EPFL

Master in Computational Science and Engineering

Mathematics Institute of Computational Science and Engineering

Finite elements-based Padé approximants for Helmholtz frequency response problems

Author:
Davide Pradovera

Supervisor:
Prof. Fabio Nobile

Academic Year 2016-2017

Autumn semester

Abstract

The focus of this project is the application of rational functions in the approximation of the solution map of a parametric Helmholtz problem with homogeneous Dirichlet boundary conditions. Such an approximation can be applied, for instance, in frequency response problems, where one wants to understand how the solution of the Helmholtz problem changes with respect to the wavenumber.

In this report it is proven that the Helmholtz equation, endowed with Dirichlet, Neumann or mixed Dirichlet-Neumann conditions, is meromorphic in $\mathbb{C}$. Moreover, the regularity of the scattering problem, i.e. the Helmholtz equation coupled with Bohr-Sommerfield radiation conditions, is discussed in some detail. The hypothesis that the solution map of such a problem may still be meromorphic is formulated after the analysis of a simple example.

A rational Padé approximant, which relies on the solution of an optimization problem, is defined for Hilbert space-valued meromorphic functions. Several original algorithms, based on slight variations of the original definition, are described in detail. Numerical tests are used to compare the different algorithms.

Contents

1 Introduction

Many engineering applications require the computation of a quantity of interest not only for a certain set of physical parameters, but on a whole range of their values. The simplest solution to this problem is the direct evaluation of the quantity over a grid of values, followed by some kind of interpolation to achieve an approximation over a continuous domain.

The main issue with this approach is the need to repeat the computation of the target function on each grid point. If the evaluation function is expensive, e.g. because it is a functional of the solution of a PDE, the high computational cost makes the "parameter sweep" method very impractical. This is particularly true for the specific case of frequency response problems concerning the Helmholtz equation, due to the presence of numerical effects, which require very fine meshes or high polynomial interpolation degrees (see e.g. [11]).

For smooth functions, projections on low-dimension subspaces (e.g. through truncated Taylor series) usually prove to be quite effective. Still, this approach cannot be easily generalized to the case of non-smooth functions. In particular, in the case of meromorphic functions (e.g. the solution map of the Helmholtz equation) the presence of poles leads to approximations which have a very small convergence region in the complex plane. In particular, a reasonable approximant of such a function should be able to predict correctly (at least locally) the position of its poles.

A very powerful tool which satisfies this requirement is *Padé approximation.* Indeed, by using rational approximants, Padé approximants are able to provide an arbitrarily close approximation of a given meromorphic function on any (compact) subset of its domain. The only requirement is that the approximant must have enough degrees of freedom (see e.g. [4], Theorem 6.2.2).

These results are true for univariate and multivariate (see e.g. [5]) complex-valued functions, and are supported by a vast literature. The extension to vector-valued or Hilbert space-valued functions is much scarcer. The possible reason for this is described in [6], where it is suggested not to apply Padé approximants to approximate the target function, but only to find its poles.

Regardless of such claims, this report follows [2] by defining rational approximants for Hilbert space-valued meromorphic functions, and by devising algorithms for their computation, with a particular focus on frequency response problems for the Helmholtz equation.

The outline of this report is the following.

Section 2 provides a preliminary analysis of the Helmholtz equation, endowed with Dirichlet, Neumann or mixed Dirichlet-Neumann conditions. The main result of this section is the proof that the solution map of such a problem is meromorphic in $\mathbb{C}$. A secondary result is a recursive formula for the computation of the derivatives of the solution map.

Section 3 explores the possibility to extend such results to the case of Bohr-Sommerfield radiation conditions on a portion of the boundary. Since the theory cannot provide a definitive answer, a simple example is analysed thoroughly.

In Section 4, a definition of Padé approximant for Hilbert space-valued meromorphic functions is provided, followed by the description and the analysis of two algorithms for its computation.

Section 5 aims at defining a multi-point extension of the Padé approximant, strongly related to Hermite interpolation. It follows the description and a brief analysis of an algorithm for the computation of such an approximation.

Section 6 contains some numerical examples regarding the results of the previous sections. Some of the code segments (in FreeFem++) used to obtain these results are reported in Section 8.

