), \quad 0 < y < 2, \quad (59)$$

for the data involved in the Dirichlet boundary condition on the inlet boundary. To ensure that the exact solution is the second evanescent mode

$$u(x, y) = \cos(2\pi y) \exp\left(-\sqrt{3}\pi x\right), \quad (60)$$

it is enough to take the following transparent boundary condition on the outlet boundary

$$\left(\partial_x u + \sqrt{3}\pi u\right)(2L, y) = 0, \quad 0 < y < 2. \quad (61)$$

We used a DG formulation mesh with $h_{\max} = 0.5$, corresponding to 282 triangles, and, as in the above example, we took $N_{\text{add}} = 4$ for the refinement of the BEM mesh for the local computation of the DtN operator. Only the maximum error remains meaningful

$$\text{Err}_{\infty} = 0.4 \% \quad (62)$$

and is similar to the case of propagating mode. The plot depicted in FIG. 9 shows that the exponential decay of the mode is well reproduced by the solution obtained from the BEM-STDG numerical scheme.

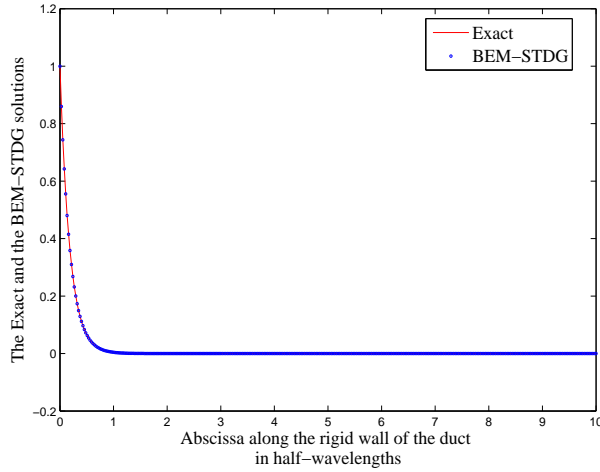


Figure 9: Exact and computed evanescent mode along the lower rigid wall of the duct.

One of the strengths of the method is its adaptation to p -refinements. For instance here, by subdividing each edge in $N_h = 12, 6, 4, 3$ for respectively a polynomial approximation of degree $m = 1, 2, 3, 4$ of the primal variable u , it is possible to keep unchanged the total number of degrees of freedom: $N = 10152$, and the filling of the matrix, while dramatically reducing the error as reported in TAB: 1.

Following a suggestion of one of the anonymous referees, we added a usual 1-norm condition number estimate, whose description is provided in the MATLABTM documentation for example, and the CPU time relative to each test. Table 1 shows in particular that the condition number does not blow up when increasing the degree of the polynomial approximation and that the improvement of the error is obtained without overburdening the computing time.

	Error (in %)	Condition Number ($\times 10^9$) (Estimate)	CPU time
$m = 1$	0.35	2.3	1060
$m = 2$	0.03	3.1	879
$m = 3$	0.01	4.7	967
$m = 4$	0.0008	8.7	1328

Table 1: Some parameters related to the solution of the duct problem on a DG formulation mesh in 282 triangles, involving 10152 unknowns, and yielding a sparse matrix with a filling of 0.6 %.

5. Long-range propagation

Now, we come to the main motivation for considering this BEM-STDG method: its ability to reduce the “*pollution effect*” and hence to perform correct numerical simulations of long-range propagation. Toward this end, we consider a homogeneous duct and its corresponding structured mesh.

5.1. The structured mesh

The structured DG formulation mesh, used for the benchmark tests related to the “*pollution effect*”, is depicted in FIG. 10. This mesh is characterized by two positive integers $N = 2L$ and $M = H$. In all these tests, κ is taken equal to π , so that the unit length is a half-wavelength. This automatically ensures that the local Dirichlet problem for the Helmholtz equation is well-posed in each element of the DG formulation mesh. All numerical experiments were performed with $M = 2$ and for N varying from 50 to 1000 by step of 50.

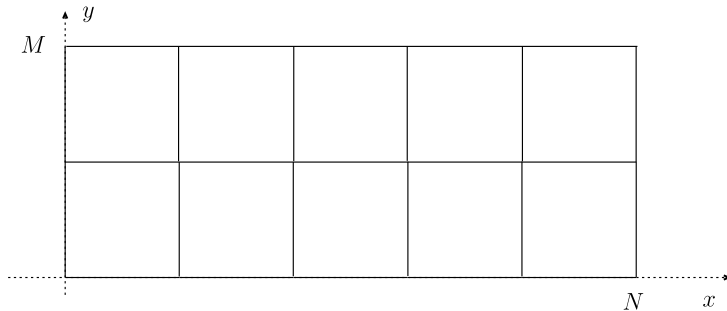


Figure 10: Type of structured DG formulation mesh used for the numerical experiments.

