Algebraic Way to Derive Discrete Absorbing Boundary Conditions for Wave Equation

Jukka Tuomela, Olivier Vacus

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Algebraic way to derive discrete absorbing boundary conditions for wave equation

Jukka Tuomela , Olivier Vacus

N° 3053

Décembre 1996

THÈME 4

ISSN 0249-6399

INRIA
ROCQUENCOURT

Algebraic way to derive discrete absorbing boundary conditions for wave equation

Jukka Tuomela *, Olivier Vacus †

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Ondes

Rapport de recherche n ° 3053 — Décembre 1996 — 37 pages

Abstract: We introduce a new algebraic framework to derive discrete absorbing boundary conditions for wave equation in the monodimensional case. The idea is to factor directly the discrete wave operator and then use one of the factors as boundary condition. We also analyse the stability of the schemes obtained this way and perform numerical simulations to estimate their practical value.

Key-words: Wave equation - absorbing boundary conditions - polynomial ideals

(Résumé : tsvp)

* Helsinki University of Technology, ESPOO, Finlande.
† INRIA, Domaine de Voluceau, B.P.105 78153 Le Chesnay cedex, France.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33 1) 39 63 55 11 – Télécopie : (33 1) 39 63 53 30

Conditions limites absorbantes pour l'équation des ondes : une approche discrète algébrique

Résumé : Nous développons dans ce rapport une approche algébrique nouvelle pour l'obtention de conditions limites absorbantes pour l'équation des ondes discrète. L'idée consiste à factoriser directement l'opérateur discret et à utiliser les facteurs pour définir les conditions limites. La stabilité des schémas ainsi obtenus est analysée, et leurs performances illustrées par de nombreuses expériences numériques.

Mots-clé : Equations des ondes - conditions limites absorbantes - idéaux de polynomes

Contents

1 Introduction

We will present here in more detail the algebraic approach to absorbing boundary conditions which was outlined in [26] and [24]. The idea is to work directly in a discrete framework and to imitate the factoring idea of the continuous case. Recall that in one-dimensional case the factorization $\partial_{tt} - \partial_{xx} = (\partial_t - \partial_x)(\partial_t + \partial_x)$ provides the absorbing boundary conditions and in many dimensional case the calculus of pseudodifferential operators is used to obtain the factors, see [4], [5], [19], [21]. Now the conditions given by pseudodifferential operators are non-local in space and in time and hence cannot be used numerically as such. To circumvent this difficulty one approximates the symbol by a rational function, which then leads to a differential, therefore implementable, operator. There are numerous articles dealing with this aspect of the problem, see for instance [4], [5], [9], [10], [12], [23] and references therein.

In spite of this, we are only aware of few articles, namely [9], [10], [12], which analyze the actual discretization of the conditions thus obtained (there is also [22] which discusses discretizations of the boundary conditions in general). So how should one go about discretizing an absorbing boundary condition? Are some ways better than others? Note that pseudodifferential calculus leads to improve the boundary condition by changing the underlying differential boundary operator.

Now all the discrete boundary conditions that will be presented below are consistent approximations of the same differential operator ; hence the 'algebraization' of the problem can be interpreted as an attempt to compare directly different discretizations of the same operator. There is also another important aspect. Recall that in usual analytical treatment of the problem the relevant statements are mostly of asymptotic nature, i.e. they are pertinent only if the discretization parameter is close to zero. However, in actual computations one would like to choose this parameter as large as possible to save the computational cost. But the algebraic structure of the problem does not change when the values of the parameter is changed, and therefore the algebraic treatment can be seen as a complement to the more traditional analysis.

Let us then mention the work of Frank [6], [7], who has also studied factorization of discrete operators. His theory of difference operators is directly inspired by the theory of pseudodifferential operators and consequently his tools are analytic throughout. Hence we believe that there is no direct relation between his theory and the ideas presented below, although more subtle connection is perhaps possible. Recall also that our goal is to design good absorbing boundary conditions and not to approximate pseudodifferential operators. However, his theory, like the Kreiss theory [16], [8], [11], could be useful in more sophisticated stability analysis of our boundary conditions. Indeed, the stability analysis given below is rather elementary and based heavily on explicit computation of various quantities.

Let us then outline the contents of the article. In section 2 we will set up the algebraic framework using some standard notions of commutative algebra and develop some tools needed in the sequel. In section 3 we show that there is a reasonable way (in fact many ways) to define the order of the absorbing boundary condition and then show that there exist conditions of any (difference) order for any discretization of the wave operator. Then we discuss the possibility of combining different orders to design a suitable boundary condition. In section 4 we give the results of actual computations which give the boundary conditions and formulate some conjectures concerning the general form of the coefficients. In section 5 we analyze the stability of the schemes obtained using asymptotic expansions ; this leads to a familiar problem of determining if the roots of some polynomials are in the left half-plane. We prove in some cases the (in)stability of the schemes for small time steps, and in others offer some numerical evidence for (in)stability. In section 6 we show the results of some actual numerical experiments and conclude with section 7 where we discuss some future perspectives.

Acknowledgment We have used extensively *Axiom* [13], *Maple* [2] and *Mathematica* [30] to perform various computations.

2 Algebraic background

2.1 Notations

Let us consider the mappings from $h\mathbb{Z}^m$ to $\mathbb{R}$, denoted by $C(h\mathbb{Z}^m, \mathbb{R})$ and introduce the operators $\mathcal{S}_{i+}$ (forward shift in the direction i) and $\mathcal{S}_{i-}$ (backward shift in the direction i). We will be interested in the operators of the form $\mathcal{T} = \sum a_k \mathcal{S}^k$ where there are only finite number of terms in the sum, $k \in \mathbb{Z}^m$ and $\mathcal{S}^k = \prod_{i=1}^m \mathcal{S}_{i+}^{k_i}$ where $\mathcal{S}_{i+}^{k_i} = \mathcal{S}_{i-}^{-k_i}$ if $k_i < 0$. We shall call such operators *discrete operators*. Let us call the *support* of the operator $\mathcal{T} = \sum a_k \mathcal{S}^k$ the set $S(\mathcal{T}) = \{k \in \mathbb{Z}^m \,|\, a_k \neq 0\}$ and define $m_i = \min\{k_i \,|\, k \in S(\mathcal{T})\}$ and

$M_i = \max\{k_i \mid k \in S(\mathcal{T})\}$. The *hull* of the support is the set $H(S(\mathcal{T})) = \{k \in \mathbb{Z}^m \mid m_i \leq k_i \leq M_i\}$. Finally we define $V(\mathcal{T}) = \prod_{i=1}^{m}(M_i - m_i + 1)$.

Now every discrete operator can be expressed in terms of differences. Let $\delta_{i+} = \mathcal{S}_{i+} - 1$ (the forward difference operator), $\delta_{i-} = 1 - \mathcal{S}_{i-}$ (the backward difference operator), $i = 1, \ldots, m$, $\delta_+ = (\delta_{1+}, \ldots, \delta_{m+})$ and $\delta_- = (\delta_{1-}, \ldots, \delta_{m-})$. Further let us introduce multi-index $\nu = (\nu_1^+, \nu_1^-, \ldots, \nu_m^+, \nu_m^-) \in \mathbb{N}^{2m}$ and define $\delta^\nu = \prod_{i=1}^{m} \delta_{i+}^{\nu_i^+} \delta_{i-}^{\nu_i^-}$. Then every operator $\mathcal{T} = \sum a_k \mathcal{S}^k$ can also be expressed as $\mathcal{T} = \sum b_\nu \delta^\nu$. It will be convenient to regard δ_{i+} etc, not only as operators, but also as indeterminates in some polynomial ring. However, we must first take care of some technical details since $\delta_{i+}\delta_{i-} = \delta_{i+} - \delta_{i-}$. We shall use some standard terminology which can be found in any textbook of algebra.

2.2 Symbol of an operator

Let $\mathbb{K}$ denote a field of characteristic zero and $\mathbb{K}[\delta_+, \delta_-]$ the ring of polynomials of $2m$ indeterminates over this field. Let $\mathcal{I}_\delta = Id(\delta_{1+}\delta_{1-} - \delta_{1+} + \delta_{1-}, \ldots, \delta_{m+}\delta_{m-} - \delta_{m+} + \delta_{m-})$ be the ideal generated by the polynomials $\delta_{1+}\delta_{1-} - \delta_{1+} + \delta_{1-}, \ldots, \delta_{m+}\delta_{m-} - \delta_{m+} + \delta_{m-}$. Then we can define

Definition 2.1 *The* symbol *of an operator $\mathcal{T} = \sum b_\nu \delta^\nu$ is $T_\delta = \sum b_\nu \delta^\nu$ considered as an element of residue class ring $\mathbb{K}[\delta_+, \delta_-]/\mathcal{I}_\delta$.*

By the support of the symbol we shall mean the support of the corresponding operator. Now the question arises, what is a convenient way to represent the elements of $\mathbb{K}[\delta_+, \delta_-]/\mathcal{I}_\delta$. Let us introduce the notations $\nu_i = (\nu_i^+, \nu_i^-) \in \mathbb{N}^2$, $|\nu_i| = \nu_i^+ + \nu_i^-$ and $|\nu| = \sum_{i=1}^{m} |\nu_i|$. We call $|\nu_i|$ the *partial degree* of the monomial δ^ν with respect to i, $|\nu|$ its *total degree*. Two monomials δ^ν and δ^μ are said to be of equal degree if $|\nu_i| = |\mu_i|$ for all i. We found the following definition useful.

Definition 2.2 *Representation of $T_\delta \in \mathbb{K}[\delta_+, \delta_-]/\mathcal{I}_\delta$ is called* efficient *if it can be written in the form $\sum a_k B_k$ and the following conditions are satisfied.*

- *every B_k is either a monomial or sum of the monomials of the same degree*
- *the degrees of B_k and B_l are different if $k \neq l$*
- *the support of each B_k belongs to $H(\{0\} \cup S(\mathcal{T}))$*

The B_k's are said to be basis elements *of the symbol.*

Evidently if $0 \in H(S(\mathcal{T}))$ then the number of terms in efficient representation is at most $V(\mathcal{T})$. Note that by sum of monomials we do not mean 'any linear combination', but 'linear combination where all coefficients are equal to one'.

It is easy to see that efficient representations exist at least sometimes and that they are not unique. For instance we have $\mathcal{S}_{i+} - 2\mathcal{S}_{i-} \cong -1 + 3\delta_{i+} + \delta_{i+}\delta_{i-} = -1 + 3(\delta_{i+} + \delta_{i-})/2 - \delta_{i+}\delta_{i-}/2$ and as can be seen both are efficient representations. Of course all representations are not efficient : for instance taking $B = \delta_{1+}\delta_{1-}$, $C_1 = \delta_{1+}^2 + \delta_{1-}^2$ and $C_2 = \delta_{1+}^2\delta_{1-}^2$ we have the identity $2B = C_1 - C_2$ where only the left hand side is efficient.

Proposition 2.1 *Given any set $S \subset \mathbb{Z}^m$ there exists an efficient basis with which all symbols whose support is in $H(S)$ can be represented.*

Proof. For the sake of simplicity we suppose that $0 \in H(S)$. Define the following sets (there are several possible choices) :

$$\begin{aligned} B_1 &= \{\nu_1 \in \mathbb{N}^2 \mid \nu_1 = (s, 0),\ s = 0, \ldots, M_1\} \\ C_1 &= \{\nu_1 \in \mathbb{N}^2 \mid \nu_1 = (M_1, s),\ s = 1, \ldots, -m_1\} \\ A_1 &= B_1 \cup C_1 \end{aligned}$$

Let us call $A_1 \subset \mathbb{N}^2$ the set of admissible ν_1. Similarly one can define $\{A_i\}_{i=1..m}$ the sets of admissible ν_i's for all i. Then we take the cartesian product of these admissible sets and call this set $A \subset \mathbb{N}^{2m}$. Now it is straightforward to check that any representation of the form $\sum_{\nu \in A} a_\nu \delta^\nu$ is an efficient representation.

It remains to prove that the symbol of any operator whose support is in $H(S)$ can be represented this way. It is clearly sufficient to prove that any given monomial can be represented this way. So given a monomial $\delta^{\hat{\nu}}$ we have to show that it can be expressed as $\delta^{\hat{\nu}} = \sum_{\nu \in A} a_\nu \delta^\nu$. Suppose that $\hat{\nu} \notin A$ - if $\hat{\nu} \in A$ there is nothing to

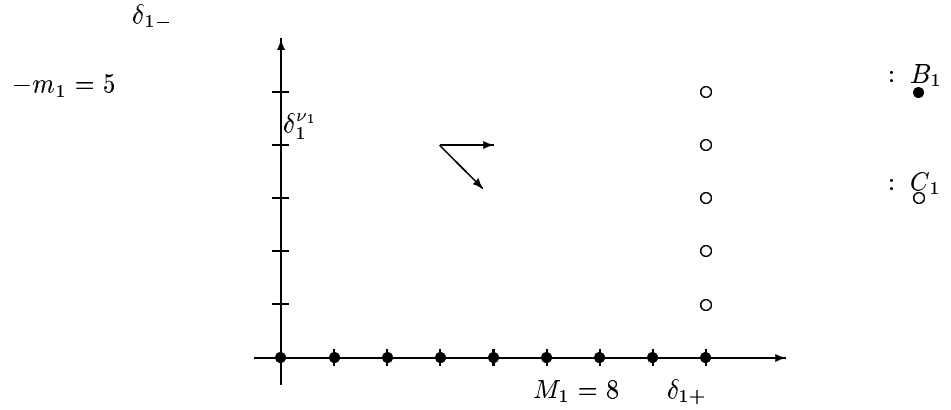

Figure 1: Pictorial proof that the reduction process terminates.

prove. By construction this means that $\hat{\nu}_i \notin A_i$ for at least on i. Without loss of generality let us suppose that $i = 1$, and that $\hat{\nu}_i = 0$ for $i > 1$ (since A has the structure of cartesian product, one can work with one index at a time).

To decompose $\delta_{1+}^{\hat{\nu}_1^+}\delta_{1-}^{\hat{\nu}_1^-}$ on our basis we use the identity

$$\delta_{1+}^{\hat{\nu}_1^+}\delta_{1-}^{\hat{\nu}_1^-} = \delta_{1+}^{\hat{\nu}_1^+ +1}\delta_{1-}^{\hat{\nu}_1^- -1} - \delta_{1+}^{\hat{\nu}_1^+ +1}\delta_{1-}^{\hat{\nu}_1^-}$$

The monomials on the right are 'closer' to being in the required basis than the one on the left as can be seen on figure 1. So continuing this process we will sooner or later have a sum on the right all of whose monomials are in the basis. □

Note that the minimal total degree in the sum in the reduction process above is the total degree of the initial monomial. This suggests the following

Definition 2.3 *Given the efficient representation of the symbol constructed above, the difference order of the operator is the minimal total degree of the monomials in the representation.*

Of course difference order can be defined in various other ways, but it is also nice to have purely algebraic definition in this context. From the computational point of view the following characterization is also useful.

Lemma 2.1 *Given any representation of the symbol, substitute repeatedly $\delta_{i+} := \delta_{i+}\delta_{i-} + \delta_{i-}$ until the lowest degree monomial is δ^ν where $\nu_i^+ = 0$ for all i ; then the difference order of the corresponding operator is $|\nu|$.*

Proof. When the lowest degree monomials are as stated one cannot anymore combine them such that only higher order monomials remain. On the other hand given any symbol, it cannot represent lower order operator than its lowest degree monomials. □

We shall encounter below some special types of symbols.

Definition 2.4 *Let $P^t(\delta_+, \delta_-) = P(-\delta_-, -\delta_+)$; P^t is called the* transpose *of P. A symbol is said to be* symmetric *(resp.* antisymmetric*) if it has a representation which satisfies $P^t = P$ (resp. $P^t = -P$).*

So in particular $\delta_{i+} = -\delta_{i-}^t$. With symmetric (antisymmetric) symbols it will be convenient to use an efficient representation where every basis element is symmetric (antisymmetric). This is always possible as the following proposition shows.