2 Problem setting

Given an open bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ (for $d = 2, 3$) and a complex value z, we consider the Helmholtz problem with homogeneous Dirichlet boundary conditions

$$\begin{cases} -\Delta u - zu = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases} \tag{1}$$

where f is a sufficiently smooth function on Ω, e.g. $f \in H^{-1}(\Omega)$. The solution u is sought in the complex-valued Hilbert space $H_0^1(\Omega)$, which in the rest of this report is indicated as V, unless specified otherwise.

The corresponding weak formulation is given by:

$$\text{find } u \in V \text{ s.t. } a_z(u, v) = F(v) \quad \forall v \in V \tag{2}$$

where

$$a_z(u, v) := \langle \nabla u, \nabla v \rangle_{L^2(\Omega)} - z \langle u, v \rangle_{L^2(\Omega)} \quad \text{and} \quad F(v) := \langle f, v \rangle \qquad \text{for } u, v \in V$$

In the expressions above, $\langle \cdot, \cdot \rangle_{L^2(\Omega)}$ represents the usual scalar product on $L^2(\Omega)$ and $\langle \cdot, \cdot \rangle$ is the result of the duality between an element of V and an element of its dual space.

If we assume that problem (2) is well-posed for all $z \in U \subset \mathbb{C}$, then there exists a map $\mathcal{S} : U \to V$ which associates the parameter z with the corrisponding solution of problem (2), which is indicated with the notation $\mathcal{S}(z)$.

Given a positive real parameter w, we consider the following scalar product on V:

$$\langle u, v \rangle_{V,w} := \langle \nabla u, \nabla v \rangle_{L^2(\Omega)} + w^2 \langle u, v \rangle_{L^2(\Omega)} \qquad \text{for } u, v \in V$$

Observe that for a given $w > 0$ the induced norm $\|u\|_{V,w} = \langle u, u \rangle_{V,w}^{1/2}$ is equivalent to the usual norm on $H^1(\Omega)$:

$$\sqrt{\min\{1, w^2\}} \|u\|_{H^1(\Omega)} \leq \|u\|_{V,w} \leq \sqrt{\max\{1, w^2\}} \|u\|_{H^1(\Omega)}$$

In the rest of this report we consider $w > 0$ fixed.

Remark 2.1. *The analysis carried out in the current and following sections can be generalized to the cases of mixed (homogeneous) Dirichlet-Neumann conditions*

$$\begin{cases} \partial_n u = 0 & \text{on } \Gamma_D \\ u = 0 & \text{on } \Gamma_N \end{cases} \qquad \text{with } \overline{\Gamma}_D \cup \overline{\Gamma}_N = \partial\Omega \text{ and } \Gamma_D \cap \Gamma_N = \emptyset$$

and (homogeneous) Neumann conditions

$$\partial_n u = 0 \quad \text{on } \partial\Omega$$

However, in this last case, one must be aware that problem (1) is ill-posed for $z = 0$, since this value belongs to the spectrum of the Laplacian operator (see e.g. [8]).

2.1 Well-posedness

In this section we prove that $\mathcal{S}$ is well defined on $\mathbb{C} \setminus \Lambda$, where $\Lambda \subset \mathbb{R}^+$ is the *discrete* spectrum of the Laplacian operator on Ω with homogeneous Dirichlet conditions.

Theorem 2.1. *Problem (2) is well-posed for $z \in \mathbb{C} \setminus \mathbb{R}^+ = \{z \in \mathbb{C} \text{ s.t. } \Im(z) \neq 0 \vee \Re(z) \leq 0\}$.*

Proof. If $z = 0$, (1) becomes a Laplace problem, and classical results can be used to show well-posedness (see e.g. [17], Chapter 3). In the case $z \neq 0$ we follow [2], Section 2.

First observe that, for any $\nu \in \mathbb{C}$ and for $0 \leq \varepsilon \leq 1$, we have

$$|\nu| \geq \frac{|\Re(\nu)| + |\Im(\nu)|}{\sqrt{2}} \geq \frac{\varepsilon\Re(\nu) + |\Im(\nu)|}{\sqrt{2}} \tag{3}$$