In this part, we characterize the BEM mesh by the number of nodes per wavelength, instead of parameter N_h used above for the validation of the method. The reasons behind the choice of this parameter will be detailed below. For instance, for the BEM, used in this experiment, whose shape functions are polynomials of degree 4, 24 nodes per wavelength correspond to a mesh-size $h = 1/3$, that is, with $N_h = 3$ segments per edge, and 16 nodes per wavelength with $h = 1/2$, that is, with $N_h = 2$ segments per edge.

As above for the case of an evanescent wave, a homogeneous duct is obtained by choosing $n_0 = 1$ in problem (50). Then, only T and R remain meaningful and have the following values

$$T = 1, R = 0. \quad (63)$$

Before focusing on the numerical experiments concerning the “*pollution effect*”, we first give a test confirming the importance of an accurate computation of the DtN map.

5.2. Approximation of the DtN operator on the refined BEM mesh

The plots in FIG. 11 depict the maximum error in % for a duct having a length of 500 wavelengths versus the number N_{add} of segments in which each segment of the BEM mesh is subdivided.

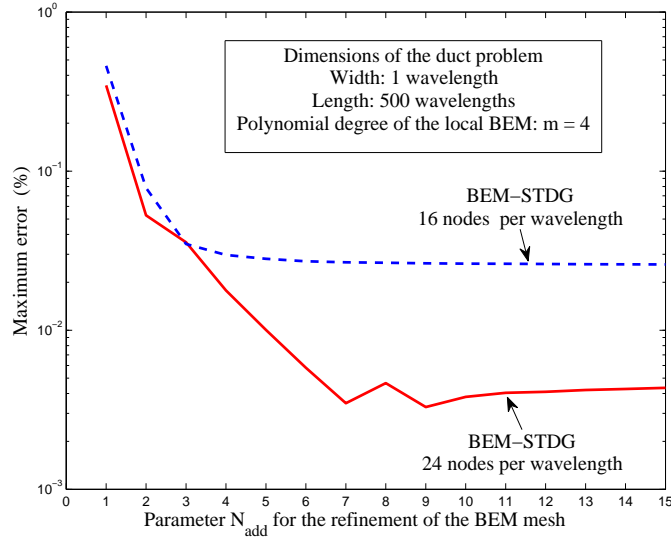


Figure 11: Maximum error in % versus N_{add}

The plots in FIG. 11 clearly demonstrate that a better approximation of the DtN operator greatly reduces the “*pollution effect*”: below $N_{\text{add}} = 3$, there is absolutely no advantage to use 24 instead of 16 nodes per wavelength.

5.3. Comparison of the BEM-STDG method with a polynomial IPDG method

To obtain an assessment of the BEM-STDG method, we compare its performances with those of a more conventional polynomial IPDG method (cf., for example, [2]). Two parameters are used for this comparison: the maximum global error in % defined earlier in (56) as a function of the length $2L$ of the duct and an estimate of the condition number. Other parameters are less meaningful for the considered duct problem since both the number of degrees of freedom, the memory usage, and the CPU time are simply linear functions of L since, as explained right below, the approximation procedure is characterized by the density of nodes being used.

It was not easy to find a common basis for comparing the two methods since the accuracy of the overall solution of the BEM-STDG method is mainly based on two meshes: the DG formulation mesh and the BEM mesh, and the polynomial IPDG method uses a structured finite element mesh in triangles as usual. Anyway, the following background seems to be a good basis for this task:

- use local polynomial approximations of the same degree for both the BEM-STDG and the polynomial IPDG method;
- assume that the degrees of freedom of the IPDG method are the nodes of the corresponding Lagrange FEM; then characterize each of these two methods by the density of nodes along each edge (number of nodes per wavelength). For instance, for a polynomial IPDG method constructed on a structured mesh in isoscele rectangular triangles whose right-angle side length is $1/N_h$, and for a BEM mesh built on the structured mesh given in FIG. 10 with N_h segments along each edge, the density, characterizing both the two methods for polynomial shape functions of degree m , is $2mN_h$ per wavelength. As a result, the IPDG method will be built on a mesh of $2N_h^2 MN$ isoscele rectangular triangles and will lead to a linear system of order $(m+1)(m+2)N_h^2 MN$, whereas the corresponding BEM-STDG method will involve $4mN_h NM$ unknowns only. It is however hard to compare the two methods on the basis on the total number of unknowns, they involve, since the matrix of the IPDG method is sparser than the BEM-STDG. This is why we chose to characterize both the h and the p refinements for each of these by $2mN_h$ which gives the genuine number of nodes by wavelength used in both the two methods.

