Computation of the exact discrete transparent boundary condition for 1D linear equations.

Sonia Fliss[a], Sébastien Imperiale[b], Antoine Tonnoir[c]

[a]POEMS (UMR 7231 CNRS-ENSTA-INRIA), ENSTA Paristech, 828 Boulevard des Maréchaux, 91762 Palaiseau, France
[b]Inria and Paris-Saclay University, 1 rue Honoré d'Estienne d'Orves, 91120 Palaiseau
[c]Normandie University, INSA Rouen Normandie, LMI, 76000 Rouen, France

Abstract

In this work, we are interested in the construction of the exact transparent boundary conditions for a semi-discretized and fully discretized 1D linear PDE. The proposed method is quite general and is based on the computation of a family of canonical functions. Several examples and numerical results to illustrate the method are presented.

Keywords: Unbounded domain, Discrete transparent boundary conditions, periodic media.

1. Introduction and model problem

In this work , we consider the general problem: Given $T > 0$, find $u \in C^m\big([0,T], H^1(-1,+\infty)\big)$ solution to

$$\left|\begin{array}{ll} D_t u(t,x) - Au(t,x) = f(t,x) & (t,x) \in [0,T] \times (-1,+\infty) \\ u(t,-1) = 0 & t \geq 0 \\ (\partial_t^i u)(0,x) = u_i(x) & (x,i) \in (-1,+\infty) \times \{0,\cdots,m-1\} \end{array}\right. \tag{P}$$

where $-A$ is a *linear* spatial differential operator of oder 2, D_t is a linear time differential operator of order $m = 1$ or 2 and f is a source term in $C^0([0,T], L^2(-1,+\infty))$. We assume the operators A and D_t to be such that problem (P) is well-posed, that is there is a unique solution that continuously depends on the data. Also, we will assume that f and the initial data $(\partial_t^i u)(0,x)$ are compactly supported in $(-1,0)$.

Let us give 3 examples covered by the general formulation (P) that motivate our study and that we will consider in this work:

- **Heat equation:** In that case, we have $D_t = \partial_t$ and $A = \partial_{xx}$

$$\left|\begin{array}{ll} \partial_t u(t,x) - \partial_{xx} u(t,x) = f(t,x) & (t,x) \in [0,T] \times (-1,+\infty) \\ u(t,-1) = 0 & t \geq 0 \\ u(0,x) = u_0(x) & x \in (-1,+\infty) \end{array}\right. \tag{H}$$

- **Klein-Gordon equation:** In that case, we have $D_t = \partial_{tt}$ and $A = \partial_{xx} - \gamma$ with $\gamma \geq 0$

$$\left|\begin{array}{ll} \partial_{tt} u(t,x) - \partial_{xx} u(t,x) + \gamma u(t,x) = f(t,x) & (t,x) \in [0,T] \times (-1,+\infty) \\ u(t,-1) = 0 & t \geq 0 \\ (\partial_t^i u)(0,x) = u_0(x) & (x,i) \in (-1,+\infty) \times \{0,1\} \end{array}\right. \tag{K}$$

- **Periodic wave equation:** In that case, we have $D_t = \partial_{tt}$ and $A = \operatorname{div}(a(x)\nabla\cdot)$

$$\left|\begin{array}{ll} \partial_{tt} u(t,x) - \operatorname{div}(a(x)\nabla u(t,x)) = f(t,x) & (t,x) \in [0,T] \times (-1,+\infty) \\ u(t,-1) = 0 & t \geq 0 \\ (\partial_t^i u)(0,x) = u_0(x) & (x,i) \in (-1,+\infty) \times \{0,1\} \end{array}\right. \tag{W}$$

  where we assume that $a(x) \in L^\infty(-1,+\infty)$ is $1-$periodic.

To numerically solve this problem and take into account the unbounded domain, several approaches can be considered. A basic idea is to compute the solution in a sufficiently large domain, provided that this solution have finite speed propagation (which is not the case for instance for the heat equation). Another method consists in solving (if possible) analytically the equation in the infinite part $(0, +\infty)$ and numerically in the bounded part $(-1, 0)$ (containing the possible perturbations, source term and initial data). Coupling the two formulations leads to construct the so-called *Dirichlet to Neumann* (DtN) operator, see for instance [1] for wave equation (for the 1D wave equation for instance, the transparent boundary condition is well-known $\partial_x u = \partial_t u$) or [5, 10] for the periodic wave equation in time harmonic regime. Usually, the exact DtN operator leads to non local boundary condition, and it can be interesting to approximate it with an absorbing condition, see for instance [9] for the heat equation and some papers in the huge literature for the wave equation [1, 8, 7, 6].

In the above mentioned works, most of the attention is paid to the construction of the absorbing / transparent boundary conditions for the continuous equation. To the best of our knowledge, quite few work has been dedicated to the construction of transparent boundary condition for discretized (D-TBC) wave equation [11]. For the Schrödinger equation or heat equation, more work have been done, see for instance [4, 2, 3]. One interest of considering the D-TBC instead of the TBC is that, since it is exact, the stability / convergence properties of the numerical scheme are directly obtained from the properties of the scheme set on the unbounded domain. In particular, this can avoid stability issues when discretizing (high order) absorbing boundary conditions. A common way to construct the D-TBC is based on the $\mathcal{Z}-$transform which mimics the role of the Fourier / Laplace transform for the time variable when constructing the continuous TBC. In this work, we propose to construct the D-TBC using a different approach which, we hope, is new and enables in particular to consider general scheme and periodic media.

Let us now give the discretization aspect for our general formulation (P). Given a time step $\Delta t > 0$ and denoting by $t_n = n\Delta t$, we consider a time discretization of (P) which leads to the semi-discretized problem: $\forall n \geq k$, find $u_n \in H^1(-1, +\infty)$ solution to

$$\left| \begin{array}{ll} \sum_{i=0}^k \alpha_i u_{n+i-k}(x) - A \sum_{i=0}^k \beta_i u_{n+i-k}(x) = \sum_{i=0}^k \beta_i f(t_{n+i-k}, x) & x \in (-1, +\infty) \\ u_i(-1) = 0 & i \geq 0 \\ u_i(x) = \widetilde{u}_i(x) & (x, i) \in (-1, +\infty) \times \{0, \cdots, k-1\} \end{array} \right. \qquad (\mathrm{P}_{\Delta t})$$

where $(\alpha_i)_{i=0,k}$ and $(\beta_i)_{i=0,k}$ are real parameters, and $(\widetilde{u}_i(x))_{i=0,k-1}$ are given initial data. We assume the time-scheme to be by *consistent* and *stable*. We mean by *consistent* that $\forall y(t, x) \in C^m([0, T], H^1(-1, +\infty))$

$$\lim_{\Delta t \to 0} \sum_{i=0}^{k} \alpha_i y(t + t_{i-k}, x) = (D_t y)(t, x) \quad \text{and} \quad \lim_{\Delta t \to 0} \sum_{i=0}^{k} \beta_i y(t + t_{i-k}, x) = y(t, x)$$

and we mean by *stable* that the norm of u_n is bounded for all n.

In the next, we will work with the equivalent variational formulation of $(\mathrm{P}_{\Delta t})$: $\forall n \geq k$, find $u_n \in H_0^1(-1, +\infty)$ s.t. $\forall v \in H_0^1(-1, +\infty)$

$$\left| \begin{array}{ll} \sum_{i=0}^k \alpha_i m(u_{n+i-k}, v) + \sum_{i=0}^k \beta_i a(u_{n+i-k}, v) = l_n(v) & \\ u_i(x) = \widetilde{u}_i(x) & i \in \{0, \cdots, k-1\} \end{array} \right. \qquad (\mathrm{P}_{\Delta t})$$

where the bilinear forms $m(\cdot, \cdot)$ and $a(\cdot, \cdot)$ are defined by

$$m(u, v) = \int_{-1}^{+\infty} u(x)v(x)\, \mathrm{d}x \quad \text{and} \quad a(u, v) = < Au, v >$$

and where $< \cdot, \cdot >$ is the dual product between $H_0^1(-1,+\infty)$ and $(H_0^1(-1,+\infty))'$. The linear form $l_n(v)$ is given by

$$l_n(v) = \sum_{i=0}^{k} \beta_i \int_{-1}^{+\infty} f(t_{n+i-k}, x)v(x)dx$$

We will assume that the bilinear forms $m(\cdot,\cdot)$ and $a(\cdot,\cdot)$ are *continuous*

$$\forall (v_1, v_2) \in H_0^1(-1,+\infty) \quad m(v_1,v_2) \le \|v_1\|_{H^1}\|v_2\|_{H^1} \quad \text{and} \quad a(v_1,v_2) \le \|v_1\|_{H^1}\|v_2\|_{H^1}$$

and the bilinear form $\alpha_k m(\cdot,\cdot) + \beta_k a(\cdot,\cdot)$ is *coercive*

$$\exists C > 0, \ \text{s.t.} \ \forall v \in H^1(-1,+\infty) \quad \alpha_k m(v,v) + \beta_k a(v,v) \ge C\|v\|_{H^1}$$

which ensures that problem ($\mathrm{P}_{\Delta t}$) is well-posed. We will also assume that the source term and the initial data are compactly supported in $(-1,0)$:

$$\forall x \ge 0, \quad f(x) = 0 \quad \text{and} \quad \widetilde{u}_i(x) = 0 \quad \forall i \in \{0,\cdots,k-1\}$$