Proposition 2.2 *Given any set $S \subset \mathbb{Z}^m$ there exists a symmetric (antisymmetric) efficient basis with which all symmetric (antisymmetric) symbols whose support is in $H(S)$ can be represented.*

Proof. Note that if P is symmetric or antisymmetric then necessarily $M_i = -m_i$ for all i, so let us first consider operators whose supports are symmetric. In this case the construction of proposition 2.1 yields a set A which can be characterized as follows

$$A = \{\nu \in \mathbb{N}^{2m} \mid 0 \leq \nu_i^+ - \nu_i^- \leq 1,\ \nu_i^+ \leq M_i\}$$

Now let us define following sets

$$A_\nu = \{\mu \in \mathbb{N}^{2m} \mid -1 \leq \mu_i^+ - \mu_i^- \leq 1,\ |\mu_i| = |\nu_i|,\ \nu \in A\}$$

and define $\Gamma_\nu = \sum_{\mu \in A_\nu} \delta^\mu$. This yields an efficient basis where every basis element is either symmetric or antisymmetric. Now given any symmetric (resp. antisymmetric) symbol and representing it in this basis, we see that the coefficients of the antisymmetric (resp. symmetric) basis elements must be zero. □

As an immediate consequence of the above construction we have

Corollary 2.1 *The difference order of a symmetric (resp. antisymmetric) operator is the minimal total degree of the basis elements in the symmetric (resp. antisymmetric) representation constructed above.*

2.3 Another kind of symbol

Every discrete operator can be written in many ways, and consequently one can associate different kinds of symbols to it. We shall consider one alternative to the 'difference approach' considered above. So instead of indeterminates $\delta_\pm$ let us introduce $\sigma_{i+} = \mathcal{S}_{i+} + 1$ ('forward mean value operator') and $\sigma_{i-} = 1 + \mathcal{S}_{i-}$ ('backward mean value operator'). Then defining $\mathcal{I}_\sigma = Id(\sigma_{1+}\sigma_{1-} - \sigma_{1+} - \sigma_{1-}, \ldots, \sigma_{m+}\sigma_{m-} - \sigma_{m+} - \sigma_{m-})$, we see that the class of symbols is (isomorphic to) $\mathbb{K}[\sigma_+, \sigma_-]/\mathcal{I}_\sigma$. Starting from this we can develop the same kind of formal properties as was done above with $\delta_\pm$ as indeterminates.

Definition 2.5 *The* integrence order *of operator $\mathcal{T}$ is n if it maps to zero all $f \in C(\mathbb{Z}^m, \mathbb{R})$ where $f(k) = (-1)^{|k|}p(k)$, p is any polynomial of (total) degree less than n and $|k| = \sum_{i=1}^m k_i$.*

The integrence order can be computed as follows (cf. lemma 2.1) :

Lemma 2.2 *Given any representation of the symbol, substitute repeatedly $\sigma_{i+} := \sigma_{i+}\sigma_{i-} - \sigma_{i-}$ until the lowest degree monomial is σ^ν where $\nu_i^+ = 0$ for all i ; then the integrence order of the corresponding operator is $|\nu|$.*

Proof. Like in the difference case, if the lowest degree monomials satisfy the given condition, they cannot be combined to yield only higher degree monomials. □

Note that an operator can have both difference and integrence orders greater than zero. As an easy example consider the symbol $T = 4\delta_{i+}\delta_{i-} + \delta_{i+}^2\delta_{i-}^2 \cong -4\sigma_{i+}\sigma_{i-} + \sigma_{i+}^2\sigma_{i-}^2$, in which case both orders are two.

3 Factoring of the discrete wave operator

3.1 Difference factors

We will now use the machinery developed above to analyse the implementation of the exact absorbing boundary condition $u_t + u_x = 0$. Let us consider the symbol of the scheme of order $2p$ which can be written as

$$T_p^\delta = \delta t_+ \delta t_- - \alpha^2 \sum_{k=1}^{p} \gamma_k \delta x_+^k \delta x_-^k = \delta t_+ \delta t_- - \alpha^2 T_p^{\delta x} \tag{3.1}$$

where $\alpha = h_t/h$, the time-step divided by space-step, and the coefficients γ_k are given by

$$\gamma_k = (-1)^k \frac{2}{(2k)!} \prod_{l=1}^{k-1} (l^2 - \alpha^2)$$

The stability condition for all p is $\alpha \leq 1$, see [20] and [27]. We shall use the notation δx_+ instead of δ_{i+} etc when it is more convenient ; sometimes we shall also use the notation $O(\delta^n)$ (resp. $O(\delta x^n)$) to denote the terms whose orders (resp. partial orders with respect to x) are at least n.

Recall that the exact absorbing boundary condition was obtained by factoring the wave operator. We shall apply the same idea in the discrete setting and try to factor the symbol T_p^δ (or the corresponding operator) in such a way that the other factor represents left going signal and the other the right going signal. For definiteness let us try to determine the right going factor and then use this factor as a boundary operator. Let us agree on the following terminology.

Definition 3.1 *The boundary operator $\mathcal{B}$ is called* practical *if its symbol does not contain any positive powers of δx_+ and if it contains δt_+ and not higher powers of δt_+.*

Consequently a practical operator uses only information which has already been computed. We would like to find P and $Q \in \mathbb{K}[\delta_+, \delta_-]/\mathcal{I}_\delta$ such that P is practical and $R^\delta = T_p^\delta - PQ$ is 'as small as possible'. It is rather easy to see that there are no P and Q such that $R^\delta = 0$. One possible way to characterize smallness algebraically is the following

Definition 3.2 *If P is practical, it is said to be a* difference factor *of order n if $R^\delta = O(\delta^{n+2})$.*

Note that T_p^δ itself is second order symbol. Then we have

Theorem 3.1 *For all n there exist difference factors of order n.*

Proof. We shall construct the factors inductively. Let $P_1 = \delta t_+ + \alpha \delta x_-$ and $Q_1 = \delta t_- - \alpha \delta x_+$. Then

$$\begin{aligned} R_1^\delta &= T_p^\delta - P_1 Q_1 = \alpha(\delta x_- \delta t_- - \delta x_+ \delta t_+) \\ &= -\alpha(\delta t_- \delta t_+ \delta x_- + \delta t_- \delta x_- \delta x_+ + \delta t_- \delta t_+ \delta x_- \delta x_+) \end{aligned}$$

and the order of the remainder is three as it should. Let us then suppose that for $P_{n-1} = \delta t_+ + \alpha \bar{P}_{n-1}$ and $Q_{n-1} = \delta t_- - \alpha \bar{Q}_{n-1}$ we have

$$T_p^\delta - P_{n-1} Q_{n-1} = O(\delta^{n+1})$$

We construct P_n and Q_n using the ansatz

$$\begin{aligned} \bar{P}_n &= \bar{P}_{n-1} + \tilde{P}_n = \bar{P}_{n-1} + a^p_{(n,0)} \delta x_-^n + \sum_{k=1}^{n} a^p_{(n,k)} \delta x_-^{n-k} \delta t_-^{k-1} \delta t_+ \\ \bar{Q}_n &= \bar{Q}_{n-1} + (-1)^{n+1} \tilde{Q}_n = \bar{Q}_{n-1} + (-1)^{n+1} \Big(a^p_{(n,0)} \delta x_+^n + \sum_{k=1}^{n} a^p_{(n,k)} \delta x_+^{n-k} \delta t_+^{k-1} \delta t_- \Big) \end{aligned}$$

Note that $\bar{Q}_n = -\bar{P}_n^t$. Then we compute

$$\begin{aligned} T_p^\delta - P_n Q_n &= \alpha \Big(\delta t_+ \bar{Q}_{n-1} - \delta t_- \bar{P}_{n-1} + (-1)^{n+1} \delta t_+ \tilde{Q}_n - \delta t_- \tilde{P}_n \Big) + \\ &\quad \alpha^2 \Big(- T_p^{\delta x} + \bar{P}_{n-1} \bar{Q}_{n-1} + (-1)^{n+1} \bar{P}_{n-1} \tilde{Q}_n + \bar{Q}_{n-1} \tilde{P}_n + (-1)^{n+1} \tilde{P}_n \tilde{Q}_n \Big) \\ &= \alpha \Big(\delta t_+ \bar{Q}_{n-1} - \delta t_- \bar{P}_{n-1} + (-1)^{n+1} \delta t_+ \tilde{Q}_n - \delta t_- \tilde{P}_n \Big) + \\ &\quad \alpha^2 \Big(- T_p^{\delta x} + \bar{P}_{n-1} \bar{Q}_{n-1} + (-1)^{n+1} \delta x_- \tilde{Q}_n + \delta x_+ \tilde{P}_n \Big) + O(\delta^{n+2}) \end{aligned}$$

To simplify further we must consider separately the cases n odd and n even. Using the lemma 3.1 below we get

$$\begin{aligned} T_p^\delta - P_{2s-1} Q_{2s-1} &= \alpha^2 \Big(- T_p^{\delta x} + \bar{P}_{2s-2} \bar{Q}_{2s-2} + \delta x_- \tilde{Q}_{2s-1} + \delta x_+ \tilde{P}_{2s-1} \Big) + O(\delta^{2s+1}) \\ T_p^\delta - P_{2s} Q_{2s} &= \alpha \Big(\delta t_+ \bar{Q}_{2s-1} - \delta t_- \bar{P}_{2s-1} - \delta t_+ \tilde{Q}_{2s} - \delta t_- \tilde{P}_{2s} \Big) + O(\delta^{2s+2}) \end{aligned}$$

Now using the substitutions $\delta t_+ := \delta t_+ \delta t_- + \delta t_-$ and $\delta x_+ := \delta x_+ \delta x_- + \delta x_-$ sufficiently many times we have

$$\begin{aligned} -T_p^{\delta x} + \bar{P}_{2s-2} \bar{Q}_{2s-2} &= -\gamma_s \delta x_-^{2s} + \sum_{k=0}^{2s} b_k^{2s} \delta t_-^k \delta x_-^{2s-k} + O(\delta^{2s+1}) \\ \delta t_+ \bar{Q}_{2s-1} - \delta t_- \bar{P}_{2s-1} &= \sum_{k=0}^{2s+1} b_k^{2s+1} \delta t_-^k \delta x_-^{2s-k+1} + O(\delta^{2s+2}) \\ \delta x_- \tilde{Q}_{2s-1} + \delta x_+ \tilde{P}_{2s-1} &= 2 \sum_{k=0}^{2s-1} a^p_{(2s-1,k)} \delta t_-^k \delta x_-^{2s-k} + O(\delta^{2s+1}) \\ -\delta t_+ \tilde{Q}_{2s} - \delta t_- \tilde{P}_{2s} &= -2 \sum_{k=0}^{2s} a^p_{(2s,k)} \delta t_-^{k+1} \delta x_-^{2s-k} + O(\delta^{2s+2}) \end{aligned} \tag{3.2}$$

where b_k^j are some known constants and we put $\gamma_s = 0$ if $s > p$. Using these and lemma 3.2 below we obtain

$$\begin{aligned} T_p^\delta - P_{2s-1}Q_{2s-1} &= \alpha^2\Big(\Big(-\gamma_s + b_0^{2s} + 2a^p_{(2s-1,0)}\Big)\delta x_-^{2s} + \\ &\qquad \sum_{k=1}^{2s-1}\Big(b_k^{2s} + 2a^p_{(2s-1,k)}\Big)\delta t_-^k \delta x_-^{2s-k}\Big) + O(\delta^{2s+1}) \\ T_p^\delta - P_{2s}Q_{2s} &= \alpha\sum_{k=1}^{2s+1}\Big(b_k^{2s+1} - 2a^p_{(2s,k-1)}\Big)\delta t_-^k \delta x_-^{2s-k+1} + O(\delta^{2s+2}) \end{aligned}$$

Consequently $T_p^\delta - P_nQ_n = O(\delta^{n+2})$ if we choose

$$\begin{cases} a^p_{(2s-1,k)} = -b_k^{2s}/2, & 1 \le k \le 2s-1 \\ a^p_{(2s-1,0)} = (\gamma_s - b_0^{2s})/2 & \\ a^p_{(2s,k)} = b_{k+1}^{2s+1}/2, & 0 \le k \le 2s \end{cases}$$

□

Lemma 3.1

$$\begin{aligned} -T_p^{\delta x} + \bar{P}_{2s-1}\bar{Q}_{2s-1} &= O(\delta^{2s+2}) \\ -\delta x_-\tilde{Q}_{2s} + \delta x_+\tilde{P}_{2s} &= O(\delta^{2s+2}) \\ \delta t_+\bar{Q}_{2s} - \delta t_-\bar{P}_{2s} &= O(\delta^{2s+3}) \\ \delta t_+\tilde{Q}_{2s+1} - \delta t_-\tilde{P}_{2s+1} &= O(\delta^{2s+3}) \end{aligned}$$

Proof. To prove the first claim we note that $-T_p^{\delta x} + \bar{P}_{2s-1}\bar{Q}_{2s-1}$ is symmetric and that by induction hypothesis its order is at least $2s+1$. By proposition 2.2 we can represent it using an efficient symmetric basis which gives

$$-T_p^{\delta x} + \bar{P}_{2s-1}\bar{Q}_{2s-1} = \sum d_k B_k + O(\delta^{2s+2})$$

where the basis elements B_k contain only monomials with degree $2s+1$. But this is impossible in the symmetric basis and hence $d_k = 0$ for all k. In the same way $-\delta x_-\tilde{Q}_{2s} + \delta x_+\tilde{P}_{2s}$ is symmetric and the second claim follows. In the third one we have an antisymmetric symbol and writing it using the efficient antisymmetric basis we obtain

$$\delta t_+\bar{Q}_{2s} - \delta t_-\bar{P}_{2s} = \sum d_k B_k + O(\delta^{2s+3})$$

where the basis elements B_k contain only monomials with degree $2s+2$. This is impossible in the antisymmetric basis and hence $d_k = 0$ for all k. The last claim involves again antisymmetric symbol and the proof is the same as above. □

Lemma 3.2 $b_0^{2s+1} = b_{2s+1}^{2s+1} = b_{2s-1}^{2s} = b_{2s}^{2s} = 0$ *for all* s *and* $a^p_{(n,n)} = 0$ *for all* n *and* p.

Proof. Consider the second equation in (3.2) ; the form $\delta t_+\bar{Q}_{2s-1} - \delta t_-\bar{P}_{2s-1}$ implies that every term must contain δt_+ or δt_- and this property does not change while using the substitutions. Hence $b_0^{2s+1} = 0$ for all s.

The rest is proved by induction. It can readily be checked that $b_3^3 = a^p_{(1,1)} = b_3^4 = b_4^4 = 0$. So suppose that the claim is true up to $2s-1$. Hence all monomials in $\bar{P}_{2s-2}$ contain δx_-. But then all monomials in $-T_p^{\delta x} + \bar{P}_{2s-2}\bar{Q}_{2s-2}$ contain δx^2 which in turn implies that $b_{2s-1}^{2s} = b_{2s}^{2s} = 0$, and hence $a^p_{(2s-1,2s-1)} = 0$. But then all monomials in $\bar{P}_{2s-1}$ contain δx_- and consequently $\delta t_+\bar{Q}_{2s-1} - \delta t_-\bar{P}_{2s-1}$ contains δx_- which means that $b_{2s+1}^{2s+1} = 0$ and hence $a^p_{(2s,2s)} = 0$. □

Note that the proof of theorem 3.1 is algorithmic ; the actual computations described below were done exactly

as in the proof. We see also the major advantage of using differences and not shifts, namely the coefficients are solved one at a time and we never have big linear systems as is usual when 'undetermined coefficients with shifts' are used. For same reasons it is also much easier to construct high order schemes (for interior equation) with differences than with shifts as was pointed out in [27] and [28].

Note that we a priori postulated that the factors are in some sense symmetric. This choice was evidently guided by the principle that both factors should have same kind of properties ; in fact Q_n gives the absorbing boundary condition on the left (but 'backwards in time'!).

The following properties are immediate consequences of the proof of theorem 3.1.

Corollary 3.1 *The boundary conditions produced by the proof of theorem 3.1 can be written in the following form*

$$\mathcal{B}_s^p = \delta t_+ + \alpha\left(\sum_{n=1}^{s} a^p_{(n,0)}\delta x_-^n + \sum_{n=2}^{s}\sum_{k=1}^{n-1} a^p_{(n,k)}\delta x_-^{n-k}\delta t_+\delta t_-^{k-1}\right) \tag{3.3}$$

Moreover, $a^1_{(n,k)} \in \mathbb{Q}$ *and for* $p > 1$, $a^p_{(n,k)} \in \mathbb{Q}[\alpha^2]$.

Corollary 3.2 *Let* $p > q$ *; then* $a^p_{(n,k)} = a^q_{(n,k)}$ *for* $n - k \leq 2q$.

3.2 Integrence factors

Let us then consider factoring the symbol of the wave operator in $\mathbb{K}[\sigma_+, \sigma_-]/\mathcal{I}_\sigma$. For instance T_1^δ transformed into an element of $\mathbb{K}[\sigma_+, \sigma_-]/\mathcal{I}_\sigma$ is

$$T_1^\sigma = \sigma t_+\sigma t_- - \alpha^2\sigma x_+\sigma x_- - 4(1-\alpha^2)$$

As expected, this is zeroth order symbol. Then we define

Definition 3.3 *The boundary operator* $\mathcal{B}$ *is called* practical *if its symbol does not contain any positive powers of* σx_+ *and if it contains* σt_+ *and not higher powers of* σt_+.

