ELSEVIER

Applied Numerical Mathematics 33 (2000) 327–340

APPLIED NUMERICAL MATHEMATICS

www.elsevier.nl/locate/apnum

Stability and accuracy of optimal local non-reflecting boundary conditions

Avram Sidi [a], Dan Givoli [b,*]

[a] *Department of Computer Science, Technion—Israel Institute of Technology, Haifa 32000, Israel*
[b] *Department of Aerospace Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel*

Abstract

Problems in unbounded domains are often solved numerically by truncating the infinite domain via an artificial boundary $\mathcal{B}$ and applying some boundary condition on $\mathcal{B}$, which is called a Non-Reflecting Boundary Condition (NRBC). Recently, a two-parameter hierarchy of optimal local NRBCs of increasing order has been developed. The optimality is in the sense that the local NRBC best approximates the exact nonlocal Dirichlet-to-Neumann (DtN) boundary condition in the L_2 norm for functions in C^∞. The optimal NRBCs are combined with finite element discretization in the computational domain. Here the *theoretical* properties of the resulting class of schemes are examined. In particular, theorems are proved regarding the numerical stability of the schemes and their rates of convergence. © 2000 IMACS. Published by Elsevier Science B.V. All rights reserved.

Keywords: Exterior problems; Non-reflecting boundary condition; Unbounded domain; Dirichlet-to-Neumann (DtN); Finite elements

1. Introduction

There are many methods to solve boundary value problems in unbounded domains [7]. One popular class of methods is that based on the use of artificial boundary conditions, which are also called absorbing boundary conditions or Non-Reflecting Boundary Conditions (NRBCs), especially in the context of wave problems [6]. The use of NRBCs comprises of three steps:

(a) Introduce an artificial boundary $\mathcal{B}$, which divides the original infinite domain into two domains: a finite computational domain Ω and an infinite residual domain D.
(b) By analyzing the problem in D, obtain a relation on $\mathcal{B}$ (exact or approximate) involving the unknown function u and its derivatives. Use this relation as a boundary condition on $\mathcal{B}$, to obtain a well-posed problem in Ω.
(c) Solve the problem in Ω numerically.

* Corresponding author. E-mail: givolid@aerodyne.technion.ac.il

PII: S0168-9274(99)00099-9

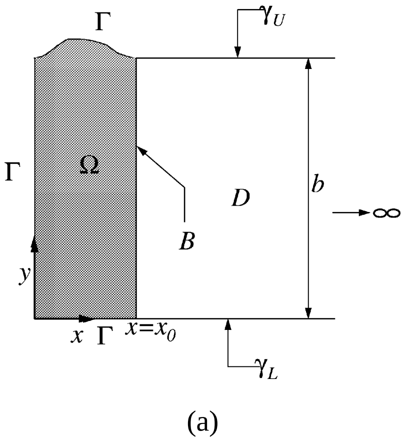

Fig. 1. Setup of the DtN method for (a) a semi-infinite strip problem, (b) an exterior problem of scattering or radiation from an obstacle.

Figs. 1(a) and (b) illustrate the typical setup for a problem in a semi-infinite strip (or a channel, or a wave guide) and in the domain exterior to an obstacle or scatterer, respectively.

Most of the NRBCs which have been proposed in the literature are *local* and *approximate*. Perhaps the most commonly used ones are the NRBCs of Engquist and Majda [4], and of Bayliss and Turkel [2]. A smaller number of *exact nonlocal* NRBCs have been devised for various problems in infinite domains [5,9,18–20,28,29]. For general linear elliptic problems, Keller and Givoli [10,22] devised an exact NRBC on an artificial boundary $\mathcal{B}$ of a simple shape (e.g., a circle in 2D or a sphere in 3D). This NRBC involves the Dirichlet-to-Neumann (DtN) map on $\mathcal{B}$, and is thus called the *DtN boundary condition*. It has been incorporated in a finite element scheme, resulting in the general *DtN Finite Element method* [7], [10,11,17]. More recently, the method was extended to treat classes of linear hyperbolic problems [8], nonlinear elliptic problems [14,25] and nonlinear hyperbolic problems [13].

Despite the fact that exact nonlocal NRBCs may be extremely useful in many situations, there are cases where *local* NRBCs may be preferred. A discussion of the relative advantages and disadvantages of local and nonlocal NRBCs can be found in [16]. The main conclusion from this comparison is that there is place for *both* types of NRBCs in computational schemes.

Recently, Givoli and Patlashenko [15] have constructed a hierarchy of *optimal local* NRBCs of increasing order. The approach is based on considering a local boundary condition of a given form (a given “order”) with unknown coefficients, and asking the following question: *What is the best choice for the unknown coefficients so that the local operator in the boundary condition be the best approximation of the DtN map*, in a certain norm? This question has been answered in [15] in a C^∞ framework, using the L_2 norm for functions that can be Fourier decomposed. The optimal conditions constitute a two-parameter hierarchy; the two parameters are N, the order of the boundary condition, and M, the number of harmonics (or Fourier modes) taken into account. Here $M \geqslant N$. The NRBCs with $M = N$ are the same as those constructed in [16,23,24] in a totally different manner.

In the computational domain Ω the finite element method is employed. If $N \geqslant 2$, special finite elements must be used, which possess high-order regularity along $\mathcal{B}$. A hierarchy of such elements in two and three dimensions has been devised in [12,16,24].

In [15] the (N,M) hierarchy was constructed and some numerical experiments demonstrated the performance of the resulting schemes. In this paper we examine the *theoretical* properties of these schemes. In particular, we analyze their *numerical stability* and *rates of convergence*.