Finally, to fully discretize ($\mathrm{P}_{\Delta t}$), let us introduce the spatial discretization. Given $\Delta x > 0$, we consider the classical Lagrange finite element space of order $r \ge 1$:

$$\mathbb{H}_0 = \{v_h \in H_0^1(-1,+\infty), \quad \text{s.t.} \quad v_h|_{[x_i,x_{i+1}]} \in \mathbb{P}_r \ \ \forall i \in \mathbb{N}\} \quad \text{where} \quad x_i = -1 + i\Delta x$$

Note that this space is (countable) infinite dimension. Then, the fully discretized problem reads: $\forall n \ge k$, find $u_{n,h} \in \mathbb{H}_0$ s.t. $\forall v_h \in \mathbb{H}_0$

$$\left| \begin{array}{ll} \sum_{i=0}^{k} \alpha_i m(u_{n+i-k,h}, v_h) + \sum_{i=0}^{k} \beta_i a(u_{n+i-k,h}, v_h) = l_n(v) & \\ u_{i,h}(x) = \mathcal{P}\widetilde{u}_i(x) & i \in \{0,\cdots,k-1\} \end{array} \right. \qquad (\mathrm{P}_{\Delta t,\Delta x})$$

where $\mathcal{P}$ is simply the Lagrange interpolation of $\widetilde{u}_i(x)$.

In the next of this paper, we will reformulate ($\mathrm{P}_{\Delta t}$) and ($\mathrm{P}_{\Delta t,\Delta x}$) in bounded domain $(-1,0)$ by constructing the D-TBC on $\{x = 0\}$. In section 2, we first reformulate the problem as a transmission problem. Then, we explain how to compute the solution in the exterior domain thanks to a family of *canonical functions* and we derive the D-TBC. In the section 3, we show numerical results to validate our method and finally, in the conclusion section, we discuss some advantage / drawback of the method, and possible extensions.

2. Bounded domain formulation

2.1. Transmission problem formulation

Let us begin by reformulating ($\mathrm{P}_{\Delta t}$) and ($\mathrm{P}_{\Delta t,\Delta x}$) as transmission problems between the domain $(-1,0)$ and $(0,+\infty)$. We denote by

$$H^- = \{v \in H^1(-1,0) \ \text{s.t.} \ v(-1) = 0\}$$
$$\text{and} \quad H^+ = H^1(0,+\infty)$$

and similarly :

$$\mathbb{H}^- = \{v_h \in H^-, \quad \text{s.t.} \quad v_h|_{[x_i,x_{i+1}]} \in \mathbb{P}_r \ \ \forall i \in \mathbb{N} \ \text{s.t.} \ x_{i+1} \le 0\}$$
$$\text{and} \quad \mathbb{H}^+ = \{v_h \in H^+, \quad \text{s.t.} \quad v_h|_{[x_i,x_{i+1}]} \in \mathbb{P}_r \ \ \forall i \in \mathbb{N} \ \text{s.t.} \ x_i \ge 0\}$$

We will denote in the next $m^\pm(\cdot,\cdot)$ and $a^\pm(\cdot,\cdot)$ the bilinear form $m(\cdot,\cdot)$ and $a(\cdot,\cdot)$ restricted to $H^\pm$ or $\mathbb{H}^\pm$

Semi-discretized problem. Let us consider the transmission problem: $\forall n \geq k$, find $u_n^{\pm} \in H^{\pm}$ and $\lambda_n \in \mathbb{R}$ s.t. $\forall v^{\pm} \in H^{\pm}$:

$$\left| \begin{array}{l} \sum_{i=0}^{k} \alpha_i m^-(u_{n+i-k}^-, v^-) + \sum_{i=0}^{k} \beta_i a^-(u_{n+i-k}^-, v^-) - (\lambda_n, v^-)_0 = l_n(v^-) \\ \sum_{i=0}^{k} \alpha_i m^+(u_{n+i-k}^+, v^+) + \sum_{i=0}^{k} \beta_i a^+(u_{n+i-k}^+, v^+) + (\lambda_n, v^+)_0 = 0 \\ u_n^-(0) = u_n^+(0) \quad \textit{(Transmission Condition)} \end{array} \right. \tag{TP$_{\Delta t}$}$$

with initial data $u_i^-(x) = \widetilde{u}_i(x)|_{(-1,0)}$ and $u_i^+(x) = 0$ for $i \in \{0, \cdots, k-1\}$. The bilinear form $(\cdot, \cdot)_0$ is simply defined by

$$(\lambda_n, v)_0 = \lambda_n v(0)$$

Proposition 2.1. *The transmission problem ($\mathrm{TP}_{\Delta t}$) is equivalent to* $(\mathrm{P}_{\Delta t})$ *in the sense that:*

- *if u_n is solution to $(\mathrm{P}_{\Delta t})$ then $(u_n^-, u_n^+, \lambda_n)$ given by*

$$u_n^- = u_n|_{(-1,0)} \quad u_n^+ = u_n|_{(0,+\infty)} \quad \lambda_n = -\sum_{i=0}^{k} \alpha_i m^+(u_{n+i-k}^+, v^+) - \sum_{i=0}^{k} \beta_i a^+(u_{n+i-k}^+, v^+)$$

  for any $v^+ \in H^-$ s.t. $v^+(0) = 1$ is solution to ($\mathrm{TP}_{\Delta t}$). Note that λ_n is independent of v^+ provided that $v^+(0) = 1$.

- *if $(u_n^-, u_n^+, \lambda_n)$ is solution to ($\mathrm{TP}_{\Delta t}$), then :*

$$u_n(x) = \left| \begin{array}{ll} u_n^-(x) & \textit{if} \quad x \in (-1, 0) \\ u_n^+(x) & \textit{if} \quad x \in (0, +\infty) \end{array} \right.$$

  is solution to $(\mathrm{P}_{\Delta t})$.

A way to show this result is to come back to strong formulation of the problem. Let us underline that $\lambda_n = \partial_x \sum_{i=0}^{k} u_{n+i-k}^-(0) = \partial_x \sum_{i=0}^{k} u_{n+i-k}^+(0)$.

Fully-discretized problem. Similarly, Let us consider the transmission problem: $\forall n \geq k$, find $u_{n,h}^{\pm} \in \mathbb{H}^{\pm}$ and $\lambda_{n,h} \in \mathbb{R}$ s.t. $\forall v_h^{\pm} \in \mathbb{H}^{\pm}$:

$$\left| \begin{array}{l} \sum_{i=0}^{k} \alpha_i m^-(u_{n+i-k,h}^-, v_h^-) + \sum_{i=0}^{k} \beta_i a^-(u_{n+i-k,h}^-, v_h^-) - (\lambda_{n,h}, v_h^-)_0 = l_n^-(v_h^-) \\ \sum_{i=0}^{k} \alpha_i m^+(u_{n+i-k,h}^+, v_h^+) + \sum_{i=0}^{k} \beta_i a^+(u_{n+i-k,h}^+, v_h^+) + (\lambda_{n,h}, v_h^+)_0 = 0 \\ u_{n,h}^-(0) = u_{n,h}^+(0) \quad \textit{(Transmission Condition)} \end{array} \right. \tag{TP$_{\Delta t,\Delta x}$}$$

with initial data $u_{i,h}^-(x) = \mathcal{P}\widetilde{u}_i(x)|_{(-1,0)}$ and $u_{i,h}^+(x) = 0$ for $i \in \{0, \cdots, k-1\}$.

Proposition 2.2. *The transmission problem ($\mathrm{TP}_{\Delta t,\Delta x}$) is equivalent to* $(\mathrm{P}_{\Delta t,\Delta x})$ *in the sense that:*

- *if $u_{n,h}$ is solution to $(\mathrm{P}_{\Delta t,\Delta x})$ then $(u_{n,h}^-, u_{n,h}^+, \lambda_{n,h})$ given by*

$$u_{n,h}^- = u_{n,h}|_{(-1,0)} \quad u_{n,h}^+ = u_{n,h}|_{(0,+\infty)} \quad \lambda_{n,h} = -\sum_{i=0}^{k} \alpha_i m^+(u_{n+i-k,h}^+, v_h^+) - \sum_{i=0}^{k} \beta_i a^+(u_{n+i-k,h}^+, v_h^+)$$

  for any $v_h^+ \in \mathbb{H}^+$ s.t. $v_h^+(0) = 1$ is solution to ($\mathrm{TP}_{\Delta t,\Delta x}$). Note that $\lambda_{n,h}$ is independent of v_h^+.