The above maximum global error, as a function of the length L of the duct generally fits well with a straight line, at least for lengths large enough. The Least Square Grow Rate (LSGR) is the slope of this straight line, obtained by the least square method. It is used as an indicator of the impact of the “*pollution effect*”. Below, we successively compare the two methods from low degree polynomial approximations corresponding to $m = 1$ to higher degree ones corresponding to $m = 4$ for various densities of nodes per wavelength and for ducts with lengths up to 500 wavelengths.

5.4. Lowest polynomial degree

For the lowest polynomial degree $m = 1$, the BEM-STDG method widely outclasses the usual polynomial IPDG method. The error of the latter even with a double density of nodes per wavelength is 10 times higher. To be able to plot the error curves corresponding to the two methods in FIG. 12, we have had to use two axes at two different scales. Clearly, as indicated by the reported LSGR, the improvement gained by the BEM-STDG method is mainly due to a much better reduction of the “*pollution effect*”.

5.5. Higher polynomial degrees

For polynomial degrees from $m = 2$ up to $m = 4$, we have done three benchmark tests: the nearest densities to respectively one, one and a half, and twice the rule of thumb of 12 nodes per wavelength.

The results are reported in TAB. 2 and the most featuring are depicted in FIG. 13, FIG. 14, FIG. 15, and FIG. 16. The negative LSGR for $m = 3$ and a density of 24 nodes per wavelength is certainly due to rounding errors (see also FIG. 14 below).

Polynomial degree	Density (nodes / λ)	Method	Error	LSGR
$m = 2$	12	IPDG	72 %	$4.1 \cdot 10^{-1}$
		BEM-STDG	22 %	$4.3 \cdot 10^{-2}$
	16	IPDG	67 %	$1.3 \cdot 10^{-1}$
		BEM-STDG	5.6 %	$1.1 \cdot 10^{-2}$
	24	IPDG	13 %	$2.7 \cdot 10^{-2}$
		BEM-STDG	0.8 %	$1.5 \cdot 10^{-3}$
$m = 3$	12	IPDG	19 %	$3.7 \cdot 10^{-2}$
		BEM-STDG	1.6 %	$3.0 \cdot 10^{-3}$
	18	IPDG	1.7 %	$3.5 \cdot 10^{-3}$
		BEM-STDG	0.1 %	$1.0 \cdot 10^{-4}$
	24	IPDG	0.3 %	$6.2 \cdot 10^{-4}$
		BEM-STDG	0.02 %	$-2.6 \cdot 10^{-10}$
$m = 4$	8	IPDG	1.8 %	$3.9 \cdot 10^{-3}$
		BEM-STDG	10.4 %	$2.0 \cdot 10^{-2}$
	16	IPDG	0.17 %	$3.0 \cdot 10^{-4}$
		BEM-STDG	0.02 %	$4.3 \cdot 10^{-6}$
	24	IPDG	0.007 %	$1.3 \cdot 10^{-5}$
		BEM-STDG	0.004 %	$3.0 \cdot 10^{-12}$

Table 2: Maximum error in % for a duct of 500 wavelengths and Least Square Grow Rate of the error as a function of the duct length. The error related to the IPDG for $m = 2$ was reached for a duct of length 175λ .

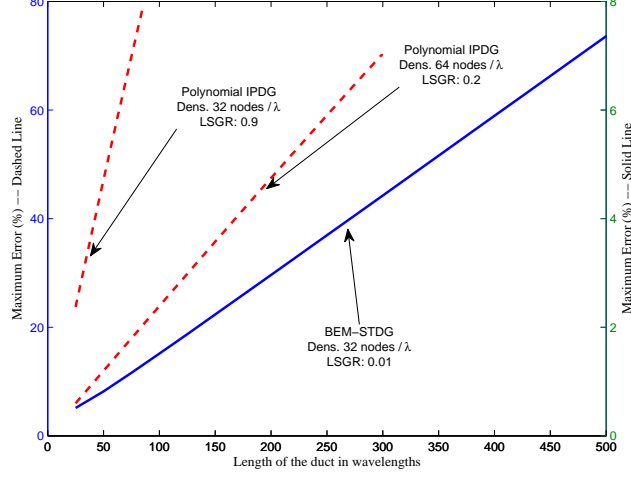


Figure 12: Maximum error in % for polynomial approximations of degree $m = 1$. The left y -axis corresponds to the error curves of the IPDG method and the right y -axis to the BEM-STDG method.