Following is the outline of the paper. In Section 2 we briefly recall the construction of the optimal local NRBCs. In Section 3 we prove a theorem on the numerical stability of the (N,M) schemes. In Section 4 we derive an error estimate for the (N,M) schemes, which is an extension of the estimate given in [16] for the case $N=M$. We close with some remarks in Section 5. In the Appendix we discuss the properties of the optimal coefficients, and we prove some lemmas which are used in the body of the paper.

2. Optimal local NRBCs

In this section we briefly summarize the construction of the optimal local (N,M) conditions introduced in [15]. To fix ideas, we concentrate on the case shown in Fig. 1(b), namely a two-dimensional exterior problem, where $\mathcal{B}$ is a circle. However, these ideas carry over in a straightforward manner to other configurations and to three dimensions.

We let R be the radius of the artificial boundary $\mathcal{B}$. In the DtN method, we impose the DtN boundary condition on $\mathcal{B}$. This condition has the form

$$-\frac{\partial u}{\partial r}(R,\theta)=(\mathcal{M}u)(\theta)\equiv\sum_{n=0}^{\infty}{}'\int_0^{2\pi} m_n(\theta,\theta')\,u(R,\theta')\,\mathrm{d}\theta', \tag{1}$$

where

$$m_n(\theta,\theta')=\frac{Z_n}{\pi R}\cos n(\theta-\theta'). \tag{2}$$

In (1), $\mathcal{M}$ is the DtN map, which is a *nonlocal* operator, and the prime after the sum indicates that a factor of $1/2$ multiplies the term with $n=0$. In (2), the Z_n are constants. For example, for Laplace equation, $Z_n=n$. See [7,10,11,22], and the summary in [16] for further details.

We wish to approximate the nonlocal operator $\mathcal{M}$ by an Nth-order *local* operator. Thus, we replace (1) by

$$-\frac{\partial u}{\partial r}(R,\theta)=(L_N u)(\theta)\equiv\frac{1}{R}\sum_{n=0}^{N}A_n\frac{\partial^{2n}u}{\partial\theta^{2n}}. \tag{3}$$

Here the A_n are constant coefficients. Now we ask the following question: how should the coefficients A_n, $n=0,\dots,N$, be chosen, such that L_N is the "best approximation" of $\mathcal{M}$?

We answer this question in the L_2 norm for functions in C^∞. We consider the Fourier expansion of u,

$$u=\sum_{j=0}^{M}{}'\left(u_j^c\cos j\theta+u_j^s\sin j\theta\right). \tag{4}$$

Here, M is a chosen number of modes, and $M\geqslant N$. Thus, the procedure described here constitutes a two-parameter (N,M) family of schemes.

Now, we apply both operators $\mathcal{M}$ and L to the function u whose expansion is given in (4). We take the difference between the two, and calculate the $L_2(\mathcal{B})$ norm of this difference. This yields

$$||(\mathcal{M}-L_N)u||_0^2 = \frac{\pi}{R^2}\sum_{j=0}^{M}{}'\left(|u_j^c|^2+|u_j^s|^2\right)|\beta_j^N|^2, \tag{5}$$

where

$$\beta_j^N = Z_j - \sum_{k=0}^{N} A_k(-1)^k j^{2k}. \tag{6}$$

Suppose we have some estimate of the relative importance of the different modes of u (see examples in [15]). In this case, we let W_j be given positive weights, where

$$W_0 \simeq \frac{\pi}{2}\left(|u_0^c|^2+|u_0^s|^2\right), \qquad W_j \simeq \pi\left(|u_j^c|^2+|u_j^s|^2\right), \quad j=1,\ldots,M. \tag{7}$$

Then (5) gives

$$||(\mathcal{M}-L_N)u||_0^2 \simeq \frac{1}{R^2}\sum_{j=0}^{M} W_j|\beta_j^N|^2 \equiv \frac{1}{R^2}||\boldsymbol{\beta}^N||_W^2. \tag{8}$$

Here $||\boldsymbol{\beta}^N||_W$ is the *weighted* Euclidean norm of the vector $\boldsymbol{\beta}^N$, whose entries are defined by (6). If nothing is known about the relative amplitudes of the modes of u, one should simply take all the weights to have value unity.

A necessary condition for the minimum of $||\boldsymbol{\beta}^N||_W^2$ is $\partial||\boldsymbol{\beta}^N||_W^2/\partial A_l = 0$ for $l=0,\ldots,N$. This gives the linear symmetric system of equations

$$\boldsymbol{B}\boldsymbol{A} = \boldsymbol{P}, \tag{9}$$

where

$$P_l = \sum_{j=0}^{M} W_j Z_j(-1)^l j^{2l}, \qquad l=0,\ldots,N, \tag{10}$$

$$B_{lk} = \sum_{j=0}^{M} W_j(-1)^{(k+l)} j^{2(k+l)}, \quad k,l=0,\ldots,N. \tag{11}$$

The solution of this system yields the desired coefficients A_k in the local boundary condition (3).