Now we would like to find P and $Q \in \mathbb{K}[\sigma_+, \sigma_-]/\mathcal{I}_\sigma$ such that P is practical and $R^\sigma = T_p^\sigma - PQ$ is 'small'. Like in difference case this leads to the following

Definition 3.4 *If* P *is practical, it is said to be an* integrence factor *of order* n *if* $R^\sigma = O(\sigma^n)$.

It is easy to compute practical factors using the same algorithm as in the proof of theorem 3.1. We calculated some of them and found that they yield unstable schemes, which was perhaps to be expected. However, we shall mix the two approaches and modify the difference factors obtained above to get a factor whose integrence order is as big as possible. More precisely we make the following ansatz

$$\mathcal{C} = \mathcal{B}_s^p + \alpha\left(a\delta x_-^{s+1} + b\delta x_-^s\delta t_+ + c\delta x_-^{s-1}\delta t_+\delta t_-\right)$$

Note that this does not modify the difference order of the boundary condition. Then we write the symbol of $\mathcal{C}$ in $\mathbb{K}[\sigma_+, \sigma_-]/\mathcal{I}_\sigma$ using the substitutions $\delta x_+ \to \sigma x_+ - 2$, $\delta x_- \to 2 - \sigma x_-$, $\delta t_+ \to \sigma t_+ - 2$ and $\delta t_- \to 2 - \sigma t_-$. Then we determine the integrence order of R^σ using the lemma 2.2. We have 3 free parameters and hence we hope to able to annihilate zeroth and first order monomials. This turns out to be case, so $\mathcal{C}$ is an integrence factor of order 2.

In this case we really get a system of 3 equations, one of which is quadratic and others linear. Hence there are two solutions and the resulting operators are denoted as follows

$$\mathcal{C}_s^{(p,l)} = \mathcal{B}_s^p + \alpha\left(\tilde{a}^{(p,l)}_{(s+1,0)}\delta x_-^{s+1} + \tilde{a}^{(p,l)}_{(s+1,1)}\delta x_-^s\delta t_+ + \tilde{a}^{(p,l)}_{(s+1,2)}\delta x_-^{s-1}\delta t_+\delta t_-\right) \tag{3.4}$$

(n,p)	$a^p_{(n,0)}$	(n,p)	$a^p_{(n,0)}$
(3,2)	$(\alpha^2+8)/24$	(7,2)	$(\alpha^6-108\alpha^4+3048\alpha^2+3296)/27648$
(4,2)	$(\alpha^2+4)/16$	(5,3)	$(\alpha^4+140\alpha^2+384)/1920$
(5,2)	$(-\alpha^4+92\alpha^2+224)/1152$	(6,3)	$(\alpha^4+60\alpha^2+128)/768$
(6,2)	$(-5\alpha^4+220\alpha^2+352)/2304$	(7,3)	$(-\alpha^6+116\alpha^4+3656\alpha^2+6624)/46080$

Table 1: Computed values of $a^p_{(n,0)}$ for $3 \le n \le 7$ and $p = 2, 3$.

4 Computation of the boundary conditions

The explicit calculations of the boundary conditions (3.3) suggested the following

Conjecture 4.1 *If $p = 1$ then*

$$\begin{aligned} a^1_{(n,0)} &= \frac{1}{4^{n-1}} \begin{pmatrix} 2n-2 \\ n-1 \end{pmatrix} \\ a^1_{(n,1)} &= \frac{1}{2^{2n-3}} \begin{pmatrix} 2n-4 \\ n-2 \end{pmatrix} \\ a^1_{(n,k)} &= -\frac{1}{2^{2n-3}k} \begin{pmatrix} 2k-2 \\ k-1 \end{pmatrix} \begin{pmatrix} 2n-2k-2 \\ n-k-1 \end{pmatrix} \end{aligned}$$

In the last formula $1 < k < n$. In general we have

$$a^p_{(n+1,k+1)} = \varepsilon_k \frac{2k-1}{2k+2} a^p_{(n,k)}$$

where $\varepsilon_k = -1$ for $k = 0, 1$ and $\varepsilon_k = 1$ for $k > 1$.

We checked these formulas up to $n \le 7$ and $p \le 3$. Note that when $p > 1$ we should still guess the values $a^p_{(n,0)}$ to get completely explicit expressions for coefficients. The computed values are given in table 1. Note that in any case corollary 3.2 implies that when passing to a bigger value of p, the number of new coefficients to be calculated is rather small.

As far as the boundary conditions (3.4) are concerned we have the following

Conjecture 4.2 *When $p = 1$ we get the two solutions*

$$\begin{cases} \tilde{a}^{(1,1)}_{(s+1,0)} = -\dfrac{s}{2^s} \\ \tilde{a}^{(1,1)}_{(s+1,1)} = \dfrac{1}{4^{s-1}} \begin{pmatrix} 2s-2 \\ s-1 \end{pmatrix} - \dfrac{1}{2} - \dfrac{1}{2^{s-1}} \\ \tilde{a}^{(1,1)}_{(s+1,2)} = \dfrac{1}{2} - \dfrac{s}{2^s} \end{cases} \qquad \begin{cases} \tilde{a}^{(1,2)}_{(s+1,0)} = 0 \\ \tilde{a}^{(1,2)}_{(s+1,1)} = \dfrac{1}{4^{s-1}} \begin{pmatrix} 2s-2 \\ s-1 \end{pmatrix} - \dfrac{1}{2} \\ \tilde{a}^{(1,2)}_{(s+1,2)} = \dfrac{1}{2} \end{cases}$$

In general $\tilde{a}^{(p,l)}_{(s+1,0)}$ and $\tilde{a}^{(p,l)}_{(s+1,2)}$ are obtained from $\tilde{a}^{(p,l)}_{(s+1,1)}$ by formulas

$$\tilde{a}^{(p,l)}_{(s+1,i)} = \frac{s}{2} \tilde{a}^{(p,l)}_{(s+1,1)} + c(\alpha, s, i)$$

where $i = 0$ or 2. The coefficients $\tilde{a}^{(p,l)}_{(s+1,1)}$ are the two roots of the second degree polynomial $P(z;\alpha,p,s) = z^2 + c_1(\alpha,s)z + c_0(\alpha,p,s)$ where the coefficients c_j depend only on the parameters indicated. The roots of P are real for $0 \le \alpha \le 1$.

We checked the conjecture for $1 \le s \le 7$ and $p \le 3$, see [29] for the computed values of c, c_0 and c_1. Note that the coefficients $\tilde{a}^{(1,l)}_{(s+1,0)}$ in cases $l = 1$ and $l = 2$ approach each other for large s since

$$\frac{1}{4^{s-1}} \begin{pmatrix} 2s-2 \\ s-1 \end{pmatrix} = \frac{1}{\sqrt{\pi s}} + O(s^{-1})$$

5 Stability analysis

5.1 Preliminaries

We discretize the wave equation $u_{tt} - u_{xx} = 0$ in the interval $I = [-L, 0]$ (where $L > 0$) with the scheme given in (3.1) and use the Dirichlet condition $u(-L,t) = 0$ on the left boundary and the absorbing condition on the right. When $p > 1$ we need in fact p conditions, since interior operators get longer. We shall simply use the same absorbing operator p times. Let us define

$$u^n = \begin{pmatrix} u^n_0 \\ \vdots \\ u^n_{m-1} \end{pmatrix}$$

where u^n_i is an approximation to $u(-ih, nh_t)$. The whole scheme can then be written as

$$u^{n+1} = \sum_{i=1}^{r} A^i u^{n-i+1} \tag{5.1}$$

Note that only A^1 and A^2 depend on the scheme (3.1) and the form of the boundary condition implies

Lemma 5.1 *The number r in (5.1) is $\max\{2, s-1\}$ for operators $\mathcal{B}^p_s$ and $\mathcal{C}^{(p,l)}_s$.*

Further we define

$$U^n = \begin{pmatrix} u^{n-r+1} \\ \vdots \\ u^n \end{pmatrix} \qquad M = \begin{pmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I \\ A^r & A^{r-1} & A^{r-2} & \dots & A^1 \end{pmatrix}$$

Then the scheme (5.1) can simply be written as

$$U^{n+1} = MU^n \tag{5.2}$$

It will be convenient to consider M and A^i as functions of α, hence we will sometimes write $M = M(\alpha)$ and $A^i = A^i(\alpha)$. We shall denote the spectrum of M by $\sigma(M)$, the spectral radius by $\rho(M)$, the nullspace by $\mathcal{N}(M)$ and range by $\mathcal{R}(M)$.

Let us recall the

Definition 5.1 *The scheme (5.2) is (asymptotically) stable if there is some number $\hat{\alpha}$ such that $\rho(M(\alpha)) < 1$ for all $0 < \alpha \le \hat{\alpha}$. The supremum of such numbers $\hat{\alpha}$ is called the stability limit of the scheme.*

Since we are not going to use any other stability concepts, we shall drop the word asymptotically for simplicity. Naturally we are interested only in schemes which are stable for all m. Hence we define

Definition 5.2 *The boundary condition (3.3) or (3.4) is stable, if the resulting scheme (5.2) is stable for all sufficiently big m.*

Obviously some small values of m must be excluded because if the boundary condition is 'long' it may not fit in the domain for small m.

5.2 Asymptotic analysis

We shall analyse the stability of the scheme (5.2) for small α by considering the asymptotic series of the relevant eigenvalue. Let us start with some simple observations.

Lemma 5.2 *$\sigma(M(0)) = \{0,1\}$ and the algebraic multiplicity of the eigenvalue $\lambda = 1$ is $2m-p$.*

Proof. See [25]. □

By continuity of eigenvalues it is clear that the scheme (5.2) is stable for small α if the multiple eigenvalue $\lambda = 1$ bifurcates into the unit disk. For the subsequent analysis it is convenient to translate this interesting eigenvalue to the origin by defining $B := M - I$. Let us denote $B(0)$ by B_0. We will need explicitly the nullspaces of B_0 and B_0^t ; these are given in the following

Lemma 5.3 *Let us define the following vectors w^k, $\tilde{w}^k \in \mathbb{R}^{rm}$, $1 \le k \le m$*

$$w_i^k = \begin{cases} 1, & i = k + jm \\ 0, & \text{otherwise} \end{cases} \qquad \tilde{w}_i^k = \begin{cases} -1, & i = (r-2)m + k \text{ and } k > p \\ 1, & i = (r-1)m + k \\ 0, & \text{otherwise} \end{cases}$$

where $0 \le j \le r-1$. Then $\{w^k\}_{k=1}^m$ is an orthogonal basis for $\mathcal{N}(B_0)$ and $\{\tilde{w}^k\}_{k=1}^m$ is an orthogonal basis for $\mathcal{N}(B_0^t)$. In particular $\dim\left(\mathcal{N}(B_0)\right) = \dim\left(\mathcal{N}(B_0^t)\right) = m$. *Moreover*

$$\langle w^i, \tilde{w}^k \rangle = \begin{cases} 1, & 1 \le i = k \le p \\ 0, & \textit{otherwise} \end{cases}$$

where $\langle \cdot, \cdot \rangle$ is the inner product in $\mathbb{R}^{rm}$.

Proof. This is straightforward. □

From this it follows that zero is a defective eigenvalue for B_0. Some computations suggest that the Jordan form corresponding to zero eigenvalue has p simple one by one Jordan blocks and $m-p$ two by two blocks. This structural information is not needed in the sequel.

Let us then recall the familiar

Lemma 5.4 *$v \in \mathcal{R}(B_0)$ if and only if $\langle v, \tilde{w}^k \rangle = 0$ for all $1 \le k \le m$.*

Proof. See for example [17]. □

Let us denote the expansion of B by $B = B_0 + B_1\alpha + B_2\alpha^2 + \ldots$. We shall need the following

Lemma 5.5 *$\mathcal{R}(B_1) = \text{span}\{\tilde{w}^1, \ldots, \tilde{w}^p\}$ and moreover there exist some numbers d_j, $(j \ge 0)$ which depend only on the boundary operator such that*

$$B_1 w^i = \sum_{k=1}^{p} d_{i-k} \tilde{w}^k \tag{5.3}$$

where we use the convention $d_j = 0$ if $j < 0$.

Proof. Recall that we use the same boundary operator p times. □

Since these coefficients d_j will be important, let us compute them explicitly.

Lemma 5.6 *For a given boundary condition the corresponding numbers d_j of the equation (5.3) depend only on the coefficients $a^p_{(i,0)}(0)$ and $\tilde{a}^p_{(i,0)}(0)$.*

Proof. Intuitively the reason is that multiplying B_1 by w^k means 'summing at all time levels' ; hence the terms which contain time differences have no influence. For details, see [25]. □

Corollary 5.1 *Denoting the numbers d_j associated with $\mathcal{B}_s^p$ (resp. $\mathcal{C}_s^{(p,l)}$) by $d_j(s,p)$ (resp. $\tilde{d}_j(s,p,l)$) we have*

$$d_j(s,p) = (-1)^{j+1} \sum_{i=j}^{s} a^p_{(i,0)}(0) \begin{pmatrix} i \\ j \end{pmatrix}$$

$$\tilde{d}_j(s,p,l) = d_j(s,p) + (-1)^{j+1} \tilde{a}^{(p,l)}_{(s+1,0)}(0) \begin{pmatrix} s+1 \\ j \end{pmatrix}$$

In particular $d_j(s,p) = 0$ (resp. $\tilde{d}_j(s,p,l) = 0$) for $j > s$ (resp. $j > s+1$).

Proof. We merely compare the coefficients in the following formulas

$$\sum_{j=0}^{s} d_j(s,p)x^k = -\sum_{j=1}^{s} a^p_{(j,0)}(0)(1-x)^j$$
$$\sum_{j=0}^{s+1} \tilde{d}_j(s,p,l)x^j = -\tilde{a}^{(p,l)}_{(s+1,0)}(0)(1-x)^{s+1} - \sum_{j=1}^{s} a^p_{(j,0)}(0)(1-x)^j$$

□

The following lemmas are easy to check.

Lemma 5.7 *Let*

$$\hat{w}^k_i = \begin{cases} 2j-1, & i = k+jm \\ 0, & \text{otherwise} \end{cases}$$

Then $B_0\hat{w}^k = 2(w^k - \tilde{w}^k)$ *if* $1 \le k \le p$ *and* $B_0\hat{w}^k = 2w^k$ *for* $k > p$.

Lemma 5.8 *Let* $k > p$, $A^1(\alpha) = A^1_0 + A^1_1\alpha + A^1_2\alpha^2 + \ldots$ *and denote the* (i,j) *element of* A^1_2 *by* $A^1_2(i,j)$ *; then* $\langle B_2 w^i, \tilde{w}^k\rangle = A^1_2(k,i)$.