- *if* $(u^-_{n,h}, u^+_{n,h}, \lambda_{n,h})$ *is solution to* ($\mathrm{TP}_{\Delta t}$), *then :*

$$u_{n,h}(x) = \left| \begin{array}{lll} u^-_{n,h}(x) & \textit{if} & x \in (-1,0) \\ u^+_{n,h}(x) & \textit{if} & x \in (0,+\infty) \end{array} \right.$$

  is solution to $(\mathrm{P}_{\Delta t,\Delta x})$.

Before proving this result, let us underline that in that case we no more have $\lambda_{n,h} = \partial_x u^-_{n,h}(0) = \partial_x u^+_{n,h}(0)$.

Proof:
$(\Rightarrow)$ Let $u_{n,h}$ be solution to $(\mathrm{P}_{\Delta t,\Delta x})$ and define $(u^-_{n,h}, u^+_{n,h}, \lambda_{n,h})$ as in Proposition above. Then by construction we have $u^-_{n,h}(0) = u^+_{n,h}(0)$ (*transmission condition*). Moreover, we have $\forall v^-_h \in \mathbb{H}^-$ s.t. $v^-_h(0) = 0$:

$$\begin{aligned} & \sum_{i=0}^{k} \alpha_i m^-(u^-_{n+i-k,h}, v^-_h) + \sum_{i=0}^{k} \beta_i a^-(u^-_{n+i-k,h}, v^-_h) - (\lambda_{n,h}, v^-_h)_0 = l^-_n(v^-_h) \\ \Leftrightarrow \quad & \sum_{i=0}^{k} \alpha_i m(u_{n+i-k,h}, \bar{v}^-_h) + \sum_{i=0}^{k} \beta_i a(u_{n+i-k,h}, \bar{v}^-_h) = l_n(\bar{v}^-_h) \quad \text{where } \bar{v}^-_h = \left| \begin{array}{lll} v^-_h & \text{if} & x \le 0 \\ 0 & \text{if} & x > 0 \end{array} \right. \end{aligned}$$

The last equality is true since $u_{n,h}$ is solution to $(\mathrm{P}_{\Delta t,\Delta x})$ and since $\bar{v}^-_h \in \mathbb{H}$. One can show exactly a similar result for $u^+_{n,h}$ taking $\forall v^+_h \in \mathbb{H}^+$ s.t. $v^+_h(0) = 0$.

Now, for any $v^+_h \in \mathbb{H}^+$ s.t. $v^+_h(0) = 1$, denoting by

$$v^+_0 := \left| \begin{array}{lll} 1 - \frac{x}{\Delta x} & \text{if} & x \in (0, \Delta x) \\ 0 & \text{if} & x \ge \Delta x \end{array} \right.$$

we have $v^+_0 \in \mathbb{H}^+$ and

$$\begin{aligned} \lambda_{n,h} &= -\sum_{i=0}^{k} \alpha_i m^+(u^+_{n+i-k,h}, v^+_h - v^+_0 + v^+_0) - \sum_{i=0}^{k} \beta_i a^+(u^+_{n+i-k,h}, v^+_h - v^+_0 + v^+_0) \\ &= -\sum_{i=0}^{k} \alpha_i m^+(u^+_{n+i-k,h}, v^+_0) - \sum_{i=0}^{k} \beta_i a^+(u^+_{n+i-k,h}, v^+_0) \qquad (1) \\ &= -\sum_{i=0}^{k} \alpha_i m(u_{n+i-k,h}, v^+_0) - \sum_{i=0}^{k} \beta_i a(u_{n+i-k,h}, v^+_0) \end{aligned}$$

since $v^+_h - v^+_0 \in \mathbb{H}^+$ satisfies $(v^+_h - v^+_0)(0) = 0$. This shows that $\lambda_{n,h}$ is independent of the test function v^+_h.

Moreover, setting

$$v^-_0 := \left| \begin{array}{lll} 1 + \frac{x}{\Delta x} & \text{if} & x \in (-\Delta x, 0) \\ 0 & \text{if} & x \in (-1, -\Delta x) \end{array} \right.$$

we have

$$\begin{aligned} & \sum_{i=0}^{k} \alpha_i m^-(u^-_{n+i-k,h}, v^-_0) + \sum_{i=0}^{k} \beta_i a^-(u^-_{n+i-k,h}, v^-_0) - (\lambda_{n,h}, v^-_0)_0 = l^-_n(v^-_0) \\ \Leftrightarrow \quad & \sum_{i=0}^{k} \alpha_i m^-(u^-_{n+i-k,h}, v^-_0) + \sum_{i=0}^{k} \beta_i a^-(u^-_{n+i-k,h}, v^-_0) + \sum_{i=0}^{k} \alpha_i m^+(u^+_{n+i-k,h}, v^+_0) + \sum_{i=0}^{k} \beta_i a^+(u^+_{n+i-k,h}, v^+_0) = l^-_n(v^-_0) \\ \Leftrightarrow \quad & \sum_{i=0}^{k} \alpha_i m(u_{n+i-k,h}, v_0) + \sum_{i=0}^{k} \beta_i a(u_{n+i-k,h}, v_0) = l_n(v_0) \quad \text{where} \quad v_0 = \left| \begin{array}{lll} v^-_0 & \text{if} & x \in (-1,0) \\ v^+_0 & \text{if} & x \in (0,+\infty) \end{array} \right. \end{aligned}$$

The last equality is true because $v_0 \in \mathbb{H}$. To conclude this first part, we simply need to remark that for any $v_h^- \in \mathbb{H}^-$ s.t. $v_h^-(0) = 1$, we have $v_h^- = v_h^- - v_0^- + v_0^-$ and $(v_h^- - v_0^-)(0) = 0$ and use the same idea as in (1).

($\Leftarrow$) Let us now show the reciprocal. Assume that $(u_{n,h}^-, u_{n,h}^+, \lambda_{n,h})$ is solution to ($\mathrm{TP}_{\Delta t}$), then summing equations satisfied by $u_{n,h}^-$ and $u_{n,h}^+$, we get $\forall v^\pm \in \mathbb{H}^\pm$

$$\begin{aligned}
&\sum_{\pm}\left(\sum_{i=0}^{k} \alpha_i m^\pm(u_{n+i-k,h}^\pm, v^\pm) + \sum_{i=0}^{k} \beta_i a^\pm(u_{n+i-k,h}^\pm, v^\pm)\right) + (\lambda_{n,h}, v^+)_0 - (\lambda_{n,h}, v^-)_0 = l_n^-(v_0^-)\\
\Leftrightarrow\quad &\sum_{\pm}\left(\sum_{i=0}^{k} \alpha_i m^\pm(u_{n+i-k,h}, v^\pm) + \sum_{i=0}^{k} \beta_i a^\pm(u_{n+i-k,h}, v^\pm)\right) + (\lambda_{n,h}, v^+ - v^-)_0 = l_n(v_0^-)
\end{aligned}$$

using u_n defined as in Proposition. To conclude the proof, it suffices to remark that any $v \in \mathbb{H}$ can be decomposed as follows

$$v(x) = \left|\begin{array}{lll} (v(x) - v(0)v_0^-(x)) + v(0)v_0^-(x) & \text{if} & x \in (-1, 0) \\ (v(x) - v(0)v_0^+(x)) + v(0)v_0^+(x) & \text{if} & x \in (0, +\infty) \end{array}\right.$$

where $(v(x) - v(0)v_0^\pm(x)) + v(0)v_0^\pm(x) \in \mathbb{H}^\pm$. ■

2.2. The canonical function and the transparent boundary condition

Let us now focus on the unbounded part. The goal in this section is to give an "explicit" expression of u_n^+ (resp. $u_{n,h}^+$) in function of its boundary value $u_n^+(0) = u_n^-(0)$ (resp. $u_{n,h}^+(0) = u_{n,h}^-(0)$).

Semi-discretized problem. Let us introduce the *canonical function* $e_n^+(x)$: $\forall n \geq k$, $e_n^+ \in H^+$ satisfies $\forall v^+ \in H^+$ s.t. $v^+(0) = 0$

$$\left|\begin{array}{l} \sum_{i=0}^{k} \alpha_i m^+(e_{n+i-k}^+, v^+) + \sum_{i=0}^{k} \beta_i a^+(e_{n+i-k}^+, v^+) = 0 \\[2mm] e_n^+(0) = \delta_{n,k} \end{array}\right. \tag{2}$$

where $e_n^+(x) = 0$ for $n \leq k-1$. In particular, for $n = k$, we get

$$\left|\begin{array}{l} \alpha_k m^+(e_k^+, v^+) + \beta_k a^+(e_k^+, v^+) = 0 \\[2mm] e_k^+(0) = 1 \end{array}\right. \tag{3}$$

Let us emphasize that this problem has a unique solution $e_k^+ \in H^+$ if and only if $\beta_k \neq 0$. Indeed, otherwise we should have $e_k^+(x) = \delta_{x,0} \notin H^+$. This implies that the time-scheme for the *semi-discretized problem* must be implicit. In fact, this condition comes from the coercivity of the bilinear form $\alpha_k m^+(\cdot,\cdot) + \beta_k a^+(\cdot,\cdot)$.