It is convenient to define $\alpha_k^{(NM)} = (-1)^k A_k$, emphasizing the fact that the optimal A_k obtained from (9) depend on the chosen parameters N and M. Then (9)–(11) are rewritten as

$$\sum_{k=0}^{N}\sum_{j=0}^{M} W_j j^{2(k+l)}\alpha_k^{(NM)} = \sum_{j=0}^{M} W_j Z_j j^{2l}, \quad l=0,\ldots,N. \tag{12}$$

With this notation, the local condition (3) is rewritten as

$$-\frac{\partial u}{\partial r}(R,\theta) = (L_{NM}u)(\theta) \equiv \frac{1}{R}\sum_{n=0}^{N}(-1)^n\alpha_n^{(NM)}\frac{\partial^{2n}u}{\partial\theta^{2n}}. \tag{13}$$

We have derived this result for the two-dimensional exterior problem (Fig. 1(b)). However, similar results hold for other configurations as well. In the case of a *semi-infinite strip* (Fig. 1(a)) with the boundary condition $u = 0$ on the two semi-infinite rays (a case which will be referred to later), it is natural to change the ranges of the indices in (12) (see [15,16]): j varies from 1 to M, whereas k and l vary from 0 to $N - 1$. Thus, (12) is replaced by the linear system

$$\sum_{k=1}^{N}\sum_{j=1}^{M} W_j\, j^{2(k+l)-4}\alpha_k^{(NM)} = \sum_{j=1}^{M} W_j Z_j\, j^{2l-2}, \quad l = 1, \ldots, N. \tag{14}$$

Also, (13) is replaced by

$$-\frac{\partial u}{\partial x}(x_0, y) = (L_{NM}u)(y) \equiv \sum_{n=1}^{N}(-1)^n \alpha_n^{(NM)} \frac{\partial^{2(n-1)}u}{\partial y^{2(n-1)}}, \tag{15}$$

on the boundary $\mathcal{B}$, which is defined by $x = x_0$, $0 \leqslant y \leqslant b$ (see Fig. 1(a)).

We remark that in the special case $N = M$, the optimal coefficients $\alpha_k^{(NM)}$ reduce to those obtained (in a totally different manner) in [16,23,24]. In this case it is easily shown that the weights W_j drop out of the formulation.

3. Numerical stability

We now consider the numerical scheme that consists in using the finite element method to solve the given PDE in Ω together with the local boundary condition (13) or (15) on $\mathcal{B}$. We assume that the partial differential operator governing in Ω is self-adjoint and positive. Thus, Laplace's equation, the modified Helmholtz equation and the equations of static linear elasticity fit into the analysis that follows, whereas the Helmholtz equation and the advection–diffusion equation do not. For concreteness we focus our attention on the case of a semi-infinite strip (Fig. 1(a)) with the boundary condition $u = 0$ on the two semi-infinite rays, although the analysis can easily be extended to other configurations.

The weak formulation of the problem can be written in the form: Find $u \in \mathcal{S}$ such that for all $w \in \mathcal{S}_0$,

$$a(w, u) + b_{NM}(w, u) = c(w). \tag{16}$$

Here $\mathcal{S}$ and $\mathcal{S}_0$ are appropriate function spaces, $a(\cdot,\cdot)$ and $b_{NM}(\cdot,\cdot)$ are symmetric bilinear forms, and $c(\cdot)$ is a linear form. The $a(\cdot,\cdot)$ and $c(\cdot)$ terms are standard, whereas the $b_{NM}(\cdot,\cdot)$ term is the contribution of the condition (15) on $\mathcal{B}$. The expression for the latter is easily obtained (cf. [16]):

$$b_{NM}(w, u) = \int_B w L_{NM} u \,\mathrm{d}\mathcal{B} = \sum_{n=1}^{N} \alpha_n^{(NM)} \int_{\mathcal{B}} \left(\frac{\partial^{n-1}w}{\partial y^{n-1}}\right)\left(\frac{\partial^{n-1}u}{\partial y^{n-1}}\right) \mathrm{d}y. \tag{17}$$

The numerical stability of standard finite element schemes for well-posed boundary value problems with symmetric positive operators is closely related to the fact that $a(\cdot,\cdot)$ is positive definite, namely that $a(w^h, w^h) > 0$ for any function w^h in the finite element space that is not identically zero [21,27]. Then the $a(\cdot,\cdot)$-form induces the "energy norm" $||w^h|| = (a(w^h, w^h))^{1/2}$. On the discrete level this property implies that the finite element stiffness matrix is positive definite. When the boundary term $b_{NM}(w, u)$ is included in the weak form (16), it is desirable to know whether $a(\cdot,\cdot) + b_{NM}(\cdot,\cdot)$ remains positive definite. A sufficient condition for this is that $b_{NM}(w^h, w^h) \geqslant 0$ for any w^h, which we prove next.

First we prove:

Lemma 1. *Let* $\alpha_n^{(NM)}$ *satisfy* (14) *with* $Z_j = f(j^2)$, *where* $f \in C[0,\infty]$ *and* $f \in C^N(0,\infty)$ *such that* $(-1)^{j-1} f^{(j)}(x) > 0$ *for all* $x > 0$.

(i) *If* $N = 1, 2$ *and* M *is arbitrary, or* $N = M = 4, 6, 8, \ldots$, *then*

$$\sum_{n=1}^{N} \alpha_n^{(NM)} k^{2(n-1)} > 0, \quad k = 1, 2, \ldots . \tag{18}$$

(ii) *In case* $M > N = 4, 6, 8, \ldots$, *then* (18) *holds with* $k = M, M+1, M+2, \ldots$.