Now the stability of the scheme for the small α depends on how the zero eigenvalue of B_0 bifurcates into $2m-p$ distinct eigenvalues. To analyse this we need also eigenvectors, i.e. we have to analyse the solutions of the eigensystem equations $B(\alpha)v(\alpha) = \lambda(\alpha)v(\alpha)$ for small α. To this end we look for the asymptotic expansions of λ and v. Now since zero is a defective eigenvalue, it is known that in general the expansion for λ contains fractional powers of α and that the eigenvectors need not even be continuous, [17], [15]. Nevertheless, we have

Theorem 5.1 *The following asymptotic expansions are valid*

$$\lambda(\alpha) = \lambda_1\alpha + O(\alpha^2)$$
$$v(\alpha) = v_0 + v_1\alpha + O(\alpha^2)$$

Proof. We make an ansatz that the expansions are indeed of the above form, and then compute the initial coefficients. Now expanding $Bv = \lambda v$ we get

$$B_0v_0 + (B_1v_0 + B_0v_1)\alpha + (B_2v_0 + B_1v_1 + B_0v_2)\alpha^2 + \cdots = \lambda_1v_0\alpha + (\lambda_2v_0 + \lambda_1v_1)\alpha^2 + \ldots$$

Hence $v_0 \in \mathcal{N}(B_0)$ and by lemma 5.3 we have $v_0 = \sum c_i w^i$ for some constants c_i. The next power of α gives

$$B_0v_1 = (\lambda_1 I - B_1)v_0 \tag{5.4}$$

and thus $(\lambda_1 I - B_1)v_0 \in \mathcal{R}(B_0)$. Using lemma 5.4 we get

$$\langle(\lambda_1 I - B_1)v_0, \tilde{w}^k\rangle = \lambda_1\sum_{i=1}^{m} c_i\langle w^i, \tilde{w}^k\rangle - \sum_{i=1}^{m} c_i\langle B_1w^i, \tilde{w}^k\rangle = 0$$

Now using lemmas 5.3 and 5.5 we see that the above equation is identically satisfied if $k > p$ and for $1 \le k \le p$ one obtains

$$\lambda_1 c_k - \sum_{i=1}^{m} c_i d_{i-k} = 0 \qquad 1 \le k \le p \tag{5.5}$$

Then to compute v_1 we make an ansatz that it can be written as

$$v_1 = \hat{v} + \sum_{k=1}^{m} a_k\hat{w}^k$$

where $\hat{v}$ is some element in $\mathcal{N}(B_0)$ and $\hat{w}^k$ are as defined in lemma 5.7. Further, using lemma 5.7 and the condition (5.5) one gets

$$B_0v_1 = 2\sum_{k=1}^{m} a_kw^k - 2\sum_{k=1}^{p} a_k\tilde{w}^k = \lambda_1\Big(\sum_{k=1}^{m} c_kw^k - \sum_{k=1}^{p} c_k\tilde{w}^k\Big) = (\lambda_1 I - B_1)v_0$$

which implies that $a_k = \lambda_1 c_k/2$. We still have only p equations for λ_1 and c_k and consequently we have to use the second order terms which give

$$B_0 v_2 = (\lambda_2 I - B_2)v_0 + (\lambda_1 I - B_1)v_1$$

Now $\langle v_0, \tilde{w}^k \rangle = 0$ if $k > p$; hence the remaining equations are

$$\langle -B_2 v_0 + (\lambda_1 I - B_1)v_1, \tilde{w}^k \rangle = 0 \qquad p < k \leq m$$

Then combining $\langle v_1, \tilde{w}^k \rangle = 2a_k = \lambda_1 c_k$, $\langle B_1 v_1, \tilde{w}^k \rangle = 0$ and lemma 5.8 we get

$$\lambda_1^2 c_k = \sum_{i=1}^{m} c_i A_2^1(k,i), \quad p < k \leq m \tag{5.6}$$

Combined with (5.5) we have then a system of m equations which is linear and homogeneous with respect to c_i, hence there are non trivial solutions if and only if the corresponding determinant vanishes. But this determinant is a polynomial of degree $2m-p$ with respect to λ_1, which implies that in general the zero eigenvalue bifurcates into $2m-p$ distinct eigenvalues.

Since $a_k = \lambda_1 c_k/2$, these solutions also determine v_1 up to an element in $\mathcal{N}(B_0)$. To compute this element we should consider higher order terms in the expansion. However, this information is not needed in the sequel. □

This result obviously implies that the scheme (5.2) is stable for small α if and only if $\Re\lambda_1 < 0$. To analyze the situation more conveniently let us denote λ_1 by z and introduce $c = (c_1, \ldots, c_m)$. Moreover we define $D = \mathrm{diag}(z, \ldots, z, z^2, \ldots, z^2) \in \mathbb{R}^{m\times m}$ where we have z p times and z^2 $m-p$ times, and $E \in \mathbb{R}^{m\times m}$ where $e_{ij} = d_{j-i}$ for $1 \leq i \leq p$ and $e_{ij} = A_2^1(i,j)$ for $p < i \leq m$. Then the equations (5.5) and (5.6) can simply be written as $(D-E)c = 0$. To write explicitly the dependence on parameters we write $E = E(p,s)$ (resp. $E = E(p,s,l)$) if we use boundary condition $\mathcal{B}_s^p$ (resp. $\mathcal{C}_s^{(p,l)}$). Then we define

$$q(z;m,p,s) = \det\big(D - E(p,s)\big)$$
$$\tilde{q}(z;m,p,s,l) = \det\big(D - E(p,s,l)\big)$$

With these notations the stability criterion can be formulated as follows.

Proposition 5.1 *The boundary condition $\mathcal{B}_s^p$ (resp. $\mathcal{C}_s^{(p,l)}$) is stable if and only if the zeros of $q(z;m,p,s)$ (resp. $\tilde{q}(z;m,p,s,l)$) are in the left half-plane for all $m > s$.*

This implies

Corollary 5.2 *Suppose that $1 \leq s \leq 7$; then $\mathcal{B}_s^1$ is stable if and only if $\mathcal{C}_s^{(1,2)}$ is stable.*

Proof. Now $q(z;m,1,s) = \tilde{q}(z;m,1,s,2)$ because $\tilde{a}^{(1,2)}_{(s+1,0)} = 0$ for all $1 \leq s \leq 7$. □

5.3 Case $p = 1$

This is considerably simpler than the other cases because the relevant polynomials can be computed easily by the following recursion formula.

Lemma 5.9 *If $m > s$ then*

$$q(z;m+2,1,s) - (z^2+2)q(z;m+1,1,s) + q(z;m,1,s) = 0$$

The same is valid for $\tilde{q}$ also.

Proof. The matrix E has the following form

$$E = \begin{pmatrix} * & * & * & \ldots & 0 \\ 1 & -2 & 1 & \ldots & 0 \\ 0 & 1 & -2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 & -2 \end{pmatrix}$$

where $*$ denotes possibly non-zero terms. Hence $D-E$ is tridiagonal (except for the first row) and consequently using elementary properties of the determinant we get the stated recursion. Because of the first row the recursion may not be valid for small m. □

We shall compute explicitly the polynomials $q(z;m,1,s)$ in various cases. To this end we introduce

Definition 5.3 *Let $q^s_m(z;x^1,x^0)=\sum_{i=0}^{2m-1} c^s_{(i,m)} z^i$ be a family of polynomials whose coefficients are defined by*

$$\begin{cases} c^s_{(2k,m)} = \displaystyle\sum_{j=0}^{s-1} x^1_j \begin{pmatrix} m+k-j-1 \\ m-k-1 \end{pmatrix} \\ c^s_{(2k+1,m)} = \displaystyle\sum_{j=0}^{s-1} x^0_j \begin{pmatrix} m+k-j \\ m-k-1 \end{pmatrix} \end{cases}$$

Recall the convention $\begin{pmatrix} m \\ k \end{pmatrix} = 0$ if $m<0$ or $k>m$.

Lemma 5.10 *The polynomials defined above satisfy the recursion*

$$q^s_{m+2} - (z^2+2)q^s_{m+1} + q^s_m = 0$$

Proof. This is easy to check using the following binomial identity

$$\begin{pmatrix} m \\ k \end{pmatrix} = \begin{pmatrix} m-1 \\ k \end{pmatrix} + \begin{pmatrix} m-1 \\ k-1 \end{pmatrix} \tag{5.7}$$

For details, see [25]. □

Recall that by corollary 5.2 we need only consider the case $l=1$.

Lemma 5.11 *The polynomials $q(z;m,1,s)$ (resp. $\tilde{q}(z;m,1,s,1)$) are obtained from the family $q^s_m(z;x^1,x^0)$ by putting $x^0_0=1$, $x^0_j=0$ for $0<j<s$ and $x^1_j = a^1_{(j+1,0)}$ for $0\le j<s$ (resp. $x^1_j = a^1_{(j+1,0)}$ for $0\le j<s$ and $x^1_s = \tilde{a}^{(1,1)}_{(s+1,0)}$).*

Proof. By previous lemma we know that the polynomials indicated satisfy the right recursion formula. Hence it is sufficient to verify that given formulas verify also the right initial conditions, which is straightforward, though tedious. □

To prove the stability we use the Routh-Hurwitz criterion as it is usually given in the control engineering literature, see for instance [3]. A thorough discussion of Routh-Hurwitz criterion in general can be found in [18]. Let us consider the polynomial $P(z)=\sum_{i=0}^{2m-1} c_i z^i$ to which we associate the following array

$$\begin{array}{ccccc} c_{2m-1} & c_{2m-3} & \cdots & c_3 & c_1 \\ c_{2m-2} & c_{2m-4} & \cdots & c_2 & c_0 \\ c^1_{2m-3} & c^1_{2m-5} & \cdots & c^1_1 & \\ c^1_{2m-4} & c^1_{2m-6} & \cdots & c^1_0 & \\ \vdots & \vdots & & & \\ c^{m-1}_1 & & & & \\ c^{m-1}_0 & & & & \end{array}$$

where the coefficients c^1_i are calculated using the formulas

$$c^1_{2k+1} = \frac{c_{2m-2}c_{2k+1} - c_{2m-1}c_{2k}}{c_{2m-2}}$$
$$c^1_{2k} = \frac{c^1_{2m-3}c_{2k} - c_{2m-2}c^1_{2k-1}}{c^1_{2m-3}}$$

Then we proceed analogously until in the last two rows there is only one element. The Routh-Hurwitz criterion is then

Theorem 5.2 *All the elements in the first column of the above array are positive if and only if the polynomial* $P(z) = \sum_{i=0}^{2m-1} c_i z^i$ *is stable, i.e. all its roots have negative real parts.*

Now let us compute the Routh-Hurwitz array for the polynomials q_m^s. For notational simplicity we denote $c^s_{(i,m)}$ by c_i. Let us define

$$\begin{aligned} x_j^{\nu+1} &= \sum_{i=0}^{j} x_i^{\nu-1} - \frac{\sum_{l=0}^{s-1} x_l^{\nu-1}}{\sum_{l=0}^{s-1} x_l^{\nu}} \sum_{i=0}^{j} x_{i-1}^{\nu} \\ \kappa_\nu &= \sum_{l=0}^{t_\nu - 1} x_l^{\nu} \end{aligned} \tag{5.8}$$

where $t_\nu := \min\{s, 2m-\nu\}$. Then we have

Theorem 5.3 *The polynomial* q_m^s *is stable if and only if* $\kappa_\nu > 0$ *for* $0 \le \nu < 2m$.

Proof. We have to show that the numbers κ_ν are the first column of the Routh-Hurwitz array. It is easily checked that $\kappa_0 = c_{2m-1}$ and $\kappa_1 = c_{2m-2}$. Then by induction one can show that

$$c^i_{2k+1} = \sum_{j=0}^{s-1} x_j^{2i} \begin{pmatrix} m+k-j-i \\ m-k-i-1 \end{pmatrix}$$

$$c^i_{2k} = \sum_{j=0}^{s-1} x_j^{2i+1} \begin{pmatrix} m+k-j-i-1 \\ m-k-i-1 \end{pmatrix}$$

¿From this we see that $c^i_{2(m-i)-1} = \kappa_{2i}$ and $c^i_{2(m-i)-2} = \kappa_{2i+1}$. □

Note that the iteration (5.8) has the following property

$$x_{s-1}^{\nu+1} = \frac{\sum_{l=0}^{s-1} x_l^0}{\sum_{l=0}^{s-1} x_l^{\nu}} x_{s-1}^1 \tag{5.9}$$

Moreover we have

Lemma 5.12 *For all boundary conditions* $x_0^\nu = 1$ *for all* ν.

Proof. It is easily seen that $x_0^{2\nu} = x_0^0$ and $x_0^{2\nu+1} = x_0^1$ for all ν. On the other hand $x_0^0 = x_0^1 = 1$ for all boundary conditions. □

Then starting with the simplest boundary condition we get

Proposition 5.2 *The boundary conditions* $\mathcal{B}_1^1$ *and* $\mathcal{C}_1^{(1,2)}$ *are stable.*

Proof. By the above lemma $x_0^\nu = \kappa_\nu = 1$ for all ν. □

Proposition 5.3 *The boundary conditions* $\mathcal{B}_2^1$ *and* $\mathcal{C}_2^{(1,2)}$ *are stable.*

Proof. In this case we have $x_0^0 = 1$, $x_1^0 = 0$, $x_0^1 = 1$ and $x_1^1 = 1/2$. Hence by (5.9) and by lemma 5.12 we get

$$x_1^{\nu+1} = \frac{1}{2 + 2x_1^\nu}$$

It is obvious that x_1^ν stays positive for all $\nu > 0$ which implies that $\kappa_\nu > 0$. □

Proposition 5.4 *The boundary conditions* $\mathcal{B}_3^1$ *and* $\mathcal{C}_3^{(1,2)}$ *are stable.*

Proof. In this case we have $x_0^0 = 1$, $x_1^0 = x_2^0 = 0$, $x_0^1 = 1$, $x_1^1 = 1/2$ and $x_2^1 = 3/8$. Defining $y_\nu := 1 + x_1^\nu + x_2^\nu$ we get

$$y_{\nu+1} = 1 + y_{\nu-1}\Big(1 - \frac{1}{y_\nu}\Big) + \frac{3}{8}\Big(\frac{1}{y_\nu} - \frac{1}{y_{\nu-2}}\Big)$$

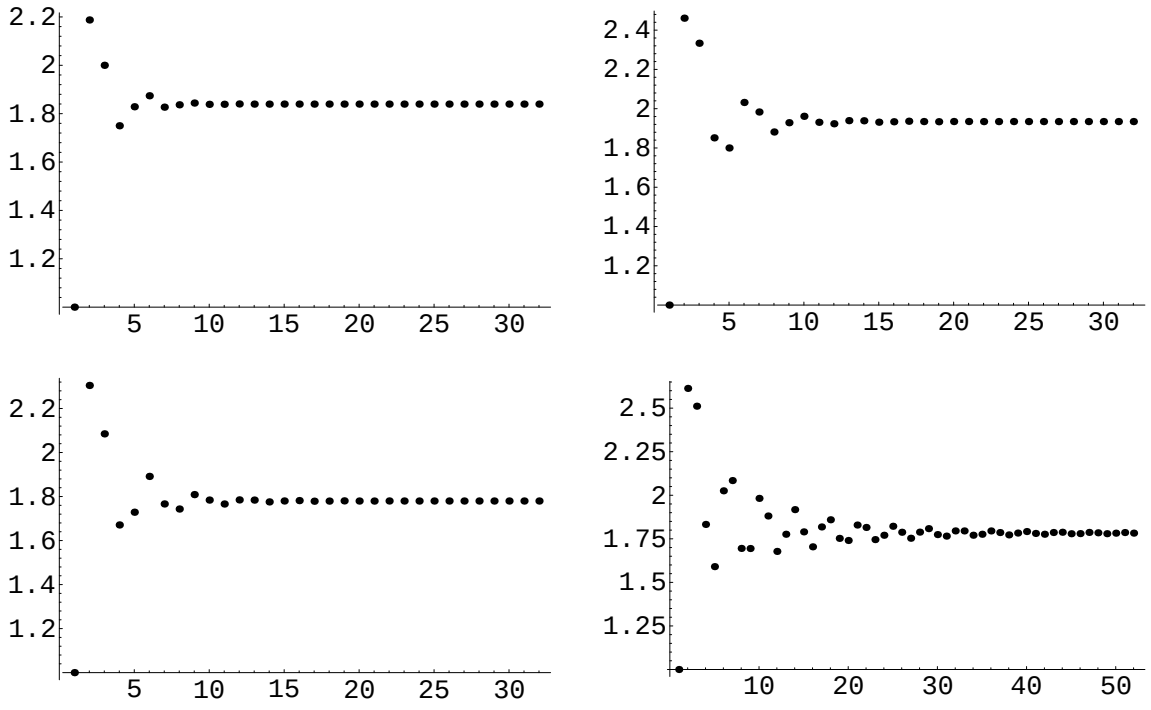

Figure 1: The iteration of y_ν for $\mathcal{B}_4^1$, $\mathcal{B}_5^1$, $\mathcal{C}_5^{(1,1)}$ and $\mathcal{C}_6^{(1,1)}$.