All these benchmark tests, except the one corresponding to a polynomial degree $m = 4$ and a density of 8 nodes per wavelength depicted in FIG. 15, confirm that the BEM-STDG method is able to reduce the “*pollution effect*” much more efficiently than the usual polynomial IPDG method. The case where the BEM-STDG method succeeded less than the polynomial IPDG method is that where the density was only of 8 nodes per wavelength, hence being less than the usual rule of thumb of 12 nodes per wavelength. This suggests that the BEM-STDG method requires a minimal density of nodes to be efficient.

It must also be noticed that the BEM-STDG method succeeded to practically rub out the “*pollution effect*” up to 500 wavelengths for polynomial approximations $m = 3$ and $m = 4$ with 24 nodes per wavelength (see FIG. 14 and FIG. 16), contrary to the IPDG method for which this error is still present even at a low level in some cases.

Following the suggestion of one of the anonymous referees, who drew our attention to the fact that plane wave DG methods may yield an explosive growth in the condition number as the mesh is refined, we added TAB. 3 which includes the 1-norm ($\|v\|_1 = \sum_{i=1}^N |v_i|$ of \mathbb{C}^N) condition number estimates for a duct of length 50 and 500 wavelengths, as well as the CPU time for a duct of length 500 wavelengths. It is important to note that, since the runs for each of the two methods were carried out on different platforms and using different programming languages, only the increase in CPU time for a h or a p refinement for each of the two approaches remain significant.

It can be argued from the values reported in TAB. 3 that the growth of the condition number for the BEM-STDG method is similar to that of the IPDG method. It does not blow up with a h or p refinement. The growth in CPU time when dealing with large domains seems to be more advantageous to the BEM-STDG method.

Degree	Density (nodes/ λ)	Method	Condition number		CPU time 500 λ
			50 λ	500 λ	
$m = 2$	12	IPDG	$8.8 \cdot 10^{10}$	$4.7 \cdot 10^{11}$	2.54
		BEM-STDG	$6.06 \cdot 10^7$	$1.00 \cdot 10^8$	4.76
	16	IPDG	$2.7 \cdot 10^{11}$	$7.6 \cdot 10^{11}$	6.01
		BEM-STDG	$2.01 \cdot 10^8$	$6.25 \cdot 10^8$	5.58
	24	IPDG	$1.2 \cdot 10^{12}$	$2.76 \cdot 10^{12}$	19.03
		BEM-STDG	$5.96 \cdot 10^8$	$9.42 \cdot 10^9$	8.5
$m = 3$	12	IPDG	$4.2 \cdot 10^{11}$	$8.9 \cdot 10^{11}$	2.13
		BEM-STDG	$2.1 \cdot 10^8$	$1.5 \cdot 10^9$	4.75
	18	IPDG	$1.41.2 \cdot 10^{12}$	$1.0 \cdot 10^{13}$	8.83
		BEM-STDG	$5.35 \cdot 10^8$	$3.19 \cdot 10^{10}$	6.2
	24	IPDG	$2.0 \cdot 10^{12}$	$6.2 \cdot 10^{13}$	20.81
		BEM-STDG	$9.52 \cdot 10^8$	$8.07 \cdot 10^{10}$	8.4
$m = 4$	8	IPDG	$1.43 \cdot 10^{11}$	$3.78 \cdot 10^{11}$	0.66
		BEM-STDG	$1.13 \cdot 10^8$	$1.08 \cdot 10^8$	3.89
	16	IPDG	$1.48 \cdot 10^{12}$	$6.73 \cdot 10^{13}$	5.10
		BEM-STDG	$7.84 \cdot 10^8$	$7.15 \cdot 10^{10}$	5.6
	24	IPDG	$2.38 \cdot 10^{12}$	$2.41 \cdot 10^{14}$	17.91
		BEM-STDG	$1.7 \cdot 10^9$	$1.7 \cdot 10^{11}$	8.41

Table 3: Behavior of the condition number and the CPU time for h and p refinements.