Proposition 2.3. *The solution of the problem: $\forall n \geq k$, find $u_n^+ \in H^+$ satisfying $\forall v^+ \in H^+$ s.t. $v^+(0) = 0$*

$$\left|\begin{array}{l} \sum_{i=0}^{k} \alpha_i m^+(u_{n+i-k}^+, v^+) + \sum_{i=0}^{k} \beta_i a^+(u_{n+i-k}^+, v^+) = 0 \\[2mm] u_n^+(0) = u_n^-(0) \end{array}\right. \tag{4}$$

with $u_n^+ = 0$ for $n \in \{0, \cdots, k-1\}$, is given by the convolution

$$\forall n \geq k, \quad u_n^+(x) = \sum_{j=0}^{n-k} e_{n-j}^+(x)\, u_{j+k}^-(0) \tag{5}$$

Remark 2.4. *The condition* $u_n^+ = 0$ *for* $n \in \{0, \cdots, k-1\}$ *is necessary and means that the initial data* $\widetilde{u}_n$, $n = \{0, \cdots, k-1\}$, *in problem* $(\mathrm{P}_{\Delta t})$ *are compactly supported in* $(-1, 0)$. *Yet, for the* heat problem (H), *due to* infinite speed propagation, *we cannot expect* $u_n^+ = \widetilde{u}_n|_{(0+\infty)} = 0$ *for* $n = \{0, \cdots, k-1\}$ *if* $k \geq 2$. *In other words, for problems with* infinite speed propagation *we must consider for the semi-discretized problem a* one step implicit time scheme.

Proof: First, taking u_n^+ defined by (9), we easily see that the boundary condition is satisfied:

$$\forall n \geq k, \quad u_n^+(0) = \sum_{j=0}^{n-k} \underbrace{e_{n-j}^+(0)}_{=\delta_{n-j,k}} u_{j+k}^-(0) = u_j^-(0)$$

Then, recalling that $u_i^+ = 0$ for $i \in \{0, \cdots, k-1\}$, we have $\forall n \geq k$

$$\begin{aligned}\sum_{i=0}^{k} \alpha_i u_{n+i-k}^+ &= \sum_{i=0}^{k} \alpha_i \sum_{j=0}^{n+i-2k} e_{n+i-k-j}^+ u_{j+k}^-(0) \\ &= \sum_{j=0}^{n-2k} u_{j+k}^-(0) \sum_{i=0}^{k} \alpha_i e_{n+i-k-j}^+ + \sum_{j=n-2k+1}^{n-k} u_{j+k}^-(0) \sum_{i=j-(n-2k)}^{k} \alpha_i e_{n+i-k-j}^+\end{aligned}$$

We get a similar result with β_i and we thus deduce that $\forall v^+ \in \mathbb{H}^+$ s.t. $v^+(0) = 0$:

$$\begin{aligned}\sum_{i=0}^{k} \alpha_i m^+(u_{n+i-k}^+, v^+) + \sum_{i=0}^{k} \beta_i a^+(u_{n+i-k}^+, v^+) &= \\ &\sum_{j=0}^{n-2k} u_{j+k}^-(0) \underbrace{\left(\sum_{i=0}^{k} \alpha_i m^+(e_{n+i-k-j}^+, v^+) + \sum_{i=0}^{k} \beta_i a^+(e_{n+i-k-j}^+, v^+) \right)}_{=0} \\ &+ \sum_{j=n-2k+1}^{n-k} u_{j+k}^-(0) \underbrace{\left(\sum_{i=j-(n-2k)}^{k} \alpha_i (e_{n+i-k-j}^+, v^+) + \sum_{i=j-(n-2k)}^{k} \beta_i a^+(e_{n+i-k-j}^+, v^+) \right)}_{=0} \\ &= 0\end{aligned}$$

by definition of the *canonical functions* (2), which proves the result. ■

Fully-discretized problem. Let us now discuss the case of fully-discretized problem. We introduce the *canonical solution* $e_{n,h}^+(x)$: $\forall n \geq k$, $e_{n,h}^+ \in \mathbb{H}^+$ satisfies $\forall v_h^+ \in \mathbb{H}^+$ s.t. $v_h^+(0) = 0$

$$\left| \begin{array}{l} \sum_{i=0}^{k} \alpha_i m^+(e_{n+i-k,h}^+, v_h^+) + \sum_{i=0}^{k} \beta_i a^+(e_{n+i-k,h}^+, v_h^+) = 0 \\ e_{n,h}^+(0) = \delta_{n,k} \end{array} \right. \tag{6}$$

where $e_{n,h}^+(x) = 0$ for $n \leq k-1$. In particular, for $n = k$, we get

$$\left| \begin{array}{l} \alpha_k m^+(e_{k,h}^+, v_h^+) + \beta_k a^+(e_{k,h}^+, v^+) = 0 \\ e_{k,h}^+(0) = 1 \end{array} \right. \tag{7}$$

In that case, this problem has a unique solution $e^+_{k,h} \in \mathbb{H}^+$ even if $\beta_k = 0$. As a consequence, for the fully discretized problem, we can consider explicit scheme.

Proposition 2.5. *The solution of the problem: $\forall n \geq k$, find $u^+_{n,h} \in \mathbb{H}^+$ satisfying $\forall v^+_h \in \mathbb{H}^+$ s.t. $v^+_h(0) = 0$*

$$\left| \begin{array}{l} \sum_{i=0}^k \alpha_i m^+(u^+_{n+i-k,h}, v^+_h) + \sum_{i=0}^k \beta_i a^+(u^+_{n+i-k,h}, v^+_h) = 0 \\ u^+_{n,h}(0) = u^-_{n,h}(0) \end{array} \right. \tag{8}$$

with $u^+_{n,h} = 0$ for $n \in \{0, \cdots, k-1\}$, is given by the convolution

$$\forall n \geq k, \quad u^+_{n,h}(x) = \sum_{j=0}^{n-k} e^+_{n-j,h}(x)\, u^-_{j+k,h}(0) \tag{9}$$

Remark 2.6. *The condition $u^+_n = 0$ for $n \in \{0, \cdots, k-1\}$ is necessary and means that the initial data $\widetilde{u}_n$, $n = \{0, \cdots, k-1\}$, in problem $(\mathrm{P}_{\Delta t, \Delta x})$ are compactly supported in $(-1, 0)$. On contrary to the case of semi-discretized formulation, for explicit scheme, we have* finite speed propagation *at the discrete level. Therefore, even for the* heat problem (H)*, we can consider* multi-step time scheme*.*

The proof of Proposition 2.5 is exactly the same as the proof of Proposition 2.3.

Now, given the explicit expression of u^+_n (resp. $u^+_{n,h}$) on $(0, +\infty)$ in terms of the *canonical functions*, we can rewrite the transmission problem ($\mathrm{TP}_{\Delta t}$) (resp. ($\mathrm{TP}_{\Delta t, \Delta x}$)). Let us detail the case of the semi-discretized problem, the case of the fully discretized problem being similar. Thanks to Proposition 2.3, we get that problem ($\mathrm{TP}_{\Delta t}$) is equivalent to:

$$\left| \begin{array}{l} \sum_{i=0}^k \alpha_i m^-(u^-_{n+i-k}, v^-) + \sum_{i=0}^k \beta_i a^-(u^-_{n+i-k}, v^-) - (\lambda_n, v^-)_0 = l_n(v^-) \\ u^+_n(x) = \sum_{j=0}^{n-k} e^+_{n-j}(x)\, u^-_{j+k}(0) \\ \lambda_n = -\sum_{i=0}^k \alpha_i m^+(u^+_{n+i-k}, v^+_0) - \sum_{i=0}^k \beta_i a^+(u^+_{n+i-k}, v^+_0) \end{array} \right.$$

for any function $v^+_0 \in H^1$ s.t. $v^+_0(0) = 1$. In particular, replacing the expression of u^+_n in the formula of λ_n, we get the expression of the so-called DtN operator:

$$\begin{aligned} \lambda_n &= -\sum_{i=0}^k \alpha_i m^+\Big(\sum_{j=0}^{n+i-2k} e^+_{n+i-k-j}\, u^-_{j+k}(0), v^+\Big) - \sum_{i=0}^k \beta_i a^+\Big(\sum_{j=0}^{n+i-2k} e^+_{n+i-k-j}\, u^-_{j+k}(0), v^+\Big) \\ &= \sum_{j=0}^{n-k} \mu_{n-k-j}\; u^-_{j+k}(0) \end{aligned} \tag{$\mathrm{DtN}_{\Delta t}$}$$

where the DtN coefficient μ_j are given $\forall j \in \{0, \cdots, n-k\}$ by:

$$\mu_j = \left| \begin{array}{ll} -\sum_{i=0}^k \alpha_i m^+(e^+_{j+i}, v^+_0) - \sum_{i=0}^k \beta_i a^+(e^+_{j+i}, v^+_0) & \text{if} \quad j \geq k \\ -\sum_{i=k-j}^k \alpha_i m^+(e^+_{i+j}, v^+_0) - \sum_{i=k-j}^k \beta_i a^+(e^+_{i+j}, v^+_0) & \text{if} \quad j < k \end{array} \right. \tag{10}$$

and v^+_0 is an arbitrary function of H^+ that satisfies $v^+_0(0) = 1$ (we recall that λ_n, and similarly the coeffcient μ_j does not depend on the choice of v^+_0, see the proof of Proposition 2.2). Let us emphasize that λ_n given by the formula above corresponds exactly to the DtN operator since it maps the Dirichlet data $u^-_j(0)$ to the