Since $x_2^\nu = 3/(8y_\nu)$, it is sufficient to prove that $y_\nu \geq 1$ for all ν. This is the case for $y_0 = 1$, $y_1 = 15/8$ and $y_2 = 5/3$. Let us consider the function

$$f(a,b,c) = 1 + b\Big(1 - \frac{1}{c}\Big) + \frac{3}{8}\Big(\frac{1}{c} - \frac{1}{a}\Big)$$

We have to prove that $\min_{a,b,c\geq 1} f \geq 1$. Calculating the gradient we find that $\nabla f = \big(3/(8a^2), 1 - 1/c, (8b - 3)/(8c^2)\big)$. All the components of the gradient are positive in the region of interest. Hence the global minimum is $f(1,1,1) = 1$. □

Conjecture 5.1 *The boundary conditions $\mathcal{B}_4^1$, $\mathcal{C}_4^{(1,2)}$, $\mathcal{B}_5^1$, $\mathcal{C}_5^{(1,2)}$, $\mathcal{C}_5^{(1,1)}$ and $\mathcal{C}_6^{(1,1)}$ are stable.*

Proof.(Numerical evidence) Similarly as above we have calculated the corresponding sequences y_ν for these boundary conditions, and the results are given in figure 1. Also it was observed that all x_j^ν seemed to converge to such values which would imply stability. □

Proposition 5.5 *The boundary conditions $\mathcal{B}_6^1$, $\mathcal{C}_6^{(1,2)}$, $\mathcal{B}_7^1$, $\mathcal{C}_7^{(1,2)}$, $\mathcal{C}_1^{(1,1)}$, $\mathcal{C}_2^{(1,1)}$, $\mathcal{C}_3^{(1,1)}$, $\mathcal{C}_4^{(1,1)}$ and $\mathcal{C}_7^{(1,1)}$ are unstable.*

Proof. It is sufficient to compute the Routh-Hurwitz array for the polynomials $q(z;8,1,6)$, $q(z;8,1,7)$, $\tilde{q}(z;3,1,1,1)$, $\tilde{q}(z;4,1,2,1)$, $\tilde{q}(z;5,1,3,1)$, $\tilde{q}(z;8,1,4,1)$ and $\tilde{q}(z;9,1,7,1)$. In all cases it was found that the negative element in the first column appeared in the initial part which is independent of m. Hence these boundary conditions cannot be stable for any m. □

5.4 Cases $p = 2$ and $p = 3$

Here the polynomials do not satisfy any simple recursion formula and we have not been able to prove any rigorous results. However, numerical computations suggest that

Conjecture 5.2 *The boundary conditions* $\mathcal{B}_7^p$, $\mathcal{C}_7^{(p,2)}$, $\mathcal{C}_1^{(p,1)}$, $\mathcal{C}_2^{(p,1)}$, $\mathcal{C}_3^{(p,1)}$, $\mathcal{C}_4^{(p,1)}$ *and* $\mathcal{C}_7^{(p,1)}$ *are unstable for* $p = 2$ *and* $p = 3$. *The other conditions are stable.*

Proof.(Numerical evidence) We computed the relevant polynomials in all cases for $m = 12$ and in the cases given above there were some zeros with positive real parts. We tested also some cases for other values of m, and the results were the same. Note that here we cannot directly conclude that the boundary conditions will stay unstable also for $m > 12$, since a priori there is no reason why the initial part of the Routh-Hurwitz array should be independent of m. However, numerical calculations suggest that this is indeed the case, which would then imply instability for all m. As an illustration it seems that

$$q(z; m, 2, 2) = z^{2m-2} + 3z^{2m-3} + \frac{10m-11}{4} z^{2m-4} + \frac{30m-71}{4} z^{2m-5} + \dots$$

This would imply that the third element of the first column of the Routh-Hurwitz array is 19/6 for any m. There are similar relations also in other cases. □

Comparing to the case $p = 1$ we see that the conditions $\mathcal{B}_6^p$ and $\mathcal{C}_6^{(p,2)}$ which were unstable for $p = 1$ seem to be stable here, and more generally it appears that $\mathcal{B}_s^p$ is stable if and only if $\mathcal{C}_s^{(p,2)}$ is stable.

6 Numerical results

We present below a representative sample of the extensive numerical experiments which are given fully in [29].

6.1 Smooth signal

We take a Gaussian (see figure 1) such that the signal arriving at the boundary has the amplitude one. We represent the reflected signal due to the absorbing boundary condition. We consider successively $\mathcal{B}_s^p$, $\mathcal{C}_s^{(p,1)}$ and $\mathcal{C}_s^{(p,2)}$ for p from 1 to 3 and for s from 1 to 6. The two main parameters of our experiments are :

- $\alpha = 0.5$: this does not indicate the stability limit of the scheme, we have merely taken the same value of α in all cases to ease the comparisons between the conditions, and this value is small enough to avoid the special effects which occur when α is nearly equal to 1. See below for a discussion about stability limits, in particular table 8.
- N the number of points per wavelength (or rather signal width). For all values of p the results are visually very similar, so we give here only the case $p = 2$ (for other cases, see [29]).

In the first series of figures (figures 2 to 4) we have taken N equal to 20. In the second series (figures 5 to 7) N is taken equal to 7 ; hence the signal is not numerically very smooth, and one can see that the reflections are more important in this case. Curiously all reflected signals are similar when $s \leq 2$, but when $s \geq 3$, $\mathcal{C}_s^{(p,1)}$ produces a signal which a mirror image of the others with respect to horizontal axis. Note also that if the initial signal has just one peak then the reflected signal has $s + 1$ peaks. This kind of phenomenon was also observed in [14]: increasing the order of the absorbing layer produced reflections with more peaks. However, the order in that article was totally different from order which is used here.

We have computed the ratio between the norm of the reflected signal and the norm of the incident signal in L^2 norm in tables 2 and 4, and in energy norm in tables 3 and 5. We also computed the ratio in L^∞ norm, but the results in this case were nearly the same as in L^2 case. Curiously the absorption for the boundary conditions $\mathcal{C}_3^{(2,1)}$ and $\mathcal{C}_3^{(3,1)}$ is 'unreasonably' good, see tables 4 and 5.

6.2 Fourier analysis

We represent in figure 8 the spectra of the reflected signals obtained for the 3 conditions of same difference order p – $\mathcal{B}_s^p$, $\mathcal{C}_s^{(p,1)}$, $\mathcal{C}_s^{(p,2)}$ – in various values of s and p. The initial signal is shown in figure 1. The solid line corresponds to $\mathcal{B}_s^p$, short dashes to $\mathcal{C}_s^{(p,1)}$ and long dashes to $\mathcal{C}_s^{(p,2)}$. One can see that the conditions of integrence order 2 are more efficient – when they are – mainly in high frequency domain. Concerning low orders, there is only one of them better than the "pure difference" condition, but for high orders both of them are always better.

s	1	2	3	4	5	6
$\mathcal{B}_s^1$	1.69	0.334	0.0974	0.0304	0.0111	0.00448
$\mathcal{C}_s^{(1,1)}$	2.53	0.147	0.0121	0.00486	0.00413	0.00253
$\mathcal{C}_s^{(1,2)}$	3.39	0.690	0.166	0.0351	0.00626	0.000306
$\mathcal{B}_s^2$	1.67	0.281	0.0771	0.0219	0.00726	0.00275
$\mathcal{C}_s^{(2,1)}$	2.51	0.0987	0.0174	0.0154	0.00822	0.00429
$\mathcal{C}_s^{(2,2)}$	3.35	0.632	0.150	0.0278	0.00306	0.00146
$\mathcal{B}_s^3$	1.65	0.276	0.0748	0.0216	0.00739	2.41
$\mathcal{C}_s^{(3,1)}$	2.48	0.0964	0.0172	0.0144	0.00765	0.00373
$\mathcal{C}_s^{(3,2)}$	3.32	0.621	0.146	0.0271	0.00351	0.000840

Table 2: Ratio reflected/incident signal in % in L^2 norm, $N = 20$.

s	1	2	3	4	5	6
$\mathcal{B}_s^1$	2.87	0.722	0.249	0.0880	0.0354	0.0155
$\mathcal{C}_s^{(1,1)}$	4.30	0.321	0.0319	0.0137	0.0130	0.00867
$\mathcal{C}_s^{(1,2)}$	5.77	1.47	0.420	0.101	0.0200	0.00114
$\mathcal{B}_s^2$	2.84	0.608	0.197	0.0636	0.0232	0.00933
$\mathcal{C}_s^{(2,1)}$	4.26	0.217	0.0435	0.0440	0.0260	0.0146
$\mathcal{C}_s^{(2,2)}$	5.71	1.35	0.381	0.0803	0.00989	0.00482
$\mathcal{B}_s^3$	2.81	0.595	0.191	0.0625	0.0236	8.16
$\mathcal{C}_s^{(3,1)}$	4.22	0.212	0.0431	0.0412	0.0243	0.0128
$\mathcal{C}_s^{(3,2)}$	5.65	1.32	0.370	0.0783	0.0112	0.00296

Table 3: Ratio reflected/incident smooth signal in % in energy norm, $N = 20$

s	1	2	3	4	5	6
$\mathcal{B}_s^1$	5.28	2.70	2.10	1.89	2.09	2.92
$\mathcal{C}_s^{(1,1)}$	7.88	1.40	0.464	0.238	0.553	1.11
$\mathcal{C}_s^{(1,2)}$	11.0	4.93	2.92	1.76	0.981	0.277
$\mathcal{B}_s^2$	5.07	2.20	1.61	1.32	1.27	1.45
$\mathcal{C}_s^{(2,1)}$	7.68	1.01	0.214	0.652	1.06	1.63
$\mathcal{C}_s^{(2,2)}$	10.6	4.39	2.55	1.37	0.569	0.596
$\mathcal{B}_s^3$	4.92	2.07	1.48	1.20	1.16	1.36
$\mathcal{C}_s^{(3,1)}$	7.46	0.928	0.228	0.619	0.988	1.43
$\mathcal{C}_s^{(3,2)}$	10.2	4.17	2.37	1.24	0.516	0.430

Table 4: Ratio reflected/incident signal in % in L^2 norm, $N = 7$

s	1	2	3	4	5	6
$\mathcal{B}_s^1$	9.12	5.56	4.86	4.79	5.71	8.67
$\mathcal{C}_s^{(1,1)}$	13.5	3.05	1.22	0.656	1.46	3.21
$\mathcal{C}_s^{(1,2)}$	19.4	9.88	6.45	4.25	2.57	0.791
$\mathcal{B}_s^2$	8.73	4.57	3.79	3.43	3.60	4.48
$\mathcal{C}_s^{(2,1)}$	13.3	2.27	0.422	1.57	2.86	4.83
$\mathcal{C}_s^{(2,2)}$	18.7	8.86	5.70	3.38	1.57	1.73
$\mathcal{B}_s^3$	8.45	4.28	3.48	3.13	3.32	4.25
$\mathcal{C}_s^{(3,1)}$	12.9	2.09	0.461	1.52	2.73	4.34
$\mathcal{C}_s^{(3,2)}$	18.1	8.40	5.31	3.08	1.42	1.29

Table 5: Ratio reflected/incident smooth signal in % in energy norm, $N = 7$

s	1	2	3	4	5	6
$\mathcal{B}_s^1$	12.3	8.50	8.04	8.52	10.8	18.3
$\mathcal{C}_s^{(1,1)}$	18.1	5.0	2.38	1.51	2.76	6.31
$\mathcal{C}_s^{(1,2)}$	27.4	14.9	10.2	7.10	4.46	1.57
$\mathcal{B}_s^2$	11.9	7.15	6.53	6.50	7.52	10.9
$\mathcal{C}_s^{(2,1)}$	18.0	3.92	0.704	2.51	5.32	10.7
$\mathcal{C}_s^{(2,2)}$	26.7	13.6	9.28	5.93	3.01	3.42
$\mathcal{B}_s^3$	11.7	6.83	6.17	6.17	7.29	11.2
$\mathcal{C}_s^{(3,1)}$	17.8	3.72	0.671	2.60	5.51	10.7
$\mathcal{C}_s^{(3,2)}$	26.2	13.1	8.87	5.59	2.84	2.88

Table 6: Ratio reflected/incident noisy signal in % in L^2 norm, $N = 15$.

s	1	2	3	4	5	6
$\mathcal{B}_s^1$	15.4	11.4	11.4	12.6	16.6	29.0
$\mathcal{C}_s^{(1,1)}$	22.3	7.09	3.83	2.57	4.15	9.83
$\mathcal{C}_s^{(1,2)}$	35.2	19.7	13.9	10.1	6.60	2.45
$\mathcal{B}_s^2$	14.8	9.68	9.37	9.87	11.9	18.1
$\mathcal{C}_s^{(2,1)}$	22.4	5.69	1.02	3.37	8.02	17.3
$\mathcal{C}_s^{(2,2)}$	34.7	18.2	12.8	8.57	4.63	5.35
$\mathcal{B}_s^3$	14.5	9.24	8.86	9.40	11.6	18.5
$\mathcal{C}_s^{(3,1)}$	22.2	5.42	0.819	3.61	8.49	17.4
$\mathcal{C}_s^{(3,2)}$	34.0	17.6	12.3	8.10	4.38	4.63

Table 7: Ratio reflected/incident noisy signal in % in energy norm, $N = 15$.

6.3 Noisy signal

The aim of this third series of experiments is to compare the behavior of conditions $\mathcal{B}_s^p$ and conditions $\mathcal{C}_s^{p,l}$ when the incident signal (see figure 1) contains a lot of high frequency components. We show the case $p = 2$ in figures 9 to 11. For other cases, see [29]; the case $p = 1$ is here rather different than others because of the effect of dispersion due to low order interior scheme. We can hope to observe better results with conditions $\mathcal{C}_s^{p,l}$ because of their integrence order. We still work with $\alpha = 0.5$ but this time $N = 15$. It is not so easy to test several discretisations by reducing N because of the numerical dispersion which distorts the signal, in particular when $p = 1$. We have again computed the ratio between the norm of the reflected signal and the norm of the incident signal in tables 6 and 7. It is seen again that the results for $\mathcal{C}_3^{(2,1)}$ and $\mathcal{C}_3^{(3,1)}$ are surprisingly good.

All these results show that taking higher order interior scheme neither diminishes the absorption nor destroys the stability of the whole scheme. Hence at least in the one-dimensional case the high order schemes are as easy to work with as the low order ones.

s	$\hat{\alpha}_{(s,1)}$	$\hat{\alpha}^1_{(s,1)}$	$\hat{\alpha}^2_{(s,1)}$	$\hat{\alpha}_{(s,2)}$	$\hat{\alpha}^1_{(s,2)}$	$\hat{\alpha}^2_{(s,2)}$	$\hat{\alpha}_{(s,3)}$	$\hat{\alpha}^1_{(s,3)}$	$\hat{\alpha}^2_{(s,3)}$
1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1
3	0.95	1	1	0.96	1	1	0.92	1	1
4	0.81	1	1	0.80	1	1	0.75	1	1
5	0.62	1	1	0.61	1	1	0.56	1	1
6	0.44	0.63	1	0.44	0.61	0.77	0.40	0.60	0.77

Table 8: Approximate stability limits.

6.4 On stability limit

We have not been able to compute explicitly the stability limits in various cases. However, the numerical computations suggest the approximate results given in table 8. We use the notation $\hat{\alpha}_{(s,p)}$ (resp. $\hat{\alpha}^l_{(s,p)}$) to denote the stability limit for the boundary condition (3.3) (resp. (3.4)).

Now we have proved above that quite many of these boundary conditions are unstable for small α. Why isn't this instability observed numerically? One reason could be that although the condition is unstable for small α it might still be stable when α belongs to some interval. To explore this possibility, we computed explicitly in some cases the spectrum of M in (5.2). Evidently, the size of the domain had to be taken quite small, but qualitatively the results did not seem to depend on the size. In the following computations the number points in the domain was taken to be 20.

Now let us first take the boundary condition $\mathcal{B}^1_6$. In figure 12 we show the spectral radius as function of α, and this plot indicates that their is a non-trivial stability interval. For values $\alpha > 0.4$ the spectral radius grows rather rapidly so the upper bound is approximately the same that was observed numerically. Note that for small α the spectral radius is extremely close to one.

However, similar plot for the boundary condition $\mathcal{C}^{(1,1)}_1$ indicates that the corresponding spectral radius is *always* bigger than one. Why was then no instability observed? Now plotting the whole spectrum in the complex plane, see figure 13, shows that the spectrum follows very closely the unit circle, and moreover computing the unstable eigenvectors shows that they tend to be very oscillating, so when using rather smooth initial signal and moderate simulation times, the instability has no time to grow to the observable level. The observed limits in table 8 correspond to the event that one eigenvalue finally starts to grow rather rapidly and leaves the immediate neighborhood of the unit circle.

This weak instability is rather curious, but we suspect that it is purely one-dimensional phenomenon and could not happen in many dimensional case.

7 Conclusion

As we saw above the results of the numerical tests are quite promising. Evidently these kind of boundary conditions will be interesting in practice only if the method for deriving them can be conveniently extended to many dimensional case. Conceptually this extension is straightforward, but its actual implementation is not. The problem is that one must necessarily mix different kinds of orders ; for example it is impossible to have factors of (difference) order one in all space variables. So instead of orders, it is probably best to use directly the language of ideals, and to improve the factorization means then that the residual belongs to a smaller ideal. These kind of computations are feasible using Gröbner basis techniques [1].

Another problem, perhaps more serious, is that we have not established any direct link between algebraic properties of the residual and actual numerical error. Therefore we face the task of choosing reasonable ideals, although the one dimensional case will guide us in this choice. Finally the stability question remains largely open, i.e. how to characterize stability algebraically? These problems will be treated in future papers.