Proof. The results follow directly from Lemmas A.2 and A.4 in the Appendix. □

Now we have:

Theorem 1. *Assume the conditions of Lemma* 1. *Then, for any Fourier decomposable function* w, *we have* $b_{NM}(w,w) \geqslant 0$, *when* (i) M *is arbitrary and* $N = 1, 2$, *or* (ii) $M = N = 4, 6, 8, \ldots$, *or* (iii) M *is arbitrary and* $N = 4, 6, 8, \ldots$ $(N < M)$ *and also*

$$\sum_{n=1}^{N} \alpha_n^{(NM)} k^{2(n-1)} \geqslant 0, \quad k = 2, 3, \ldots, M-1 .$$

Proof. We expand w in a Fourier series:

$$w = \sum_{k=1}^{\infty} w_k \sin \frac{k\pi y}{b}.$$

We substitute this expansion in (17) and make use of the orthogonality of the sines to get

$$b_{NM}(w,w) = \frac{b}{2} \sum_{n=1}^{N} \alpha_n^{(NM)} \sum_{k=1}^{\infty} w_k^2 k^{2(n-1)}. \tag{19}$$

A sufficient condition for $b_{NM}(w,w)$ to be non-negative is thus

$$\sum_{n=1}^{N} \alpha_n^{(NM)} k^{2(n-1)} \geqslant 0 \quad \text{for } k = 1, 2, \ldots .$$

But this condition is indeed satisfied by Lemma 1. □

Note that the extra condition for the case $N < M$ in Theorem 1 can be verified by performing a *finite* number $(M-2)$ of computations. (See [16] for examples of such computations.) Our numerical experiments with different values of M, N and various sets of weights W_j, suggest that this condition is satisfied in general, although we do not have a proof of this at the time of writing.

A consequence of this theorem is that *the optimal local boundary conditions with* $N = 1$ *or* N *even and any* M *are numerically stable.* On the other hand, the conditions with $N \neq 1$ odd are potentially unstable. Numerical experiments [15,16] show that this is indeed the case. Moreover, the same behavior is observed for the Helmholtz equation which is beyond the scope of this analysis.

4. Error estimate

When the optimal local NRBC (13) or (15) is incorporated in a *finite element* scheme, the resulting method involves four computational parameters: h, the mesh parameter, which is roughly the size of the largest element in the mesh; p, the polynomial degree of the finite element space; N, the order of the local NRBC; and M, the number of Fourier modes taken into account. We consider here an h-type finite element method, where p is small and fixed and the error is reduced by reducing h (i.e., by refining the mesh). Our goal is to determine the rate of convergence of the method as a function of these four parameters.

In [16], Givoli et al. have derived an error estimate for a sequence of NRBCs that happen to coincide with the optimal conditions (13) and (15) *when* $N = M$ (although they were constructed in a different way in [16]). This error estimate can be stated as follows:

Theorem 2 (The case $N = M$). *If* (a) N *is 1 or even* (*i.e., the NRBC is stable*), (b) *the coefficients* $\alpha_n^{(NN)}$ *satisfy* (14), *and* (c) $u \in H^r(\Omega)$ *where* $r > \max(1, 2N-2)$ *and* $r \geqslant p+1$, *then*

$$||e||_1 \leqslant \left(C_1 h^p + C_2 N^{-\mu}\right)||u||_r, \tag{20}$$

where $\mu = \min(r-1, r-2N+2)$.

Here e is the error, i.e., the difference between the exact and finite element solutions, H^r is the Sobolev space of order r, $||\cdot||_r$ is the norm in this space, and C_1 and C_2 are constants. The first term on the right-hand side of (20) is the standard finite element error estimate, whereas the second term shows the rate of convergence with respect to N. The parameter r is the degree of smoothness of the exact solution u. If u is infinitely smooth, then $\mu \to \infty$, which indicates exponential convergence in N. Numerical experiments [16] show agreement with the estimate (20).

The proof of Theorem 2 [16] is quite long, but most of it does not depend on the properties of the coefficients α_n. The only part in the proof that involves the α_n can be summarized by the following lemma (see [16, Section 7.6]):

Lemma 2 (The case $N = M$). *For any* w_n *and* u_n,

$$T_N \equiv \left| \sum_{n=1}^{\infty} \left(\sum_{m=1}^{N} \alpha_m^{(NN)} n^{2(m-1)} \right) w_n u_n - \sum_{n=1}^{N} n w_n u_n \right| \leqslant C \sum_{n=N+1}^{\infty} n^{2(N-1)} |w_n u_n|, \tag{21}$$

where C *is a constant that does not depend on* N.

Proof. The proof in [16] does not actually show that C is independent of N, and therefore we give here the full proof. Making the substitution $Q_N(x) = \sum_{m=1}^{N} \alpha_m^{(NN)} x^{m-1}$, we first observe that

$$T_N = \left| \sum_{n=1}^{\infty} Q_N(n^2) w_n u_n - \sum_{n=1}^{N} n w_n u_n \right| \leqslant \sum_{n=N+1}^{\infty} \left| Q_N\left(n^2\right)\right| |w_n u_n|,$$

since $Q_N(n^2) = n$, $n = 1, \ldots, N$. The proof can now be completed by invoking Lemma A.3. $\Box$

Now we extend the error estimate (20) to the case $M > N$. To this end, we first extend Lemma 2 to this case:

Lemma 3 (The case $M > N$). *Let $M > N > 1$ and assume that the W_n satisfy the growth condition $W_n \sim An^{-4(N-1)}$ as $n \to \infty$, where $A > 0$ is some constant independent of N and M. If $M - N \leqslant K$, where K is fixed hence independent of N and M, and $Z_j = j$, $j = 1, 2, \ldots$, then, for any w_n and u_n,*

$$U_{NM} \equiv \left| \sum_{n=1}^{\infty} \left(\sum_{m=1}^{N} \alpha_m^{(NM)} n^{2(m-1)} \right) w_n u_n - \sum_{n=1}^{M} n w_n u_n \right| \leqslant C \sum_{n=1}^{\infty} n^{2(N-1)} |w_n u_n|, \tag{22}$$

where C is a constant that does not depend on N and M.