Neumann data $\lambda_n = \partial_x u_n^-(0)$. Very classically, this operator is non-local in the sense that we need all the previous time-step to compute λ_n via the convolution. Moreover, let us remark that, taking $v_0^+ = e_k^+$, we get for the first coefficient μ_0:

$$\mu_0 \quad = \quad -\alpha_k m^+(e_k^+, e_k^+) - \beta_k a^+(e_k^+, e_k^+) \quad < 0 \tag{11}$$

Finally, we can rewrite our problem in bounded domain $(-1, 0)$ as follows

$$\sum_{i=0}^{k} \alpha_i m^-(u_{n+i-k}^-, v^-) + \sum_{i=0}^{k} \beta_i a^-(u_{n+i-k}^-, v^-) - (\sum_{j=0}^{n-k} \mu_{n-k-j} \ u_{j+k}^-(0), v^-)_0 = l_n(v^-) \tag{$\mathrm{P}^b_{\Delta t}$}$$

Exactly the same procedure in the fully discretize case can be applied and we get

$$\lambda_{n,h} \quad = \quad \sum_{j=0}^{n-k} \mu_{n-k-j,h} \ u_{j+k,h}^-(0) \tag{$\mathrm{DtN}_{\Delta t,\Delta x}$}$$

where the coefficients $\mu_{j,h}$ are given $\forall j \in \{0, \cdots, n-k\}$ by:

$$\mu_{j,h} = \left| \begin{array}{ll} -\sum_{i=0}^{k} \alpha_i m^+(e_{j+i,h}^+, v_{0,h}^+) - \sum_{i=0}^{k} \beta_i a^+(e_{j+i,h}^+, v_{0,h}^+) & \text{if} \quad j \geq k \\ -\sum_{i=k-j}^{k} \alpha_i m^+(e_{i+j,h}^+, v_{0,h}^+) - \sum_{i=k-j}^{k} \beta_i a^+(e_{i+j,h}^+, v_{0,h}^+) & \text{if} \quad j < k \end{array} \right. \tag{12}$$

and $v_{0,h}^+$ is an arbitrary function of $\mathbb{H}^+$ that satisfies $v_{0,h}^+(0) = 1$. As for the semi-discretize case, we can see that for the first coefficient $\mu_{0,h}$, taking $v_{0,h}^+ = e_{k,h}^+$, we have

$$\mu_{0,h} \quad = \quad -\alpha_k m^+(e_{k,h}^+, e_{k,h}^+) - \beta_k a^+(e_{k,h}^+, e_{k,h}^+) \quad < 0 \tag{13}$$

2.3. Computation of the canonical function

As we have seen, the reformulation of our initial problem in bounded domain $(-1, 0)$ requires to know the *canonical functions* defined by (2) and (6). Let us now see how we can compute these functions. The key idea will be to rewrite (2) (resp. (6)) also in bounded domain, for instance $(0, 1)$, using the *same* transparent boundary condition we wish to compute. Since the coefficients of the DtN operator depend on the canonical function, this leads to a non-linear problem (only for the first canonical function, as we will see), that can be easily solved.

Semi-discretized case. Using the same procedure as in the previous section, we can show that if e_n^+ is solution to (2), then $\forall n \geq k$ it satisfies $\forall v^+ \in H^+$ s.t. $v^+(0) = 0$

$$\left| \begin{array}{l} \sum_{i=0}^{k} \alpha_i m^{01}(e_{n+i-k}^+, v^+) + \sum_{i=0}^{k} \beta_i a^{01}(e_{n+i-k}^+, v^+) - (\sum_{j=0}^{n-k} \mu_{n-k-j} \ e_{j+k}^+(1), v^+)_1 = 0 \\ e_n^+(0) = \delta_{n,k} \end{array} \right. \tag{14}$$

where the bilinear form $m^{01}(\cdot, \cdot)$ and $a^{01}(\cdot, \cdot)$ correspond to $m(\cdot, \cdot)$ and $a(\cdot, \cdot)$ restricted to $H^1(0, 1)$, and the coefficients μ_j are given by (10). The question is: is solving (14) equivalent to solve (2) ?

Remark 2.7. *For problems* (H) *and* (K)*, we could have consider problem* (14) *on any segment* $(0, c)$ *with* $c > 0$ *(instead of segment* $(0, 1)$*). This is in particular useful to reduce the computational cost of the canonical functions, as we will see.*

Let us begin with the case $n = k$, for which we have $\forall v^+ \in H^+$ s.t. $v^+(0) = 0$

$$\left| \begin{array}{l} \alpha_k m^{01}(e_k^+, v^+) + \beta_k a^{01}(e_k^+, v^+) - (\mu_0 e_k^+(1), v^+)_1 = 0 \\ \mu_0 = -\alpha_k m^+(e_k^+, v_0^+) - \beta_k a^+(e_k^+, v_0^+) \\ e_k^+(0) = 1 \end{array} \right. \tag{15}$$

where $v_0^+ \in H^+$ s.t. $v_0^+(0) = 1$. To solve the non linear problem: find $e_k^+ \in H^1(0,1)$ and $\mu_0 \in \mathbb{R}$ solution to (15), we will need to add a condition on $e^k(x)$. Indeed, as we will see the above problem as it is has two solutions.

Lemma 2.8. *The solution e_k^+ to (3) satisfies $|e^+(1)| < 1$.*

Proof: Let us denote by $v_0^+ \in H^+$ the function defined by

$$v_0^+(x) = \left| \begin{array}{lll} 1 - x & \text{if} & x \in (0,1) \\ 0 & \text{if} & x \geq 1 \end{array} \right.$$

On one hand we have:

$$\begin{array}{rcl} \mu_0 & = & -\alpha_k m^+(e_k^+, v_0^+) - \beta_k a^+(e_k^+, v_0^+) \\ & = & -\alpha_k m^{01}(e_k^+, v_0^+) - \beta_k a^{01}(e_k^+, v_0^+) \end{array}$$

because $v(x) = 0 \ \forall x \geq 1$, and on the other hand, using $e_k^+ = e_k^+ - v_0^+ + v_0^+$:

$$\begin{array}{rcl} \alpha_k m^{01}(e_k^+, e_k^+) + \beta_k a^{01}(e_k^+, e_k^+) & = & -\mu_0 + \alpha_k m^{01}(e_k^+, e_k^+ - v_0^+) + \beta_k a^{01}(e_k^+, e_k^+ - v_0^+) \\ & = & -\mu_0 + (\mu_0 e_k^+(1), e_k^+ - v_0^+)_1 \\ & = & -\mu_0(1 - (e_k^+(1)^2) \end{array}$$

because, as we said, e_k^+ satisfies (15) and $(e_k^+ - v_0^+)(0) = 0$. To conclude the proof, we simply need to remark that the above term must be strictly positive so that, since $-\mu_0 > 0$, see (13), we get

$$1 - (e_k^+(1))^2 > 0 \quad \Leftrightarrow \quad |e_k^+(1)| < 1$$

■

Remark 2.9. *This lemma could be proved for any value $x_0 > 0$, that is $|e_k^+(x_0)| < 1$.*

Theorem 2.10. *The problem: find $e_k^{01} \in H^1(0,1)$ and $\mu_0 \in \mathbb{R}$ that satisfies $\forall v^{01} \in H^1(0,1)$ s.t. $v^{01}(0) = 0$:*

$$\left| \begin{array}{l} \alpha_k m^{01}(e_k^{01}, v^{01}) + \beta_k a^{01}(e_k^{01}, v^{01}) - (\mu_0 e_k^{01}(1), v^{01})_1 = 0 \\ \mu_0 = -\alpha_k m^{01}(e_k^{01}, v_0^+) - \beta_k a^{01}(e_k^{01}, v_0^+) \\ e_k^{01}(0) = 1 \end{array} \right. \tag{16}$$

where $v_0^+ \in H^1(0,1)$ s.t. $v_0^+(0) = 1$ and $v_0^+(1) = 0$, and that also satisfies $|e_k^{01}(1)| < 1$ has a unique solution given by $e_k^{01} = e_k^+|_{(0,1)}$.