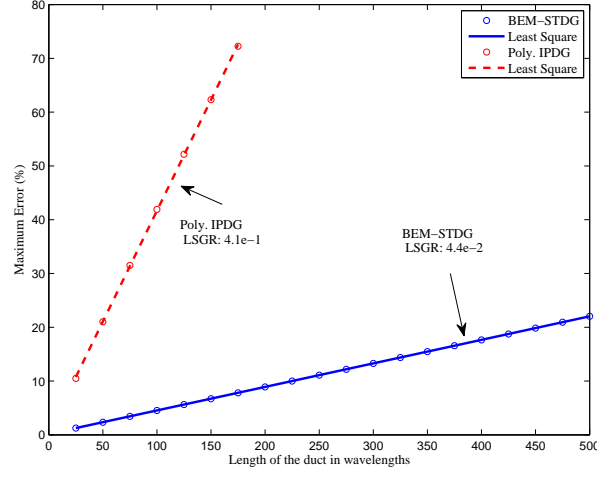


Figure 13: Maximum error in % for polynomial approximations of degree $m = 2$ and a density of 12 nodes per wavelength.

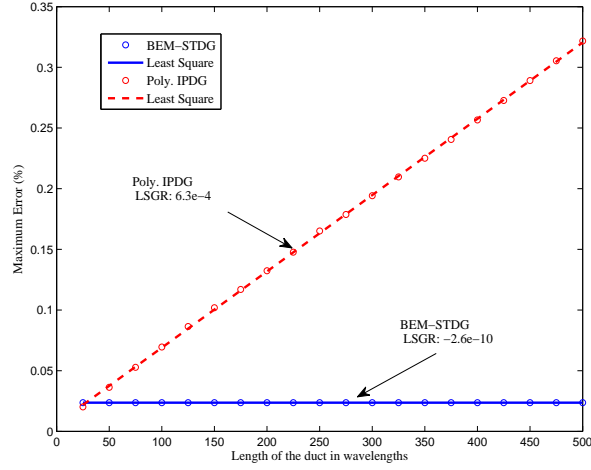


Figure 14: Maximum error in % for polynomial approximations of degree $m = 3$ and a density of 24 nodes per wavelength.

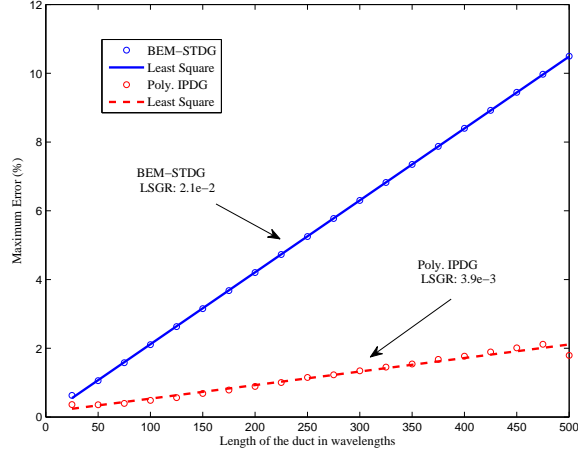


Figure 15: Maximum error in % for polynomial approximations of degree $m = 4$ and a density of 8 nodes per wavelength.

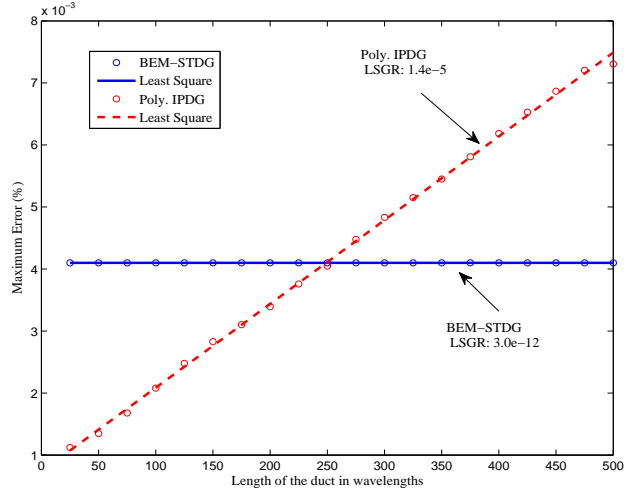


Figure 16: Maximum error in % for polynomial approximations of degree $m = 4$ and a density of 24 nodes per wavelength.