Proof. Making the substitution $P_{NM}(x) = \sum_{m=1}^{N} \alpha_m^{(NM)} x^{m-1}$, we first observe that

$$U_{NM} = \left| \sum_{n=1}^{M} \left[P_{NM}(n^2) - n \right] w_n u_n + \sum_{n=M+1}^{\infty} P_{NM}(n^2) w_n u_n \right| \leqslant S_1 + S_2,$$

where

$$S_1 = \sum_{n=1}^{M} \left| P_{NM}(n^2) - n \right| |w_n u_n| \quad \text{and} \quad S_2 = \sum_{n=M+1}^{\infty} \left| P_{NM}(n^2) \right| |w_n u_n|.$$

Applying part (iii)(b) of Lemma A.4, and using the fact that $f(x) = \sqrt{x} \leqslant x^{N-1}$ for $x \geqslant 1$ when $N > 1$, we have $|P_{NM}(x)| \leqslant E_2 x^{N-1}$ for $x \geqslant x_M = M^2$ when $N \geqslant 2$, with $E_2 > 0$ a constant independent of N and M. From this we conclude that, when $N \geqslant 2$,

$$S_2 \leqslant E_2 \sum_{n=M+1}^{\infty} n^{2(N-1)} |w_n u_n|.$$

We next proceed to S_1. First, we have

$$\begin{aligned} S_1 &\leqslant \left(\sum_{n=1}^{M} \frac{1}{\sqrt{W_n}} |w_n u_n| \right) \max_{1 \leqslant n \leqslant M} \left(\sqrt{W_n} \left| P_{NM}(n^2) - n \right| \right) \\ &\leqslant \left(\sum_{n=1}^{M} \frac{1}{\sqrt{W_n}} |w_n u_n| \right) \left(\sum_{n=1}^{M} W_n \left[P_{NM}(n^2) - n \right]^2 \right)^{1/2}. \end{aligned}$$

Now $P_{NM}(x)$ satisfies (A.6) in the Appendix. Setting $R(x) = Q_N(x) = P_{NN}(x)$ in (A.6) and recalling that $Q_N(n^2) = n$, $n = 1, \ldots, N$, we obtain

$$\left(\sum_{n=1}^{M} W_n \left[P_{NM}(n^2) - n \right]^2 \right)^{1/2} \leqslant \left(\sum_{n=N+1}^{M} W_n \left[Q_N(n^2) - n \right]^2 \right)^{1/2}.$$

Applying Lemma A.3 with $f(x) = \sqrt{x}$ to $Q_N(x)$ and again using the fact that $\sqrt{x} \leqslant x^{N-1}$ for $x \geqslant 1$ when $N > 1$, we obtain $|Q_N(x) - \sqrt{x}| \leqslant E' x^{N-1}$ for $x \geqslant x_N = N^2$, where $E' > 0$ is a constant independent of N and M. As a result,

$$\left(\sum_{n=1}^{M} W_n \left[P_{NM}(n^2) - n \right]^2 \right)^{1/2} \leqslant E' \left(\sum_{n=N+1}^{M} W_n n^{4(N-1)} \right)^{1/2} \quad \text{for } N \geqslant 2.$$

By the asymptotic growth condition imposed on W_n we have $W_n n^{4(N-1)} \sim A$ as $n \to \infty$, as a result of which

$$\left(\sum_{n=N+1}^{M} W_n n^{4(N-1)} \right)^{1/2} \leqslant E''$$

for some constant $E'' > 0$ independent of N and M whenever $M - N \leqslant K$ with K fixed. By the same growth condition on the W_n we also have

$$\frac{1}{\sqrt{W_n}} \sim \frac{1}{\sqrt{A}} n^{2(N-1)} \quad \text{as } n \to \infty,$$

from which

$$\sum_{n=1}^{M} \frac{1}{\sqrt{W_n}} |w_n u_n| \leqslant E''' \sum_{n=1}^{M} n^{2(N-1)} |w_n u_n|,$$

where $E''' > 0$ is a constant independent of N and M. Combining all this, we obtain

$$S_1 \leqslant E_1 \sum_{n=1}^{M} n^{2(N-1)} |w_n u_n|$$

with $E_1 = E' E'' E''' > 0$ a constant independent of N and M. Letting $C = \max\{E_1, E_2\}$, the result in (22) now follows. $\square$

This lemma again leads to the error estimate given in Theorem 2. Note, however, that additional assumptions are made here regarding the asymptotic behavior of the weights W_n, and that the difference $M - N$ is required to be bounded. These assumptions are reasonable in light of the numerical experiments presented in [15].

We have also proved a slightly more general version of Lemma 3, where the asymptotic behavior of the weights is $W_n \sim A n^{-4(N-1-\varepsilon)}$ as $n \to \infty$, with $0 \leqslant \varepsilon \leqslant 1$. Moreover, if $\varepsilon < 3/4$ then the requirement $M - N \leqslant K$ can be dropped, thus enabling us to vary N and M arbitrarily. The proof of this more general lemma is somewhat lengthy, and therefore we shall not give it here.

5. Concluding remarks

We have considered a two-parameter hierarchy of local approximate NRBCs that have been constructed in [15], for the numerical solution of elliptic boundary value problems in unbounded domains. These NRBCs are optimal in the sense that they best approximate the exact nonlocal Dirichlet-to-Neumann (DtN) boundary condition for C^∞ functions in the L_2 norm. The optimal local NRBC may be of low order but still represent high-order modes. We have analyzed the stability and accuracy properties of these NRBCs, and obtained some theoretical results which are compatible with those of the numerical experiments performed in [15,16].

One research direction that is worth pursuing is deriving optimal local NRBCs which are based on optimization with respect to norms other than the L_2 norm. One possibility in the latter context is to work in the space $H^{-1/2}$, where the DtN map and the local differential operator are bounded [3]. Another area of investigation is the construction of optimal local NRBCs for time-dependent wave propagation problems. This can be done within the framework of a number of schemes, including the ones described in [1,8].