References

[1] T. Becker and V. Weispfenning, *Gröbner bases : a computational approach to commutative algebra*, Graduate texts in mathematics, vol. 141, Springer, 1993.

[2] B. Char, K. Geddes, G. Bonnet, B. Leong, M. Monagan, and S. Watt, *Maple V language refence manual*, Springer, 1991.

[3] R. Dorf, *Modern control systems*, Addison-Wesley, 1967.

[4] B. Engquist and A. Majda, *Absorbing boundary conditions for the numerical simulation of waves*, Math. Comp. **31** (1977), 629–651.

[5] B. Engquist and A. Majda, *Radiation boundary conditions for acoustic and elastic wave calculations*, Comm. Pure Appl. Math. **32** (1979), 313–357.

[6] L. Frank, *Factorization for difference operators*, J. Math. Anal. Appl. **62** (1978), 170–185.

[7] L. Frank, *Toeplitz matrices and the theory of one-parameter families of difference operators*, Asympt. Anal. **3** (1991), 291–300.

[8] B. Gustafsson, H-O. Kreiss, and A. Sundström, *Stability theory of difference approximations for mixed initial boundary value problems II*, Math. Comp. **26** (1972), 649–686.

[9] L. Halpern, *Absorbing boundary conditions for discretization schemes of the one-dimensional wave equation*, Math. Comp. **38** (1982), 415–430.

[10] R. Higdon, *Absorbing boundary conditions for difference approximations to the multi-dimensional wave equation*, Math. Comp. **47** (1986), 437–460.

[11] R. Higdon, *Initial boundary value problems for linear hyperbolic systems*, SIAM Rev. **28** (1986), 177–217.

[12] R. Higdon, *Numerical absorbing boundary conditions for the wave equation*, Math. Comp. **49** (1987), 65–90.

[13] R. Jenks and R. Sutor, *Axiom, the scientific computation system*, Springer, 1992.

[14] P. Joly and J. Tuomela, *A new theoretical approach to absorbing layers*, to appear in SIAM J. Numer. Anal.

[15] T. Kato, *Perturbation theory for linear operators*, Grundlehren der Math. Wiss., vol. 132, Springer, 1966.

[16] H-O. Kreiss, *Initial boundary value problems for hyperbolic systems*, Comm. Pure Appl. Math. **23** (1970), 277–298.

[17] P. Lancaster and M. Tismenetsky, *The theory of matrices, 2nd ed.*, Acad. press, 1985.

[18] M. Marden, *Geometry of polynomials, 2nd ed*, Mathematical surveys and monographs, vol. 3, Amer. Math. Soc., 1966.

[19] L. Nirenberg, *Lectures on linear partial differential equations*, Reg. conf. ser. math., vol. 17, Amer. math. soc., Providence, 1972.

[20] R. Renaut-Williamson, *Full discretizations of $u_{tt} = u_{xx}$ and the rational approximations to* $\cosh(\mu z)$, SIAM J. Numer. Anal. **26** (1989), 338–347.

[21] M. Taylor, *Pseudodifferential operators*, Princeton mathematical series, vol. 34, Princeton university press, 1981.

[22] L. Trefethen, *Instability of difference models for hyperbolic initial boundary value problems*, Comm. Pure Appl. Math. **37** (1984), 329–367.

[23] L. Trefethen and L. Halpern, *Well-posedness of one-way wave equation and absorbing boundary conditions*, Math. Comp. **47** (1986), 421–435.

[24] J. Tuomela, *Algebraic approach to absorbing boundary conditions 1 : Basic ideas*, Research Report A335, Helsinki University of Technology, 1994.

[25] J. Tuomela, *Algebraic approach to absorbing boundary conditions 3 : On stability*, Research Report A342, Helsinki University of Technology, 1995.

[26] J. Tuomela, *Discrete absorbing boundary conditions for one dimensional wave equation*, Third international conference on mathematical and numerical aspects of wave propagation (E. Bécache, G. Cohen, P. Joly, and J. Roberts, eds.), SIAM, 1995, pp. 483–488.

[27] J. Tuomela, *A note on high order schemes for the one dimensional wave equation*, BIT **35** (1995), 394–405.

[28] J. Tuomela, *On the construction of arbitrary order schemes for many dimensional wave equation*, to appear in BIT.

[29] J. Tuomela and O. Vacus, *Algebraic approach to absorbing boundary conditions 2 : Numerical results*, Research Report A3??, Helsinki University of Technology, 1995.

[30] S. Wolfram, *Mathematica : a system for doing mathematics by computer, 2nd ed.*, Addison-Wesley, 1991.

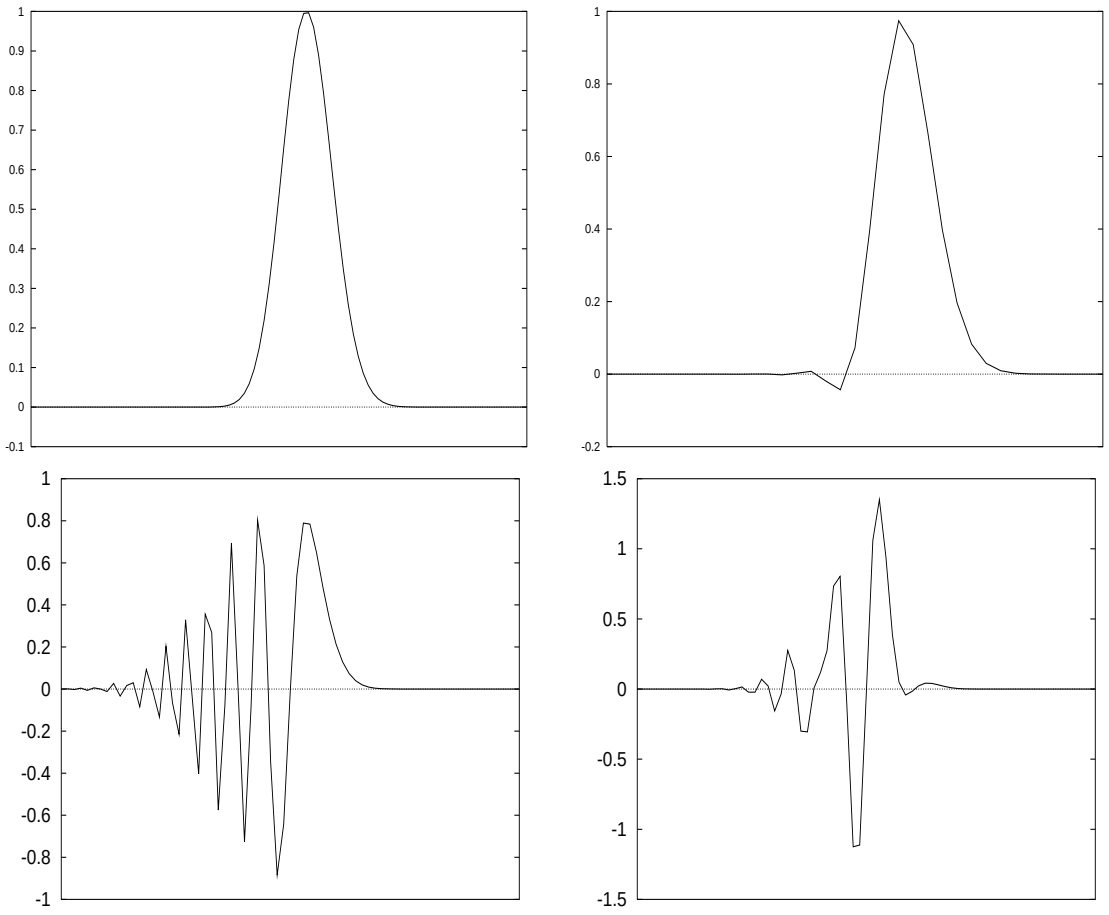

Figure 1: Smooth incident signal, $N = 20$ and $N = 7$; incident signal for spectral analysis and noisy incident signal, $N = 15$.

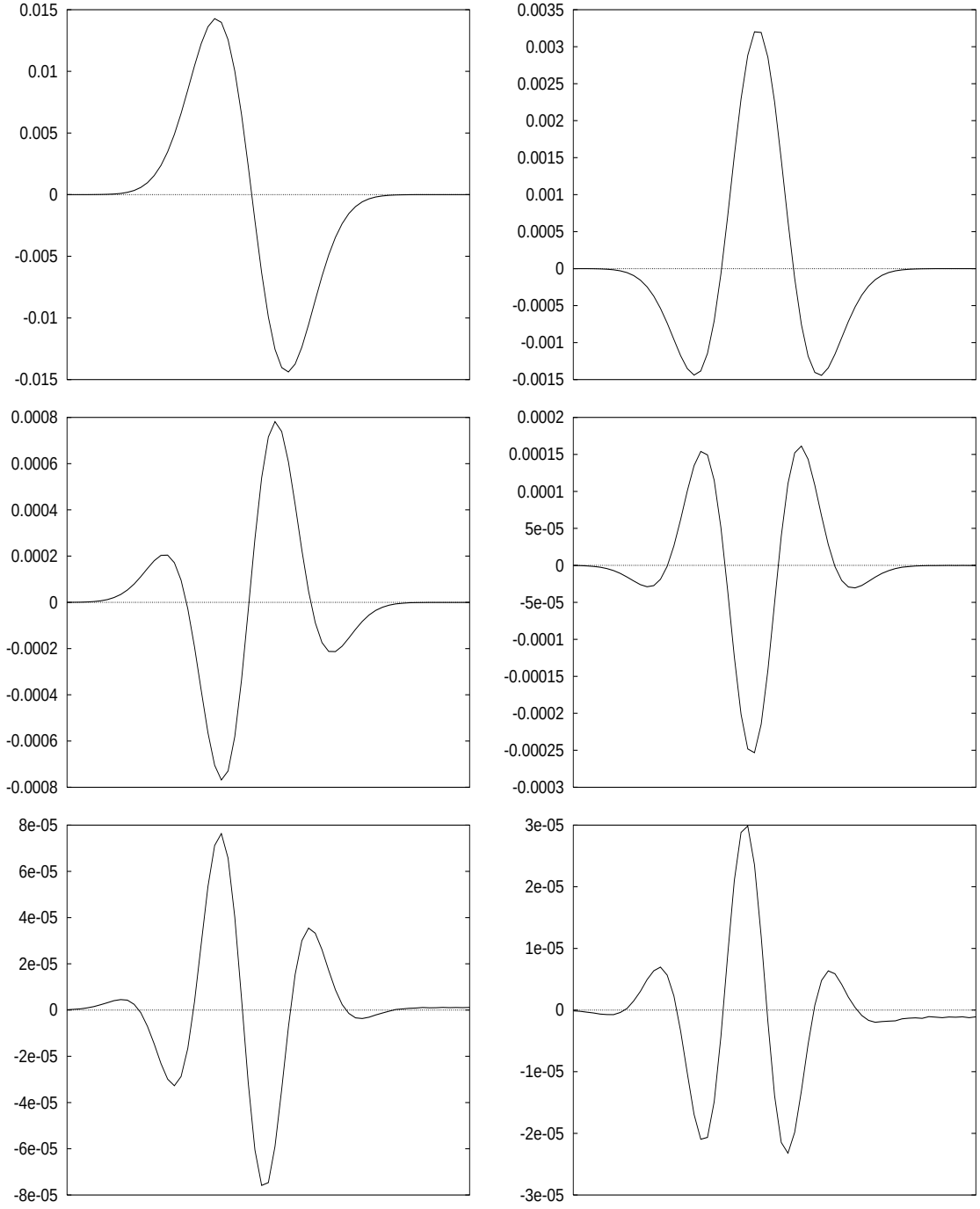

Figure 2: $\mathcal{B}_s^2$ for $s = 1$ to 6 ($N = 20$)

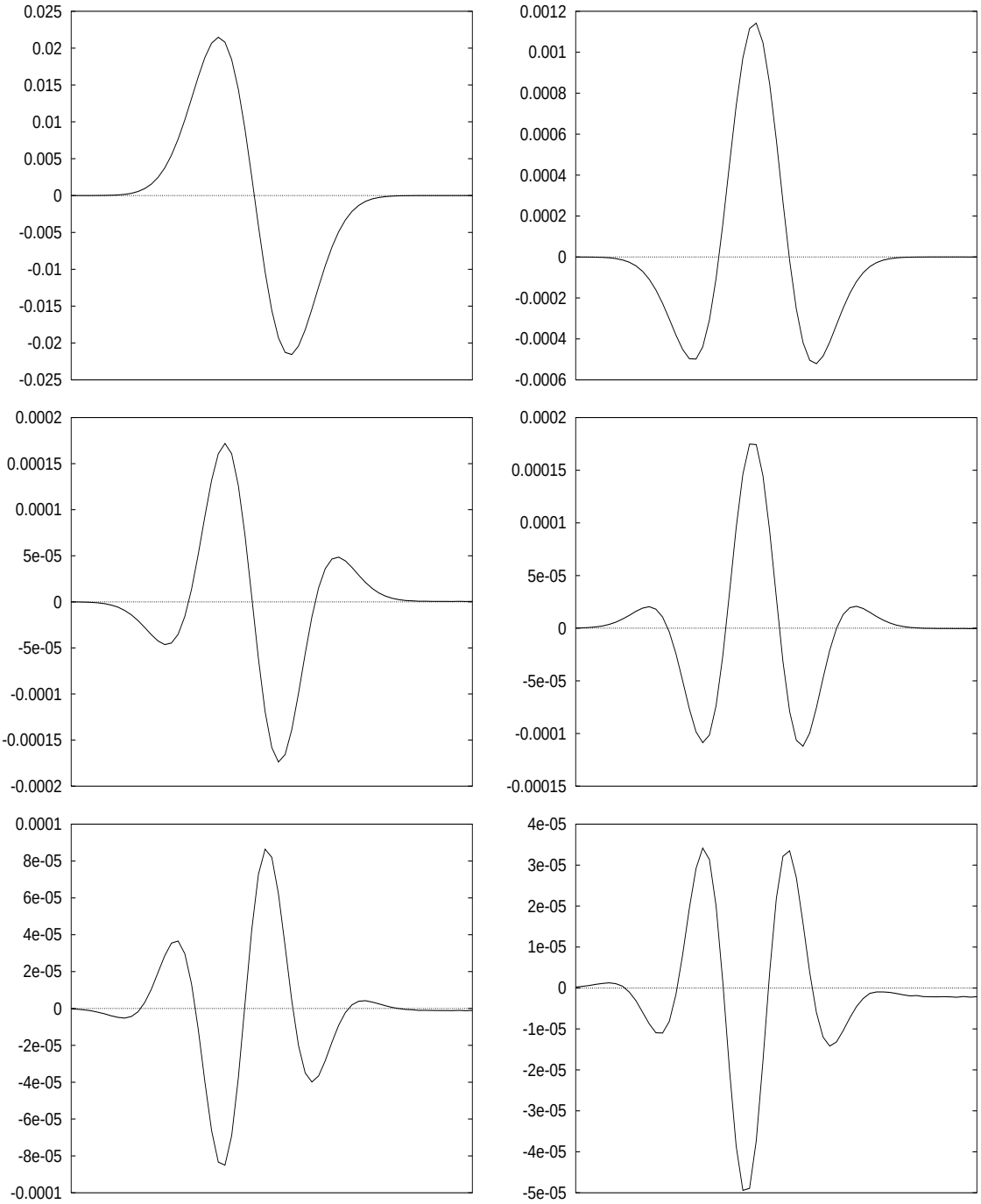

Figure 3: $\mathcal{C}_s^{(2,1)}$ for $s = 1$ to 6 ($N = 20$)