6. Concluding remarks

At first, it is worth stressing the outstanding stability of the BEM-STDG method relatively to the penalty parameters. All the results were obtained using the same set of parameters. Generally, for usual IPDG methods, these parameters have to be tuned according to geometrical features of the neighboring elements and the polynomial degree of the local shape functions.

On the other hand, this study has confirmed the expected property that a TDG method, whose local shape functions are obtained by means of a BEM, considerably reduces the so-called “*pollution effect*” instabilities. It was even shown that it is possible to completely rub out the “*pollution effect*” by a local refinement the BEM mesh, which does not affect the global size of the discrete problem, and using a BEM of modest polynomial degree. It should be noted that these excellent performances have been obtained through an extremely careful tuning of the BEM method, but done once for all when implementing the BEM code. In particular, the most difficult part of this task is an elaborate way for computing the involved singular and regular integrals. A brief idea of such a task was given above in part 3.2.7 of section 3. A complete description of the procedure will be given elsewhere. The accurate computation of the approximation of the DtN operator must be also noticed.

The current study gives also rise to several questions:

- Is it possible to replace the BEM solution by the approximation of the DtN operator through a suitable FEM? In the context of a standard FEM, such an approach is used to carry out a kind of numerical homogenization process in [53]. The results obtained here from an accurate evaluation of the DtN operator suggests that such a way to proceed should improve also this homogenization procedure.
- Is it possible to confirm the excellent reduction of the “*pollution effect*” observed for the duct problem by a study of the dispersion of the related numerical scheme, following the approach described in [4], or at least numerically as in [54]?
- Can the UWVF be dealt with using a similar procedure based on a BEM for building the local approximating functions?
- Is it possible to theoretically justify the stability of the method relatively to the size of the propagation domain?

All these issues will be studied in forthcoming papers.

- [1] F. Ihlenburg, I. Babuska, Finite element solution of the Helmholtz equation with high wave number – part i: the h-version of the FEM, *Computers Math. Applic.* 30 (9) (1995) 9–37.
- [2] X. Feng, H. Wu, Discontinuous Galerkin methods for the Helmholtz equation with large wave number., *SIAM J. Numer. Anal.* 47 (4) (2009) 2872–2896.
- [3] J. M. Melenk, A. Parsania, S. Sauter, General DG-Methods for highly indefinite Helmholtz problems, *J. Sci. Comput.* 57 (2013) 536–581.
- [4] M. Ainsworth, Dispersive and dissipative behaviour of high order discontinuous Galerkin finite element methods., *Journal of Computational Physics* 198 (2004) 106–130.
- [5] M. Ainsworth, P. Monk, W. Muniz, Dispersive and dissipative properties of discontinuous Galerkin finite element methods for the second-order wave equation., *Journal of Scientific Computing* 27 (1–3) (2006) 5–40.
- [6] A. F. Loula, G. B. Alvarez, E. G. do Carmo, F. A. Rochinha, A discontinuous finite element method at element level for Helmholtz equation, *Comput. Meth. Appl. Mech. Engrg.* 196 (2007) 867–878.