Proof: To start, let us introduce the following problem: Find $\widetilde{e} \in H^1(0,1)$ that satisfies $\forall v \in H_0^1(0,1)$

$$\left| \begin{array}{l} \alpha_k m^{01}(\widetilde{e}, v) + \beta_k a^{01}(\widetilde{e}, v) = 0 \\ \widetilde{e}(0) = d_0 \\ \widetilde{e}(1) = d_1 \end{array} \right.$$

where (d_0, d_1) are Dirichlet data. This problem is well-posed and we will denote by A^{01} the operator that maps $(d_0, d_1) \in \mathbb{R}^2 \to A(d_1, d_1) = \widetilde{e} \in H^1(0,1)$. This operator is *bilinear* and *continuous*.

This operator introduced, we can equivalently rewrite problem (16) as follows: Find $e^{01}(1)$ and μ_0 satisfying $\forall v^{01} \in H^1(0,1)$ s.t. $v^{01}(0) = 0$:

$$\left| \begin{array}{l} \alpha_k m^{01}(A(1, e_k^{01}(1)), v^{01}) + \beta_k a^{01}(A(1, e_k^{01}(1)), v^{01}) - (\mu_0 e_k^{01}(1), v^{01})_1 = 0 \\ \mu_0 = -\alpha_k m^{01}(A(1, e_k^{01}(1)), v_0^+) - \beta_k a^{01}(A(1, e_k^{01}(1)), v_0^+) \end{array} \right.$$

Injecting the expression of μ_0 in the first equation and since $A(1, e_k^{01}(1)) = A(1,0) + e_k^{01}(1)A(0,1)$, we deduce that $e_k^{01}(1)$ satisfies a second ordre polynomial equation:

$$C_2 \left(e_k^{01}(1)\right)^2 + C_1 e_k^{01}(1) + C_0 = 0$$

where

$$\begin{aligned} C_0 &= \alpha_k m^{01}(A(1,0), v^{01}) + \beta_k a^{01}(A(1,0), v^{01}) \\ C_1 &= \alpha_k m^{01}(A(0,1), v^{01}) + \beta_k a^{01}(A(0,1), v^{01}) + v^{01}(1) \left(\alpha_k m^{01} A(1,0), v_0^+) + \beta_k a^{01}(A(1,0), v_0^+)\right) \\ C_2 &= v^{01}(1) \left(\alpha_k m^{01}(A(0,1), v_0^+) + \beta_k a^{01}(A(0,1), v_0^+)\right) \end{aligned}$$

Let us remark that, by definition of A, if we take v^{01} s.t. $v^{01}(1) = 0$, then the polynomial equation becomes $0 = 0$. Let us take instead $v^{01} = A(0,1)$, it satisfies $v^{01}(0) = 0$ and $v^{01}(1) = 1$, and $v_0^+ = A(1,0)$, which satisfies $v_0^+(0) = 1$. Then, the coefficient are given by:

$$\begin{aligned} C_0 &= \alpha_k m^{01}(A(1,0), A(0,1)) + \beta_k a^{01}(A(1,0), A(0,1)) \\ C_1 &= \alpha_k m^{01}(A(0,1), A(0,1)) + \beta_k a^{01}(A(0,1), A(0,1)) + \alpha_k m^{01}(A(1,0), A(1,0)) + \beta_k a^{01}(A(1,0), A(1,0)) \\ C_2 &= \alpha_k m^{01}(A(0,1), A(1,0)) + \beta_k a^{01}(A(0,1), A(1,0)) \end{aligned}$$

We can note that $C_0 = C_2$ and since

$$\begin{aligned} & \alpha_k m^{01}(A(0,1) + A(1,0), A(1,0) + A(0,1)) + \beta_k a^{01}(A(0,1), A(0,1)) > 0 \\ \Leftrightarrow \quad & C_1 > -2C_0 \quad \Rightarrow C_1^2 - 4C_0 C_2 > 0 \end{aligned}$$

we deduce that

- either the polynomial equation has two distinct real solutions r_1 and r_2 that satisfy $r_1 r_2 = 1$. In that case, the solution of problem (16) is uniquely determined by the fact that $|e_k^+(1)| < 1$
- or $C_0 = C_2 = 0$ and there is only one root given by $e_k^+(1) = 0$.

This result concludes the proof of the Theorem. ■

Now we have explain how to construct the first canonical function e_k^{01} (which corresponds to the restriction of e_k^+ on $(0,1)$), let us see how to compute the next canonical functions e_n^+ for $n > k$. In fact, solving (14) for $n > k$ knowing μ_j and e_j^+ for $j \in \{k, \cdots, n-1\}$ consists in solving the linear problem: $\forall n > k$, find $e_n^+ \in H^1(0,1)$ and $\mu_{n-k} \in \mathbb{R}$ that satisfies $\forall v^+ \in H^1(0,1)$ s.t. $v^+(0) = 0$

$$\left| \begin{array}{l} \alpha_k m^{01}(e_n^+, v^+) + \beta_k a^{01}(e_n^+, v^+) - (\mu_0 e_n^+, v^+)_1 - (\mu_{n-k}\, e_k, v^+)_1 = g_n \\ \mu_{n-k} + \alpha_k m^{01}(e_n^+, v_0^+) + \beta_k a^{01}(e_n^+, v_0^+) = \nu_n \\ e_n^+(0) = 0 \end{array} \right. \tag{17}$$

where $v_0^+ \in H^1(0,1)$ is such that $v_0^+(0)=1$ and $v_0^+(1)=0$. The data g_n is given by:

$$g_n = -\sum_{i=0}^{k-1} \alpha_i m^{01}(e^+_{n+i-k}, v^+) - \sum_{i=0}^{k-1} \beta_i a^{01}(e^+_{n+i-k}, v^+) + (\sum_{j=1}^{n-k-1} \mu_{n-k-j}\ e^+_{j+k}, v^+)_1$$

and the data ν_{n-k} is given by:

$$\nu_n = \left| \begin{array}{ll} -\sum_{i=0}^{k-1} \alpha_i m^{01}(e^+_{n-k+i}, v_0^+) - \sum_{i=0}^{k-1} \beta_i a^{01}(e^+_{n-k+i}, v_0^+) & \text{if} \quad n \geq 2k \\ -\sum_{i=2k-n}^{k-1} \alpha_i m^{01}(e^+_{i+n-k}, v_0^+) - \sum_{i=2k-n}^{k-1} \beta_i a^{01}(e^+_{i+n-k}, v_0^+) & \text{if} \quad n < 2k \end{array} \right.$$

Proposition 2.11. *For all $n > k$, the problem: find $e_n^{01} \in H^1(0,1)$ and $\mu_{n-k} \in \mathbb{R}$ that satisfies $\forall v^{01} \in H^1(0,1)$ s.t. $v^{01}(0) = 0$*

$$\left| \begin{array}{l} \alpha_k m^{01}(e_n^{01}, v^{01}) + \beta_k a^{01}(e_n^{01}, v^{01}) - (\mu_0 e_n^{01}, v^{01})_1 - (\mu_{n-k} e_k^{01}, v^{01})_1 = g_n \\ \mu_{n-k} + \alpha_k m^{01}(e_n^{01}, v_0^{01}) + \beta_k a^+(e_n^{01}, v_0^{01}) = \nu_n \\ e_n^{01}(0) = 0 \end{array} \right. \tag{18}$$

with

$$g_n = -\sum_{i=0}^{k-1} \alpha_i m^{01}(e^{01}_{n+i-k}, v^{01}) - \sum_{i=0}^{k-1} \beta_i a^{01}(e^{01}_{n+i-k}, v^{01}) + (\sum_{j=1}^{n-k-1} \mu_{n-k-j}\ e^{01}_{j+k}, v^{01})_1$$

and

$$\nu_n = \left| \begin{array}{ll} -\sum_{i=0}^{k-1} \alpha_i m^{01}(e^{01}_{n-k+i}, v_0^+) - \sum_{i=0}^{k-1} \beta_i a^{01}(e^{01}_{n-k+i}, v_0^+) & \textit{if} \quad n \geq 2k \\ -\sum_{i=2k-n}^{k-1} \alpha_i m^{01}(e^{01}_{i+n-k}, v_0^+) - \sum_{i=2k-n}^{k-1} \beta_i a^{01}(e^{01}_{i+n-k}, v_0^+) & \textit{if} \quad n < 2k \end{array} \right.$$

and where μ_0 and e_k^{01} are given by Theorem 2.10, has a unique solution given by $e_n^{01} = e_n^+|_{(0,1)}$.