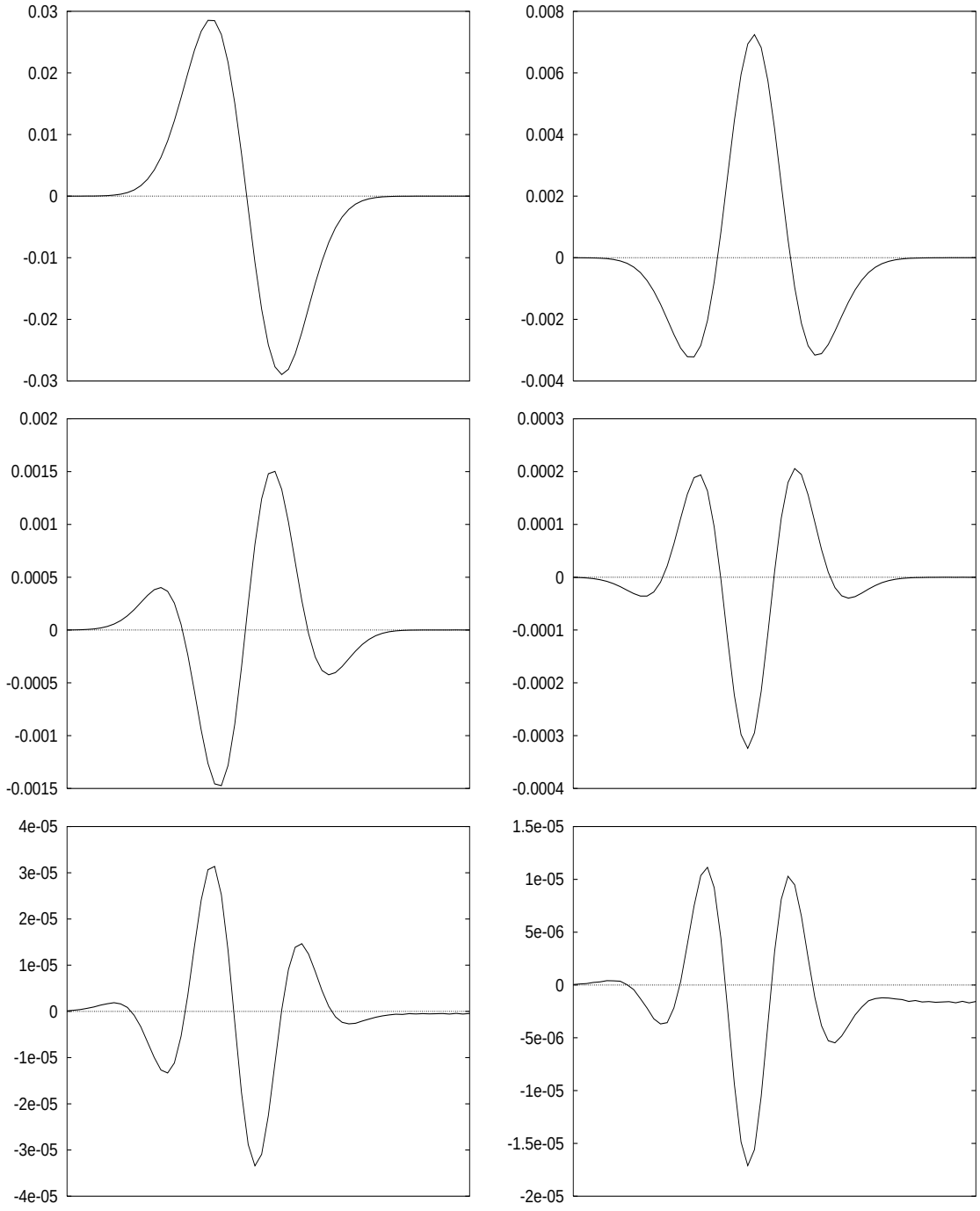

Figure 4: $\mathcal{C}_s^{(2,2)}$ for $s = 1$ to 6 ($N = 20$)

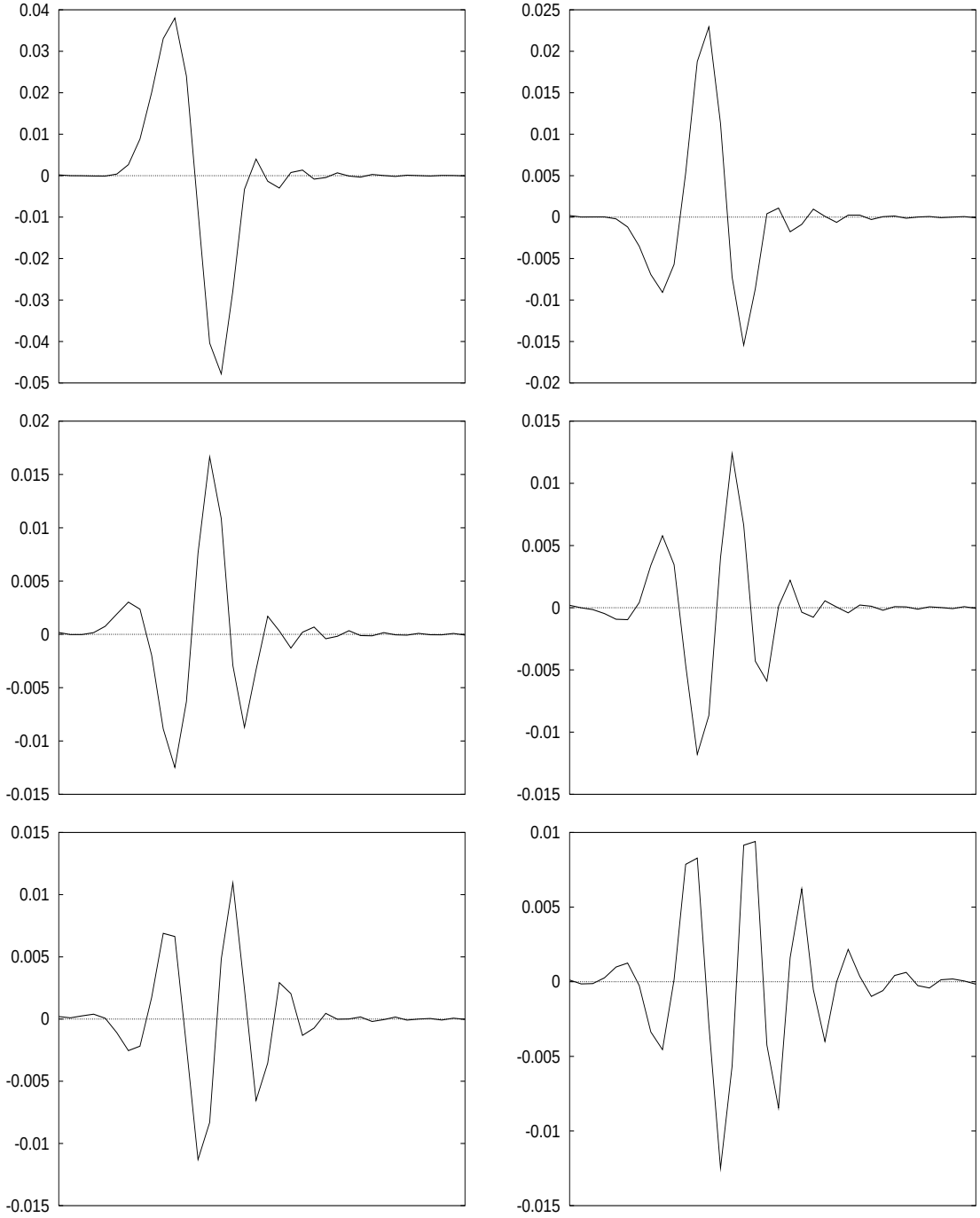

Figure 5: $\mathcal{B}_s^2$ for $s = 1$ to 6 ($N = 7$)

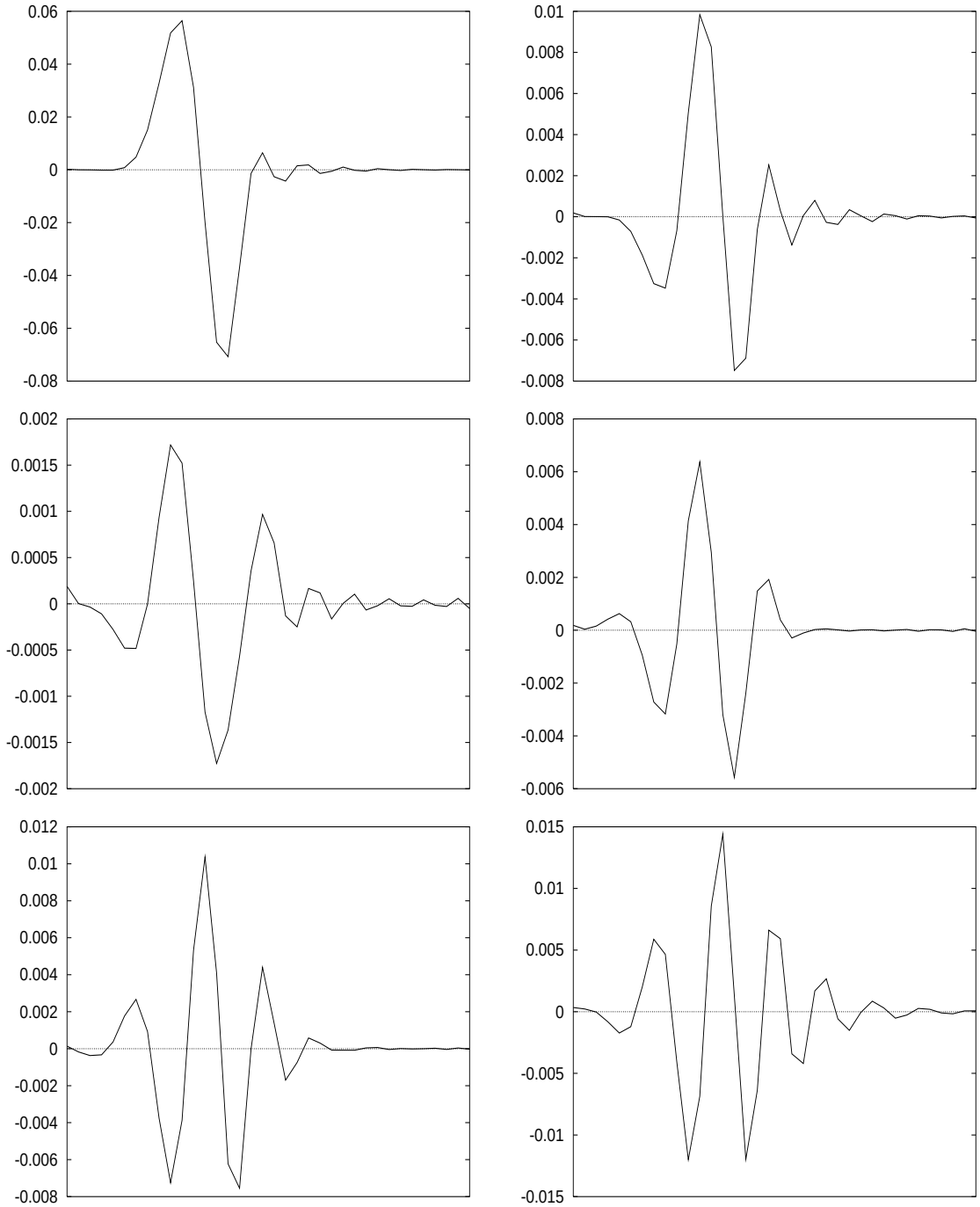

Figure 6: $\mathcal{C}_s^{(2,1)}$ for $s = 1$ to 6 ($N = 7$)

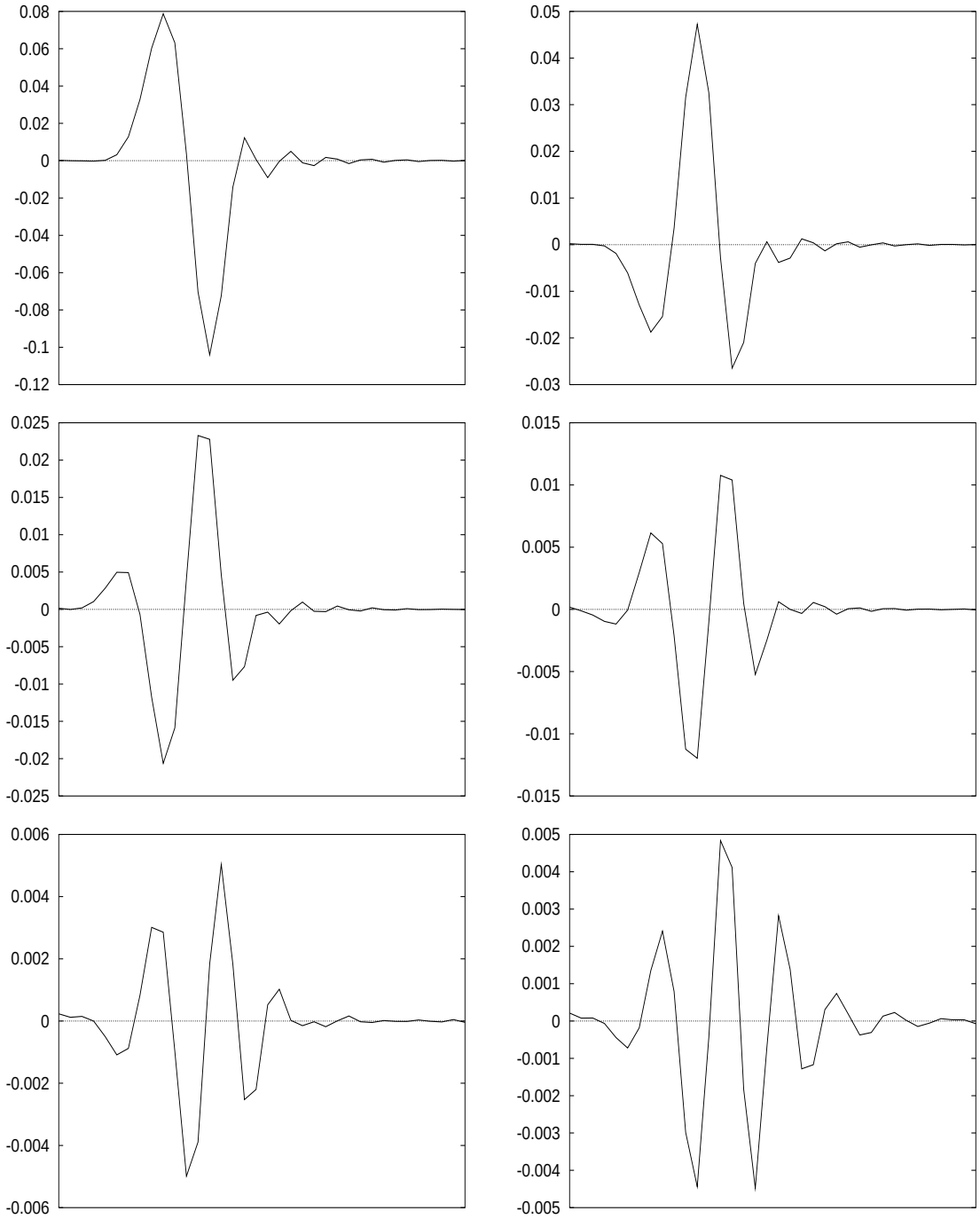

Figure 7: $\mathcal{C}_s^{(2,2)}$ for $s = 1$ to 6 ($N = 7$)

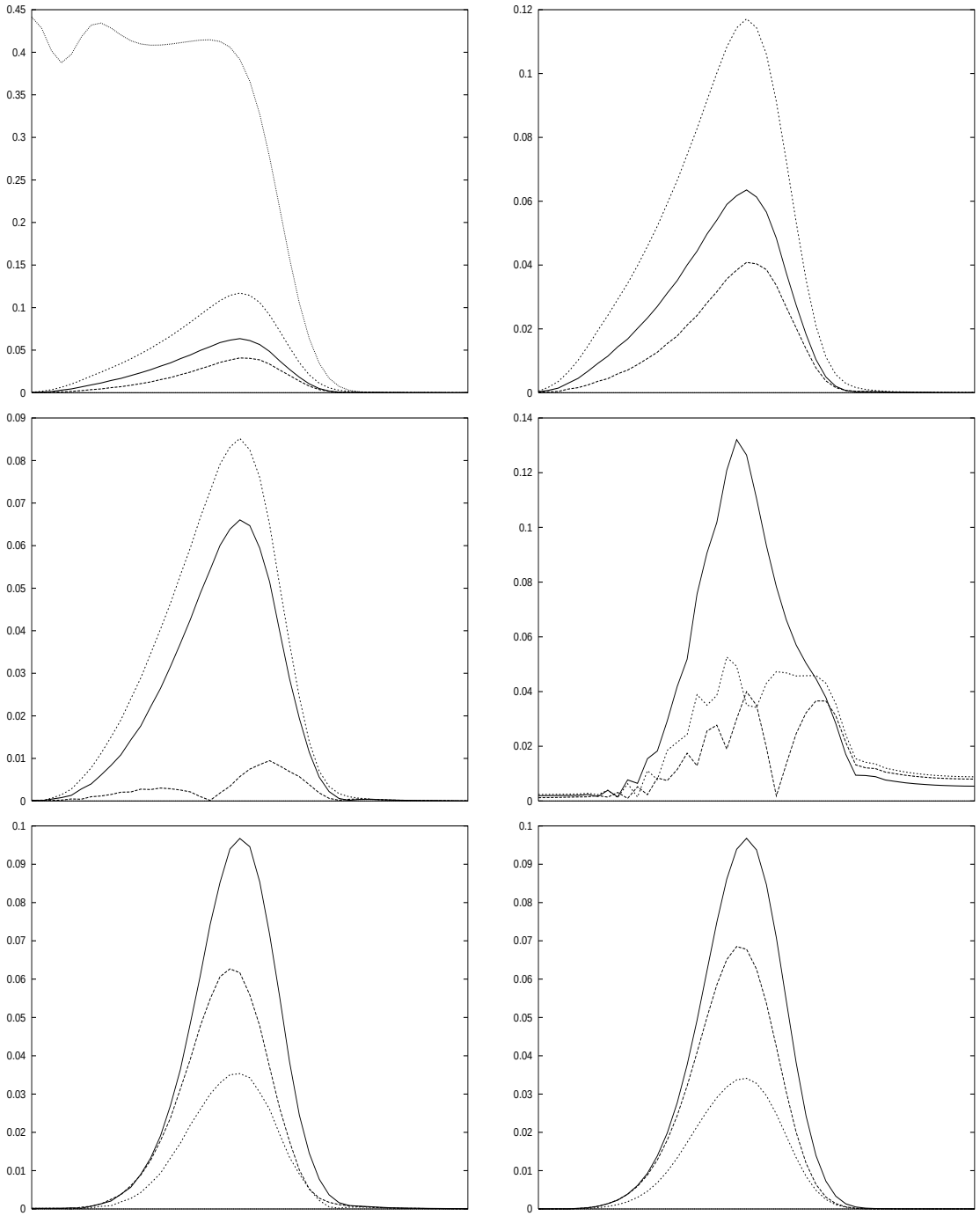

Figure 8: Spectra of the reflected signals in cases $(s,p)=(2,2)$ (also with the spectrum of the initial signal), $(s,p)=(3,2)$, $(s,p)=(5,1)$, $(s,p)=(5,2)$ and $(s,p)=(5,3)$.