- [7] R. Hiptmair, A. Moiola, I. Perugia, A survey of Trefftz methods for the Helmholtz equation, Research Report 2015-20, SAM ETH Zurich, to appear in Springer Lecture Notes on Computational Science and Engineering (2015).
- [8] B. Pluymers, B. van Hal, D. Vandepitte, W. Desmet, Trefftz-based methods for time-harmonic acoustics, *Archives for Computational Methods in Engineering* 14 (4) (2007) 343–381.
- [9] C. J. Gittelsohn, R. Hiptmair, I. Perugia, Plane wave discontinuous Galerkin methods: analysis of the h -version, *Mathematical Modelling and Numerical Analysis* 43 (2009) 297–331.
- [10] R. Hiptmair, I. Perugia, Mixed plane wave DG methods, in: M. Bercovier, M. J. Gander, R. Kornhuber, O. Windlund (Eds.), *Domain decomposition methods in science and engineering XVIII*, Lect. Notes Comput. Sci. Eng., Springer, 2008.
- [11] R. Hiptmair, A. Moiola, I. Perugia, Trefftz discontinuous Galerkin methods for acoustic scattering on locally refined meshes, *Applied Numerical Mathematics* 79 (2014) 79–91.
- [12] B. Després, Sur une formulation variationnelle ultra-faible, *Comptes Rendus de l’Académie des Sciences Série I* 318 (1994) 939–944.
- [13] O. Cessenat, B. Després, Application of an ultra weak variational formulation of elliptic pdes to the two-dimensional Helmholtz problem, *SIAM J. Num. Analysis* 35 (1) (1998) 255–299.
- [14] G. Gabard, Discontinuous Galerkin methods with plane waves for time-harmonic problems, *Journal of Computational Physics* 225 (2007) 1961–1984.
- [15] A. Buffa, P. Monk, Error estimates for the ultra weak variational formulation of the Helmholtz equation, *Mathematical Modelling and Numerical Analysis* 42 (6) (2008) 925–940.
- [16] B. Cockburn, J. Gopalakrishnan, R. Lazarov, Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems, *SIAM J. Num. Analysis* 47 (2) (2009) 1319–1365.
- [17] G. Giorgiani, S. Fernández-Méndez, A. Huerta, Hybridizable discontinuous Galerkin p -adaptivity for wave propagation problems, *Int. J. Numer. Meth. Fluids* 72 (2013) 1244–1262.
- [18] N. C. Nguyen, J. Peraire, F. Reitich, B. Cockburn, A phase-based hybridizable discontinuous Galerkin method for the numerical solution of the Helmholtz equation, *Journal of Computational Physics* 290 (2015) 318–335.
- [19] M. Amara, H. Calandra, R. Dejlouli, M. Grigoroscuta-Strugaru, A stable discontinuous Galerkin-type method for solving efficiently Helmholtz problems, *Computers and Structures* 106–107 (2012) 258–272.
- [20] P. Gamallo, R. J. Astley, A comparison of two Trefftz-type methods: The ultraweak variational formulation and the least-squares method, for solving shortwave 2-D Helmholtz problems, *International Journal for Numerical Methods in Engineering* 71 (2007) 406–432.
- [21] D. Wang, R. Tezaur, J. Toivanen, C. Ferhat, Overview of the discontinuous enrichment method, the ultra-weak variational formulation, and the partition of unity method for the acoustic scattering in the medium frequency regime and performance comparisons, *International Journal for Numerical Methods in Engineering* 89 (2012) 403–417.
- [22] L. Kovalevsky, P. Ladevèze, H. Riou, The Fourier version of the Variational Theory of Complex Rays for medium-frequency acoustics, *Comput. Methods Appl. Mech. Engrg.* 225/228 (2012) 142–153.
- [23] J. E. Caruthers, J. C. French, G. K. Raviprakash, Green function discretization for numerical solution of the Helmholtz equation, *Journal of Sound and Vibration* 187 (4) (1995) 553–568.
- [24] C. Hofreither, U. Langer, S. Weissner, Convection-adapted bem-based fem, *arXiv:1502.05954v1 [math.NA]* 20 Feb 2015 (February 2015).
- [25] C. Hofreither, A non-standard Finite Element Method using Boundary Integral Operators, Ph.D. thesis, J. Kepler University, Linz (2012).
- [26] T. J. R. Hughes, Multiscale phenomena: Green’s functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods, *Comput. Methods Appl. Mech. Engrg.* 127 (1995) 387–401.
- [27] E. G. D. do Carmo, G. B. Alvarez, A. F. D. Loula, F. A. Rochinha, A nearly optimal Galerkin projected residual finite element method for Helmholtz problem, *Comput. Meth. Appl. Mech. Engrg.* 197 (2008) 1362–1375.
- [28] D. N. Arnold, An interior penalty finite element method with discontinuous elements, *SIAM J. Num. Analysis* 19 (4) (1982) 742–760.
- [29] D. A. D. Pietro, A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, Springer, Berlin Heidelberg, 2012.
- [30] S. A. Sauter, C. Schwab, *Boundary Element Methods*, Springer-Verlag, Berlin-Heidelberg, 2011.
- [31] A. Bendali, M. Fares, Boundary integral equations methods in acoustics, in: F. Magoules (Ed.),