Appendix. Properties of the coefficients $\alpha_n^{(NM)}$

We start with the case $N = M$ for which the equations in (14) become

$$\sum_{k=1}^{N} \alpha_k^{(NN)} j^{2(k-1)} = Z_j, \quad j = 1, \ldots, N, \tag{A.1}$$

independently of the weights W_j. In this case $Q_N(x) = \sum_{k=1}^{N} \alpha_k^{(NN)} x^{k-1}$ is a polynomial of interpolation to a function $f(x)$ at the points $x_j = j^2$, $j = 1, \ldots, N$, such that $f(x_j) = Z_j$ for all j. For the general case in which $0 < x_1 < x_2 < \cdots < x_N$ but x_j are arbitrary otherwise, the following lemma has been proved in [26].

Lemma A.1. *Let $f(x)$ be such that*
(a) *$f \in C[0, X]$ for some $X > 0$ with $f(0) \geqslant 0$, and*
(b) *$f \in C^N(0, X)$ with $(-1)^{j-1} f^{(j)}(x) > 0$ for $x \in (0, X)$, $j = 1, \ldots, N$.*
Let $0 < x_1 < x_2 < \cdots < x_N < X$ with x_j arbitrary otherwise. If $Q_N(x) = \sum_{k=1}^{N} \alpha_k^{(NN)} x^{k-1}$ is a polynomial of interpolation to $f(x)$ at $x_1, \ldots, x_N$, then the following are true:
(i) *$\alpha_1^{(NN)}, \alpha_2^{(NN)}, \ldots,$ have the sign pattern $+, +, -, +, -, +, \ldots$.*
(ii) *With $f(x) = \sqrt{x}$ and $x_j = j^2$, $j = 1, 2, \ldots,$ from which $Z_j = j$ for all j, we have $|\alpha_k^{(NN)}| \leqslant C$ for all k and N, where C is some constant independent of k and N. Moreover, $\lim_{N\to\infty} \alpha_k^{(NN)} = \widehat{\alpha}_k$ for some constant $\widehat{\alpha}_k$, and $\lim_{N\to\infty} \alpha_{N-k}^{(NN)} = 0$, where k is hold fixed in both cases.*
(iii) *With $f(x)$ and x_j as in part* (ii), *the l_1 condition number of the linear system in* (A.1) *is at best* $\mathrm{O}(N^{2N-2})$ *and at worst* $\mathrm{O}(N^{2N-1})$ *as $N \to \infty$.*

For more refined statements of the results of Lemma A.1 we refer the reader to [26].

Lemma A.2. *Let $f(x)$ and x_j satisfy conditions* (a) *and* (b) *of Lemma* A.1. *Then*
(i) *for $N = 1, 2$, we have $Q_N(x) > 0$ for all $x > 0$, and*
(ii) *for $N \geqslant 4$ and even, we have $Q_N(x_j) = f(x_j) > 0$, $j = 1, \ldots, N$, and $Q_N(x) > 0$ for all $x \in (x_N, X)$ or $x \in [0, x_1)$ as well.*

Proof. Part (i) follows from the fact that $\alpha_k^{(NN)} > 0$ for $N = 1, 2$. Part (ii) follows from the fact that $f(x) > 0$ for $x \in (0, X)$ that is guaranteed by $f(0) \geqslant 0$ and $f'(x) > 0$ for $x \in (0, X)$, and from the relation

$$Q_N(x) = f(x) - \frac{f^{(n)}(\xi(x))}{N!} \prod_{i=1}^{N} (x - x_i) \quad \text{for some } \xi(x) \in \bigl(\min\{x_1, x\}, \max\{x_N, x\}\bigr), \tag{A.2}$$

which is a rearrangement of the error formula for polynomial interpolation. □

By imposing certain growth conditions on the $f^{(k)}(x)$ we can derive a useful upper bound on $|Q_N(x)|$, to which we now turn.

Lemma A.3. *Let $f(x)$ and the x_j be as in Lemma* A.2*, and assume, in addition, that*

$$\sup_{x \geqslant x_1} \left| \frac{f^{(k)}(x)}{k!} \right| \leqslant D k^{-1-\delta}, \quad k = 1, 2, \dots, \tag{A.3}$$

for some $D > 0$ and $\delta > 0$. Then there exists a constant $E > 0$ independent of N such that

$$|Q_N(x)| \leqslant E x^{N-1} \quad \textit{for } x \geqslant \max\{1, x_{N-1}\}. \tag{A.4}$$

Proof. Let us consider the Newton form of $Q_N(x)$, namely,

$$Q_N(x) = f(x_1) + \sum_{k=2}^{N} f[x_1, x_2, \dots, x_k] \prod_{i=1}^{k-1} (x - x_i),$$

where $f[x_1, x_2, \dots, x_k]$ is the divided difference of $f(x)$ over the set of points $\{x_1, x_2, \dots, x_k\}$. Obviously,

$$|Q_N(x)| \leqslant \left(|f(x_1)| + \sum_{k=2}^{N} |f[x_1, x_2, \dots, x_k]| \right) x^{N-1} \quad \text{for } x \geqslant \max\{1, x_{N-1}\}.$$

But

$$f[x_1, x_2, \dots, x_k] = \frac{f^{(k-1)}(\xi_k)}{(k-1)!} \quad \text{for some } \xi_k \in (x_1, x_k),$$

which, upon invoking the growth condition (A.3) on $f^{(k-1)}(x)$, becomes

$$|f[x_1, x_2, \dots, x_k]| \leqslant D(k-1)^{-1-\delta}, \quad k = 2, 3, \dots,$$