Proof: Since $e_n^+|_{(0,1)}$ is solution of the problem, it remains to show that the solution is unique. Indeed, we can check that the homogeneous problem has zero as unique solution, because $\mu_0 < 0$. ■

To end this part, let us finally emphasize that all what we have explain could be also done considering the reformulation (14) not on $(0,1)$ but on any interval $(0,l)$ where $l > 0$, except for the periodic-wave problem (W) for which we need to compute the canonical function on a period.

Fully discretized case. In that case, everything is the same. We can also rewrite the definition of the canonical function as (14). Note that the non-linear problem satisfied by e_k^{01} has also as unique solution $e_k^{01} = e_k^+|_{(0,1)}$. But, in some particular discretization cases, we can simply it into a linear problem. Indeed, if we consider an explicit scheme, (16) becomes

$$\left| \begin{array}{l} \alpha_k m^{01}(e_k^{01}, v^{01}) - (\mu_0 e_k^{01}(1), v^{01})_1 = 0 \\ \mu_0 = -\alpha_k m^{01}(e_k^{01}, v_0^+) \\ e_k^{01}(0) = 1 \end{array} \right. \tag{19}$$

Then, using mass-lumping technique, meaning a proper choice of the basis function $\{\varphi_0, \cdots, \varphi_N$ of $\mathbb{H}^{01}$ and appropriate quadrature formula s.t.

$$m^{01}(\varphi_j, \varphi_i) = \delta_{ij} \quad \forall (\varphi_)$$

we straightforwardly get $e_k^{01} = \varphi_0$ as solution of (19).

3. Numerical results

To finish, let us show some numerical applications of the method proposed in this work. For the discretization parameters, we take in each case $\Delta t = 10^{-3}$, a mesh size $\Delta x = 10^{-2}$ and 4^{th} order finite element. The initial condition is given in each case by $u_0(x) = e^{-100(x+0.5)^2}(= u_1$ for wave problems). To illustrate the impact of the D-TBC, we have compared the results obtained in a larger domain.

3.1. Heat problem

For the *heat equation*, we consider the following implicit scheme:

$$m(\frac{u_{n+1} - u_n}{\Delta t}, v_h) + a(\frac{u_{n+1} + u_n}{2}, v_h) = l_n(v_h) \tag{20}$$

which corresponds to take $a_1 = \frac{1}{\Delta t}$, $a_0 = -\frac{1}{\Delta t}$, and $\beta_0 = \beta_1 = \frac{1}{2}$. To compute the *canonical functions* and the DtN coefficients, we solve problem (16) and problem (18) on the segment $(0, c)$. On Figure 1, we have represented the first hundred coefficients μ_n for different values of c. We should have for each value of c the same results for μ_n. Yet, as we can see, this is true only for $c \geq 4h$. We think the problem comes from numerical stability issues.

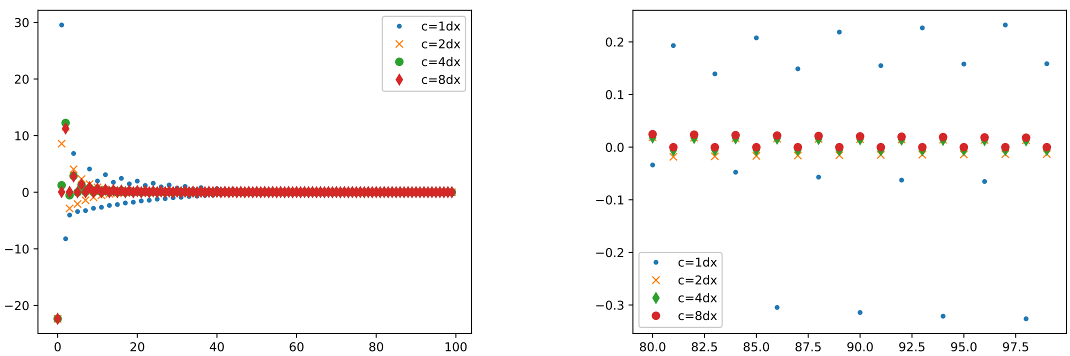

Figure 1: First hundred coefficients μ_n (on the left) and zoom on the last twenty coefficients (on the right) for $c = \{\Delta x, 2\Delta x, 4\Delta x, 8\Delta x\}$ for the Heat equation.

On figure 2, we have represented the solution of (20) (or more precisely its reformulation with D-TBC) computed with two sizes of the computational domain ($(-1, 0)$ and $(-1, 2)$) at different time. For the precomputation of the DtN coefficients, we have taken $c = 32\Delta x$. As we can see, the difference between the two computed solution in $(-1, 0)$ remains small but grows. We think this is due to numerical error in the computation of the coefficients μ_n.

3.2. Klein-Gordon problem

Here, we have considered the following time-scheme:

$$m(\frac{u_{n+1} - 2u_n + u_{n-1}}{\Delta t^2}, v_h) + a(u^n, v_h) = l_n(v_h) \tag{21}$$

which is stable under CFL condition. This corresponds to take $a_2 = a_0 = \frac{1}{\Delta t^2}$, $a_1 = -\frac{2}{\Delta t^2}$, $\beta_0 = \beta_2 = 0$ and $\beta_1 = 1$. For the value of parameter γ, we have taken $\gamma = 10$. As for the Heat equation, we have first study the computation of the DtN coefficients. On Figure 3, we represent the first hundred coefficients μ_n for various values of c. Here, we can see that the results are much better and we truly obtain similar values !

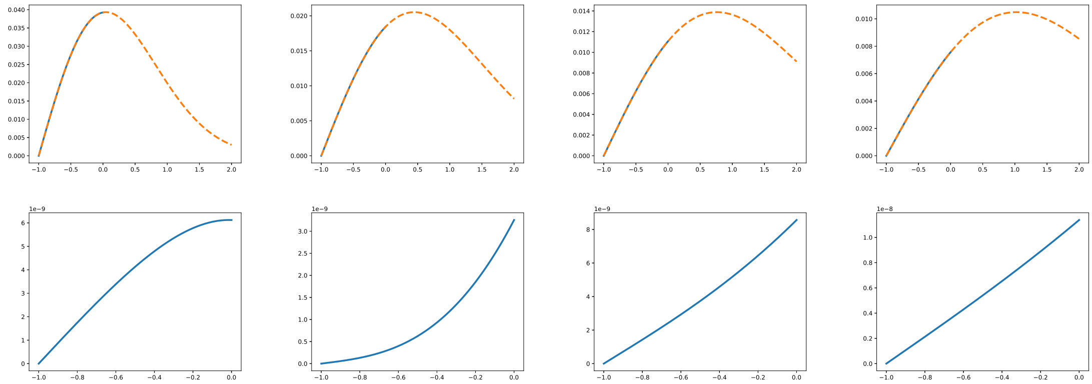

Figure 2: Comparison of the computed solution with D-TBC for the Heat equation using two size of the domain $(-1,0)$ (in blue continuous line) and $(-1,2)$ (in orange dashed line) at different time $t \in \{0.5, 1, 1.5, 2\}$ (from left to right). Below, we represented the absolute difference between the two solutions in $(-1,0)$ (magnitude of the error of order 10^{-8}).

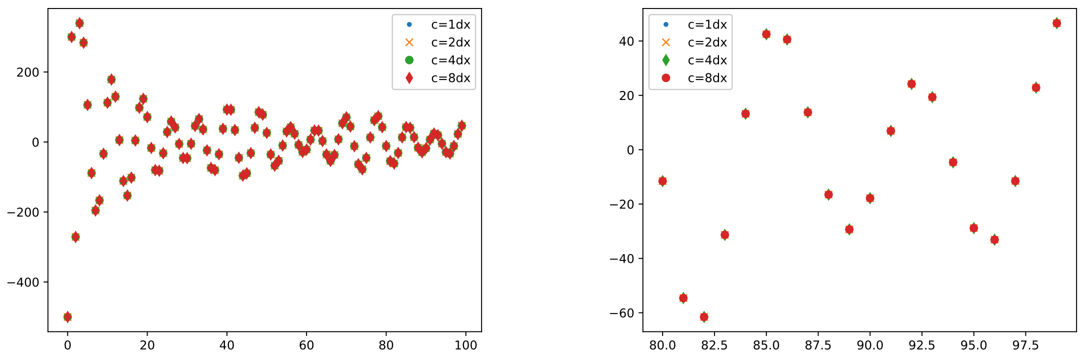

Figure 3: First hundred coefficients μ_n (on the left) and zoom on the last twenty coefficients (on the right) for $c = \{\Delta x, 2\Delta x, 4\Delta x, 8\Delta x\}$ for the KG problem.