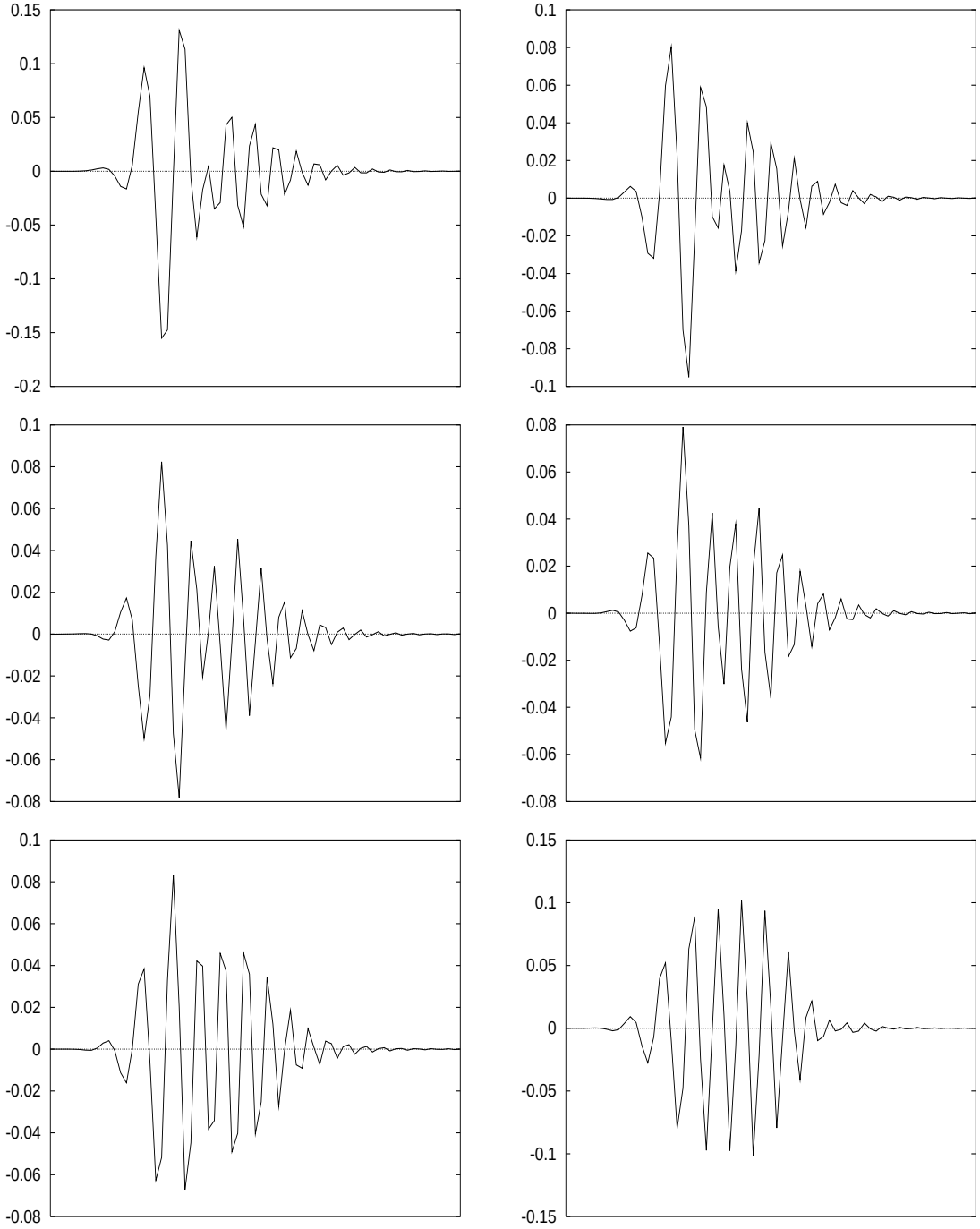

Figure 9: $\mathcal{B}_s^2$ for $s = 1$ to 6 ($N = 15$)

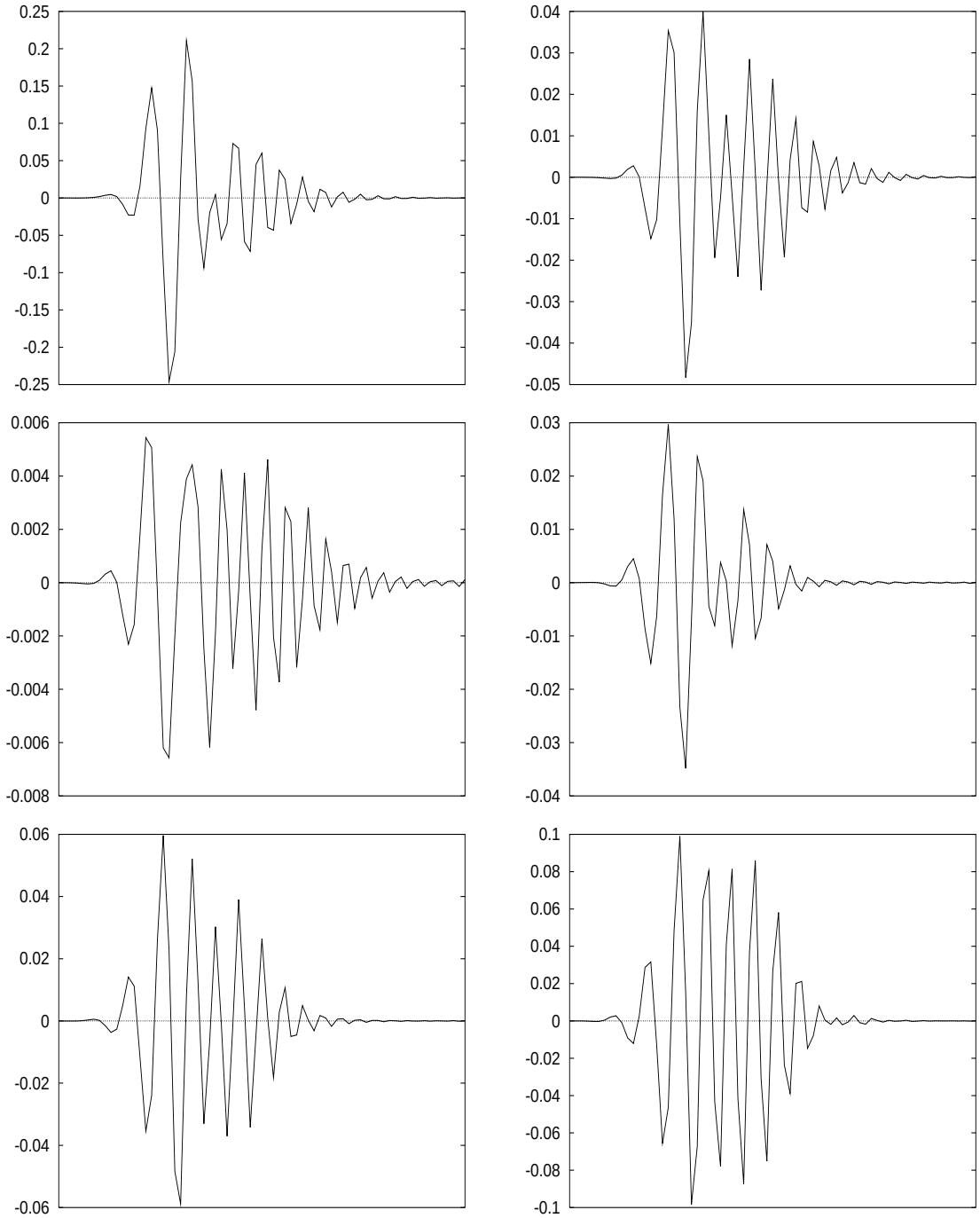

Figure 10: $\mathcal{C}_s^{(2,1)}$ for $s = 1$ to 6 ($N = 15$)

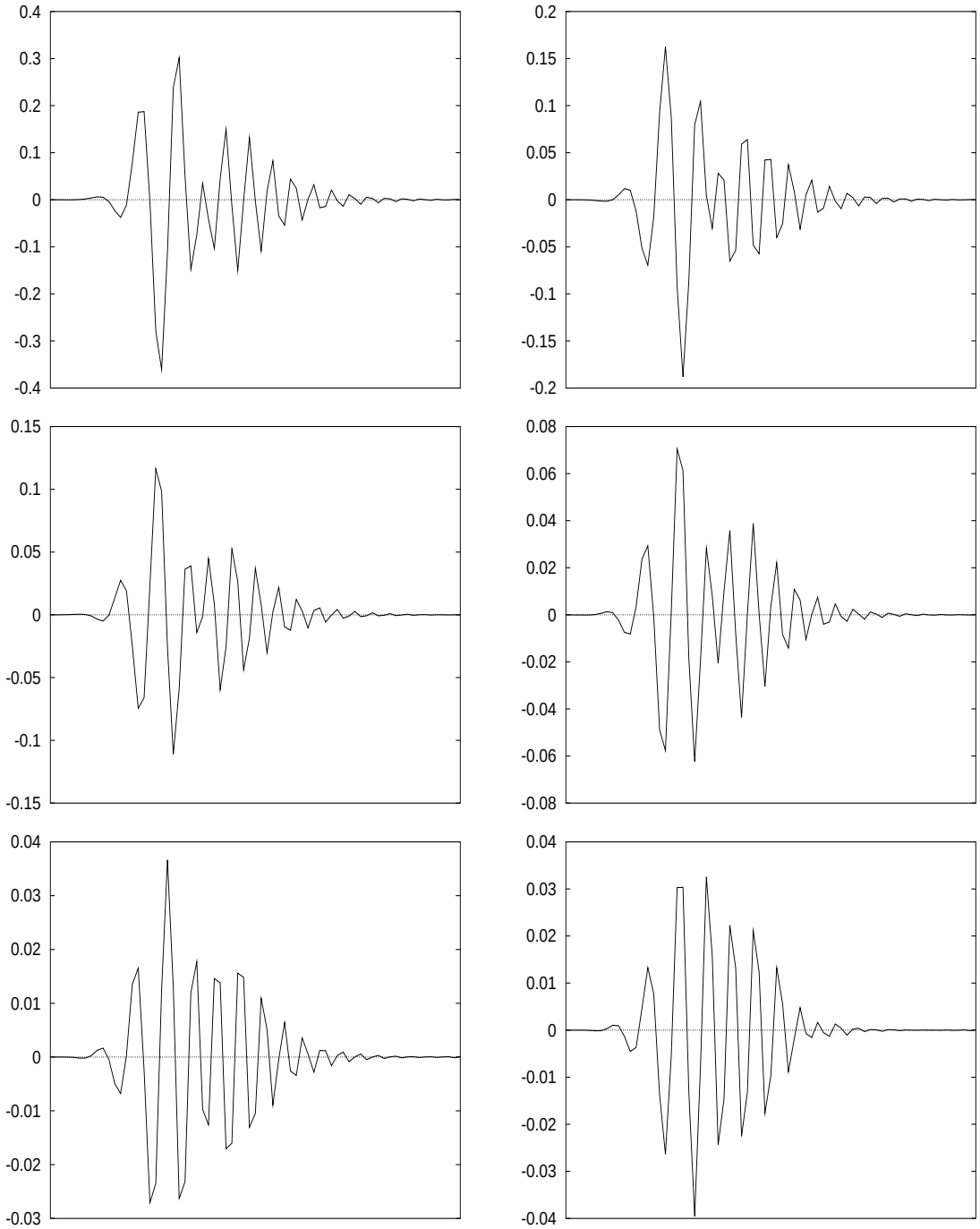

Figure 11: $\mathcal{C}_s^{(2,2)}$ for $s = 1$ to 6 ($N = 15$)

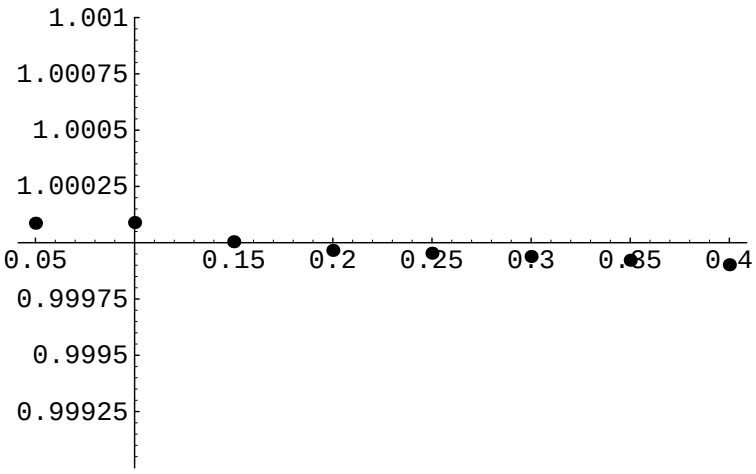

Figure 12: Spectral radius for the boundary condition $\mathcal{B}_6^1$ as function of α.

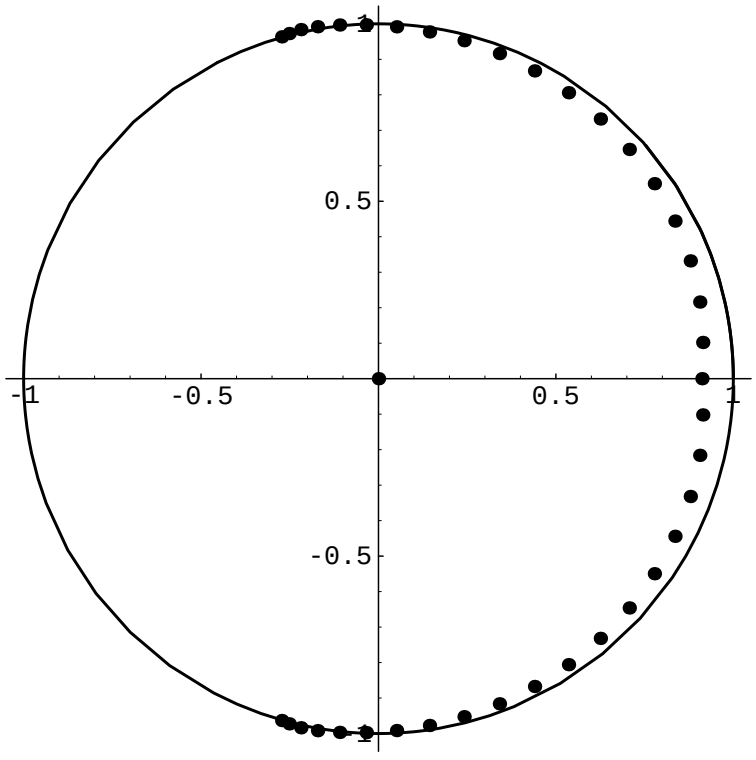

Figure 13: Spectrum for the boundary condition $\mathcal{C}_1^{(1,1)}$ for $\alpha = 0.8$.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l' Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

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