Consequently, for all $x \geqslant \max\{1, x_{N-1}\}$ we have

$$|Q_N(x)| \leqslant \left(f(x_1) + D \left(\sum_{k=2}^{N} (k-1)^{-1-\delta} \right) \right) x^{N-1} < \left(f(x_1) + D \left(\sum_{k=1}^{\infty} k^{-1-\delta} \right) \right) x^{N-1}.$$

(Note that $\sum_{k=1}^{\infty} k^{-1-\delta}$ is a convergent series.) This completes the proof with

$$E = f(x_1) + D \left(\sum_{k=1}^{\infty} k^{-1-\delta} \right). \qquad \square$$

Note that Lemma A.3 applies to the function $f(x) = x^{\delta}$, $0 < \delta < 1$, with $x_1 \geqslant 1$, for which (A.3) holds, as can easily be verified.

We now turn to the treatment of the case $M > N$ in Eq. (14). Replacing j^2 and Z_j in these equations by x_j and $f(x_j)$, respectively, we now have the more general problem

$$\sum_{j=1}^{M} W_j \left[P_{NM}(x_j) - f(x_j) \right] x_j^{l-1} = 0, \quad l = 1, \dots, N, \tag{A.5}$$

where we have denoted $P_{NM}(x) = \sum_{k=1}^{N} \alpha_k^{(NM)} x^{k-1}$. This is to say, $P_{NM}(x)$ is a least-squares polynomial approximation to $f(x)$ in the sense that

$$\sum_{j=1}^{M} W_j \left[P_{NM}(x_j) - f(x_j) \right]^2 \leqslant \sum_{j=1}^{M} W_j \left[R(x_j) - f(x_j) \right]^2 \quad \text{for any } R(x) \in \pi_{N-1}, \tag{A.6}$$

where π_r denotes the set of polynomials of degree at most r. (With this notation we also have $P_{NN}(x) = Q_N(x)$.)

Lemma A.4 contains the main results concerning $P_{NM}(x)$.

Lemma A.4. *Let $f(x)$ and x_j be as in Lemma* A.2 *and let $P_{NM}(x)$ $(N < M)$ be the solution to* (A.5). *Then the following are true*:

(i) *There exist N points $y_1^{(NM)}, \ldots, y_N^{(NM)}$ in $[x_1, x_M]$ at which $P_{NM}(x)$ interpolates $f(x)$. As a result, $\alpha_1^{(NM)}, \alpha_2^{(NM)}, \ldots, \alpha_N^{(NM)}$ have the sign pattern $+, +, -, +, -, \ldots$.*

(ii) *$P_{NM}(x) > 0$ for all $x \geqslant 0$ when $N = 1, 2$ and $P_{NM}(x) > 0$ for $x \in [x_m, X)$ or $x \in [0, x_1]$ when $N = 4, 6, 8, \ldots$.*

(iii) *If, in addition,* (A.3) *holds for some $\delta > 0$, then*

(a) *$0 < P_{1M}(x) = f(y_1^{(1M)}) \leqslant f(x)$ for $x \geqslant x_M \geqslant y_1^{(1M)}$, and*

(b) *$|P_{NM}(x)| < f(x) + \widetilde{E}x^{N-1}$ for all $x \geqslant \max\{1, x_M\}$, where $\widetilde{E} > 0$ is a constant independent of M and N, and $N \geqslant 2$.*

Proof. Part (i) follows from Lemma A.5 that we state next and from part (i) of Lemma A.1. Part (ii) follows from Lemma A.2. Part (iii)(a) is an immediate consequence of $f(x)$ being positive and increasing for $x > 0$. To prove part (iii)(b) we proceed as in the proof of Lemma A.3. We first have that

$$\left|P_{NM}(x)\right| < f\left(y_1^{(NM)}\right) + \widetilde{E}x^{N-1} \quad \text{for all } x \geqslant \max\{1, x_M\},$$

where $\widetilde{E} = D(\sum_{k=1}^{\infty} k^{-1-\delta})$. The result now follows by $f(y_1^{(NM)}) \leqslant f(x_M) \leqslant f(x)$ for $x \geqslant x_M$. □

Lemma A.5. *Let $e(x)$ be continuous on $[a, b]$, and let $a \leqslant x_1 < \cdots < x_M \leqslant b$. If*

$$\sum_{j=1}^{M} W_j e(x_j) x_j^{l-1} = 0, \quad l = 1, \ldots, N,$$

for some $N < M$ and $W_j > 0$, $j = 1, \ldots, M$, then either $e(x_j) = 0$, $j = 1, \ldots, M$, or there exist N points $y_1, \ldots, y_N$ in $[x_1, x_M]$, for which $e(y_j) = 0$, $j = 1, \ldots, N$, and $e(x)$ changes sign at the y_j.

Proof. We start with the observation that

$$\sum_{j=1}^{M} W_j e(x_j) t(x_j) = 0 \quad \text{for any } t(x) \in \pi_{N-1}. \tag{A.7}$$

Letting $t(x) \equiv 1$ in (A.7), we conclude that either (i) $e(x_j) = 0$, $j = 1, \ldots, M$, or (ii) some of the $e(x_j)$ are positive and some are negative. In case (i) there is nothing to prove. In case (ii) we have that $e(x)$ must change sign on (x_1, x_M) at least once. We will show that $e(x)$ changes sign on (x_1, x_M) at least N times. Suppose it changes sign exactly r times, $r < N$, at the points $y_1, \ldots, y_r$. Then we can write $e(x) = q(x)\widehat{e}(x)$, where $q(x) = \prod_{i=1}^{r}(x - y_i)$ and $\widehat{e}(x)$ is of one sign on (x_1, x_M). As $r \leqslant N - 1$, we can let $t(x) = q(x)$ in (A.7) and thus obtain $\sum_{j=1}^{M} W_j \widehat{e}(x_j)[q(x_j)]^2 = 0$. This implies that $\widehat{e}(x_j)[q(x_j)]^2 = 0$, $j = 1, \ldots, M$, which, in turn, implies that $e(x_j) = 0$, $j = 1, \ldots, M$, contrary to our assumption that not all $e(x_j)$ are zero. Therefore, we must have $r \geqslant N$. The rest of the proof is easy and is left to the reader. □

Acknowledgements

This work was partly supported by the Fund for the Promotion of Research at the Technion.