- Computer Methods for Acoustics Problems, Saxe-Coburg Publications, Kippen, Stirlingshire, Scotland, 2008, Ch. 1, pp. 1–36.
- [32] A. Moiola, R. Hiptmair, I. Perugia, Plane wave approximation of homogeneous Helmholtz solutions, *Z. Angew. Math. Phys.* 62 (2011) 809–837.
 - [33] E. Spence, “When all else fails, integrate by parts”. An overview of new and old variational formulations for linear elliptic PDEs., in: A. Fokas, B. Pelloni (Eds.), *Unified transform methods for boundary value problems: applications and advances.*, SIAM, Philadelphia, 2015, pp. 93–159.
 - [34] S. Rienstra, A. Hirschberg, An introduction to acoustics, S.W. Rienstra & A. Hirschberg 2004, 2004, available on line <http://www.win.tue.nl/~sjoerdr/papers/boek.pdf>.
 - [35] T. Huttunen, J. P. Kaipio, P. Monk, The perfectly matched layer for the ultra weak variational formulation of the 3D Helmholtz equation, *International Journal for Numerical Methods in Engineering* 61 (2004) 1072–1092.
 - [36] L. D. Landau, E. M. Lifschitz, *Fluid Mechanics* 2nd edition, reprinted with corrections 2009 Edition, Vol. 6 of Landau and Lifschitz: course of theoretical physics, Elsevier, Amsterdam, 1987.
 - [37] J.-C. Nédélec, *Acoustic and Electromagnetic Equations: Integral Representations for Harmonic Problems*, Springer, Berlin, 2001.
 - [38] C. H. Wilcox, *Scattering theory for the d’Alembert equation in exterior domains*, Vol. 442, Springer-Verlag, Berlin, 1975.
 - [39] J. F. Semblat, J. J. Briost, Efficiency of higher order finite element for the analysis of seismic wave propagation, *Journal of Sound and Vibration* 231 (2) (2000) 460–467.
 - [40] W. McLean, *Strongly Elliptic Systems and Boundary Integral Equations*, Cambridge University Press, Cambridge, UK, and New York, USA, 2000.
 - [41] D. N. Arnold, F. Brezzi, B. Cockburn, L. D. Marini, Unified analysis of discontinuous Galerkin methods for elliptic problems, *SIAM J. Num. Analysis* 39 (5) (2002) 1749–1779.
 - [42] K. Zhao, M. N. Vouvakis, J.-F. Lee, Solving electromagnetic problems using a novel symmetric fem-bem approach, *IEEE Transactions on Magnetics* 42 (4) (2006) 583–586.
 - [43] C. Brebbia, W. Wendland, G. Kuhn (Eds.), *Symmetric Methods for the Coupling of Finite Elements and Boundary Elements*, Springer, Berlin, 1987.
 - [44] E. Burman, A. Ern, I. Mozolevski, B. Stamm, The symmetric discontinuous galerkin method does not need stabilization in 1d for polynomial orders $p \geq 2$, *Compt. Rend. Acad. Sciences Série I* 345 (10) (2007) 599–602.
 - [45] J. S. Hesthaven, T. Warburton, *Nodal Discontinuous Galerkin Methods. Algorithms, Analysis, and Applications*, Vol. 54 of Texts in Applied Mathematics, Springer, Berlin, Heidelberg, 2008.
 - [46] O. Cessenat, *Application d’une nouvelle formulation variationnelle aux équations d’ondes harmoniques. problèmes d’Helmholtz 2D et de Maxwell 3D.*, Ph.D. thesis, University of Paris XI Dauphine (1996).
 - [47] G. C. Hsiao, W. L. Wendland, *Boundary Integral Equations*, Springer, Berlin-Heidelberg, 2008.
 - [48] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North Holland, Amsterdam, 1978.
 - [49] J. Jin, *The Finite Element Method in Electromagnetics*, Second Edition., John Wiley & Sons, New York, 2002.
 - [50] R. Cools, A. Haegemans, Algorithm 824: Cubpack: A package for automatic cubature; framework description, *ACM Transactions on Mathematical Software* 29 (3) (2003) 287–296.
 - [51] T. Huttunen, P. Gamallo, R. J. Astley, Comparison of two wave element methods for the Helmholtz equation, *Comm. Numer. Meth. Engrg.* 25 (1) (2009) 35–52.
 - [52] M. J. Gander, F. Magoules, F. Nataf, Optimized Schwarz methods without overlap for the Helmholtz equation, *SIAM Journal of Scientific Computing* 24 (2002) 38–60.
 - [53] Y. Efendiev, T. Y. Hou, *Multiscale Finite Element Methods. Theory and Applications*, Vol. 4 of Surveys and Tutorials in the Applied Mathematical Sciences, Springer, New-York, 2009.
 - [54] C. Gittelsohn, R. Hiptmair, Dispersion analysis of plane wave discontinuous methods, *International Journal for Numerical Methods in Engineering* 98 (5) (2014) 313–323.