Using the coefficients of the DtN obtained with $c = \Delta x$, we have computed the solution of (21) in two domains ($(-1,0)$ and $(-1,2)$). The comparison results are illustrated on figure 4 and, as we can see, the results are quite good since the difference between the two solutions stays with an order of 10^{-12}.

3.3. *Periodic wave problem*

Finally, let us illustrate with our final example: the periodic case. We have considered the same scheme (21) as for the KG problem. We suppose that the function $a(x)$ is 1-periodic and is given by:

$$a(x) = \left| \begin{array}{lcl} 2 & \text{if} & x \in [0.4, 0.8] \\ 1 & \text{if not} & \end{array} \right.$$

In that case, as we previously explained, we must take $c = k$ with $k \in \mathbb{N}^*$ to take into account the periodicity of the exterior domain. On Figure 5, we have represented the coefficients μ_n computed with $c = 1$ and $c = 2$. As we can expect, the results obtained are the same.

Now, as for the two previous example, we have compared the solutions computed in two domains. On Figure 6, we can see that they give very closed results, the difference being of order 10^{-12}.

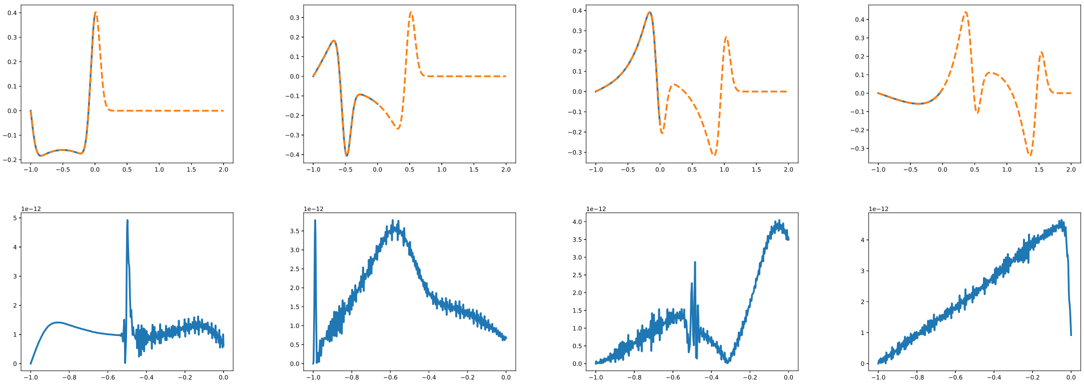

Figure 4: Comparison of the computed solution with D-TBC for the KG problem using two size of the domain $(-1,0)$ (in blue continuous line) and $(-1,2)$ (in orange dashed line) at different time $t \in \{0.5, 1, 1.5, 2\}$ (from left to right). Below, we represented the absolute difference between the two solutions in $(-1,0)$ (magnitude of the error of order 10^{-12}).

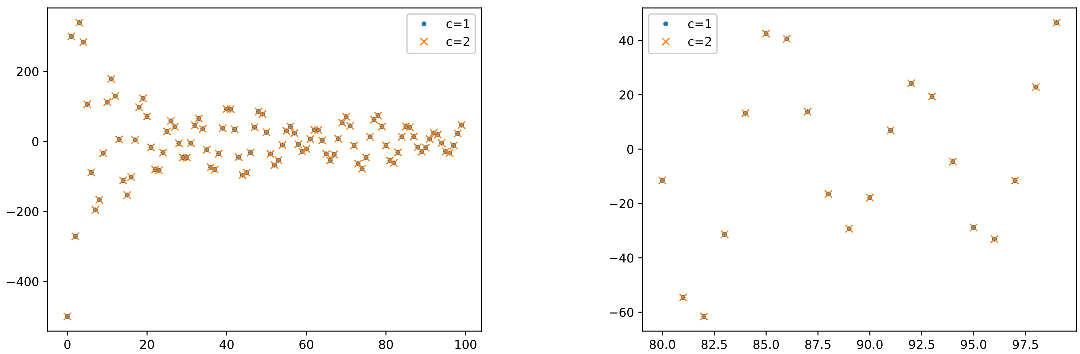

Figure 5: First hundred coefficients μ_n (on the left) and zoom on the last twenty coefficients (on the right) for $c = \{\Delta x, 2\Delta x, 4\Delta x, 8\Delta x\}$ for the Periodic wave problem.

4. Some concluding remarks

In this work, we have proposed a new approach for computing the D-TBC based on the computation of canonical functions. The main novelty is the idea of using the DtN to compute the canonical functions in the exterior domain (and therefore the D-DtN itself). We have shown that for only one problem we have to solve a non linear equation (which turns to be a polynomial of order 2). One interesting point of this method compared to $\mathcal{Z}-$transform approaches for computing D-TBC is the fact it does not required to use quadrature (in cases where numerical computations must be done to invert the $\mathcal{Z}$-transform).

Also, the idea of the method can be easily extended to waveguide problems (the coefficients μ_n would be replaced by matrices and we would solve problems on slice of the waveguide instead of interval $(0, c)$). Yet, the implementation and the analysis of the method in waveguide case are more complicated and are still under study. Clearly, the boundary conditions would also become much more costly. However, this cost can be worth for problems where no approximate or exact boundary conditions can be computed (for instance periodic waveguide or anisotropic elastic media).

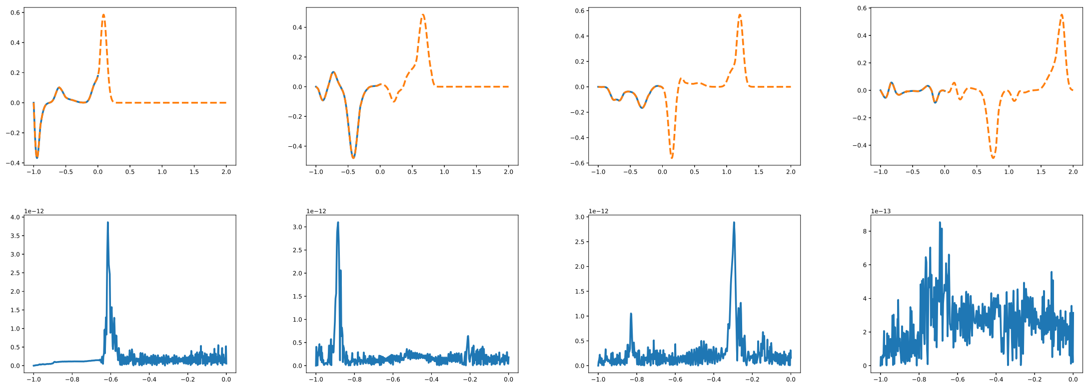

Figure 6: Comparison of the computed solution with D-TBC for the Periodic Wave problem using two size of the domain $(-1,0)$ (in blue continuous line) and $(-1,2)$ (in orange dashed line) at different time $t \in \{0.5, 1, 1.5, 2\}$ (from left to right). Below, we represented the absolute difference between the two solutions in $(-1,0)$ (magnitude of the error of order 10^{-12}).

It would be also interesting to extend the approach to non-linear equation. One difficulty in that case would be to generalize the definition of the canonical functions since it relies on the linearity of the problem.

References

[1] Bradley Alpert, Leslie Greengard, and Thomas Hagstrom. Nonreflecting boundary conditions for the time-dependent wave equation. *Journal of Computational Physics*, 180(1):270–296, 2002.

[2] Xavier Antoine, Anton Arnold, Chritophe Besse, Matthias Ehrhardt, and Achim Schädle. A review of transparent and artificial boundary conditions techniques for linear and nonlinear schrödinger equations. 2008.

[3] Anton Arnold and Matthias Ehrhardt. Discrete transparent boundary conditions for the schrödinger equation. 2001.

[4] Matthias Ehrhardt. Discrete artificial boundary conditions. 2002.

[5] Sonia Fliss and Patrick Joly. Exact boundary conditions for time-harmonic wave propagation in locally perturbed periodic media. *Applied Numerical Mathematics*, 59(9):2155–2178, 2009.

[6] Dan Givoli. *Numerical methods for problems in infinite domains*. Elsevier, 2013.

[7] Thomas Hagstrom. Radiation boundary conditions for the numerical simulation of waves. *Acta numerica*, 8:47–106, 1999.

[8] Thomas Hagstrom, Assaf Mar-Or, and Dan Givoli. High-order local absorbing conditions for the wave equation: Extensions and improvements. *Journal of computational physics*, 227(6):3322–3357, 2008.

[9] Houde Han and Zhongyi Huang. A class of artificial boundary conditions for heat equation in unbounded domains. *Computers & Mathematics with applications*, 43(6-7):889–900, 2002.

[10] Patrick Joly, Jing-Rebecca Li, and Sonia Fliss. Exact boundary conditions for periodic waveguides containing a local perturbation. *Commun. Comput. Phys*, 1(6):945–973, 2006.

[11] Ludwig Wagatha. On boundary conditions for the numerical simulation of wave propagation. *ApNM*, 1:309–314, 1985.