References

[1] P.E. Barbone, A. Cherukuri, D. Goldman, Canonical representations of complex vibratory subsystems, Internat. J. Solids Structures, to appear.
[2] A. Bayliss, E. Turkel, Radiation boundary conditions for wave-like equations, Comm. Pure Appl. Math. 33 (1980) 707–725.
[3] L. Demkowicz, F. Ihlenburg, Analysis of coupled finite-infinite element method for exterior Helmholtz problems, TICAM Report 96-52, University of Texas, Austin, Nov. 1996.
[4] B. Engquist, A. Majda, Radiation boundary conditions for acoustic and elastic calculations, Comm. Pure Appl. Math. 32 (1979) 313–357.
[5] L. Ferm, B. Gustafsson, A down-stream boundary procedure for the Euler equations, Comput. Fluids 10 (1982) 261–276.
[6] D. Givoli, Nonreflecting boundary conditions, J. Comput. Phys. 94 (1991) 1–29.
[7] D. Givoli, Numerical Methods for Problems in Infinite Domains, Elsevier, Amsterdam, 1992.
[8] D. Givoli, A spatially exact non-reflecting boundary condition for time dependent problems, Comput. Methods Appl. Mech. Engrg. 95 (1992) 97–113.
[9] D. Givoli, D. Cohen, Non-reflecting boundary conditions based on Kirchhoff-type formulae, J. Comput. Phys. 117 (1995) 102–113.
[10] D. Givoli, J.B. Keller, A finite element method for large domains, Comput. Methods Appl. Mech. Engrg. 76 (1989) 41–66.
[11] D. Givoli, J.B. Keller, Non-reflecting boundary conditions for elastic waves, Wave Motion 12 (1990) 261–279.
[12] D. Givoli, J.B. Keller, Special finite elements for use with high-order boundary conditions, Comput. Methods Appl. Mech. Engrg. 119 (1994) 199–213.
[13] D. Givoli, I. Patlashenko, Finite element solution of nonlinear time-dependent exterior wave problems, J. Comput. Phys. 143 241–258 (1998); Commun. Numer. Methods Engrg. 12 (1996) 257–267.
[14] D. Givoli, I. Patlashenko, Finite element schemes for nonlinear problems in infinite domains, Internat. J. Numer. Methods Engrg. 42 (1998) 341–360.
[15] D. Givoli, I. Patlashenko, Optimal local non-reflecting boundary conditions, Appl. Numer. Math. 27 (1998) 367–384.
[16] D. Givoli, I. Patlashenko, J.B. Keller, High order boundary conditions and finite elements for infinite domains, Comput. Methods Appl. Mech. Engrg. 143 (1997) 13–39.
[17] D. Givoli, S. Vigdergauz, Artificial boundary conditions for 2D problems in geophysics, Comput. Methods Appl. Mech. Engrg. 110 (1993) 87–101.
[18] M.J. Grote, J.B. Keller, Nonreflecting boundary conditions for time dependent scattering, J. Comput. Phys. 127 (1996) 52–65.
[19] T. Hagstrom, Conditions at the downstream boundary for simulations of viscous, incompressible flow, SIAM J. Sci. Statist. Comput. 12 (1991) 843–858.
[20] T. Hagstrom and H.B. Keller, Exact boundary conditions at an artificial boundary for partial differential equations in cylinders, SIAM J. Math. Anal. 17 (1986) 322–341.
[21] C. Johnson, Numerical Solutions of Partial Differential Equations by the Finite Element Method, Cambridge University Press, Cambridge, 1987.
[22] J.B. Keller and D. Givoli, Exact non-reflecting boundary conditions, J. Comput. Phys. 82 (1989) 172–192.

[23] I. Patlashenko, D. Givoli, Nonlocal and local artificial boundary conditions for two-dimensional flow in an infinite channel, Internat. J. Numer. Methods Heat Fluid Flow 6 (1996) 47–62.
[24] I. Patlashenko, D. Givoli, Non-reflecting finite element schemes for three-dimensional acoustic waves, J. Comput. Acoustics 5 (1997) 95–115.
[25] I. Patlashenko, D. Givoli, A numerical method for problems in infinite strips with irregularities extending to infinity, Numer. Methods Partial Differential Equations 14 (1998) 233–249.
[26] A. Sidi, A family of matrix problems: solution, Problems and Solutions, SIAM Rev., to appear. Solution to: D. Givoli, A family of matrix problems, Problems and Solutions, SIAM Rev. 39 (1997) 514.
[27] G. Strang, G. J. Fix, An Analysis of the Finite Element Method, Prentice-Hall, Englewood Cliffs, NJ, 1973.
[28] L. Ting, M.J. Miksis, Exact boundary conditions for scattering problems, J. Acoust. Soc. Amer. 80 (1986) 1825–1827.
[29] S.V. Tsynkov, E. Turkel, S. Abarbanel, External flow computations using global boundary conditions, AIAA J. 34 (1996) 700–706.