Consider the sequilinear form $a_z(u, v)$. By applying (3) we deduce

$$\begin{aligned}|a_z(u,u)| &\geq \frac{\varepsilon}{\sqrt{2}}\Re(a_z(u,u)) + \frac{1}{\sqrt{2}}|\Im(a_z(u,u))| \\ &= \frac{\varepsilon}{\sqrt{2}}(\|\nabla u\|^2_{L^2(\Omega)} - \Re(z)\|u\|^2_{L^2(\Omega)}) + \frac{1}{\sqrt{2}}|\Im(z)|\|u\|^2_{L^2(\Omega)} \\ &= \frac{\varepsilon}{\sqrt{2}}\|\nabla u\|^2_{L^2(\Omega)} + \left(\frac{|\Im(z)| - \varepsilon\Re(z)}{\sqrt{2}}\right)\|u\|^2_{L^2(\Omega)} \\ &\geq \frac{1}{\sqrt{2}}\min\left\{\varepsilon, \frac{|\Im(z)| - \varepsilon\Re(z)}{w^2}\right\}\|\nabla u\|^2_{V,w}\end{aligned}$$

for any $u \in V$, provided $0 \leq \varepsilon \leq 1$.

Now we consider three cases:

(a) If $\Re(z) < 0$, take $\varepsilon = 1$ and define $\alpha := \min\left\{1, \frac{|\Im(z)| - \Re(z)}{w^2}\right\} > 0$.

(b) If $\Re(z) = 0$, take $\varepsilon = 1$ and define $\alpha := \min\left\{1, \frac{|\Im(z)|}{w^2}\right\}$. Observe that $\alpha > 0$ since $\Im(z) \neq 0$.

(c) If $\Re(z) > 0$, take $\varepsilon = \min\left\{1, \frac{|\Im(z)|}{2\Re(z)}\right\}$ and define $\alpha := \min\left\{\varepsilon, \frac{|\Im(z)| - \varepsilon\Re(z)}{w^2}\right\}$. Observe that $\alpha > 0$ since $|\Im(z)| - \varepsilon\Re(z) > 0$.

In all cases we have proven that the sequilinear form a_z is coercive on V, with constant $\frac{\alpha}{\sqrt{2}}$. Moreover, from the triangular and Cauchy-Schwartz inequalities[1] we have

$$\begin{aligned}|a_z(u,v)| &\leq \|\nabla u\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)} + |z|\|u\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)} \\ &\leq \left(\|\nabla u\|^2_{L^2(\Omega)} + |z|\|u\|^2_{L^2(\Omega)}\right)^{1/2}\left(\|\nabla v\|^2_{L^2(\Omega)} + |z|\|v\|^2_{L^2(\Omega)}\right)^{1/2} \\ &\leq \max\left\{1, \frac{|z|}{w^2}\right\}\|u\|_{V,w}\|v\|_{V,w}\end{aligned}$$

for any $u, v \in V$. Hence the sequilinear form a_z is continuous on V, with constant $\max\left\{1, \frac{|z|}{w^2}\right\}$. Since $f \in H^{-1}(\Omega) = V'$, the well-posedness of (2) follows from the Lax-Milgram Lemma. □

The main interest in problem (1) is motivated by the behavior of $\mathcal{S}$ for *real* (positive) values of z, which is related to the phenomenon of *resonance*. In particular, the following Theorem states that the (possible) resonant frequencies depend only on the geometry of the problem.

Theorem 2.2. *Problem (2) is well-posed for $z \in \mathbb{R}^+ \setminus \Lambda$, where Λ is the spectrum of the Laplacian operator on Ω with homogeneous Dirichlet boundary conditions, which depends only on Ω.*

Proof. Let $z \in \mathbb{R}^+$. By following the same derivation used in Theorem 2.1, we can prove that a_z is continuous. Also, the sequilinear form a_z satisfies a Gårding inequality:

$$a_z(u,u) + \theta\|u\|^2_{L^2(\Omega)} \geq \gamma\|u\|^2_{V,w}$$

where $\theta = w^2 + z$ and $\gamma = 1$.

[1] In particular, the second step of the derivation follows from the Cauchy-Schwartz inequality in $\mathbb{R}^2$, i.e.

$$(ab + cd)^2 \leq (a^2 + c^2)(b^2 + d^2) \qquad \forall a, b, c, d \in \mathbb{R}$$

Hence we can apply the *Fredholm alternative* (see e.g. [17], Theorem 6.66): it is enough to check the well-posedness of the homogeneous problem

$$\text{find } u_0 \in V \text{ s.t. } a_z(u,v) = 0 \quad \forall v \in V$$

which admits a unique solution in V if $z \in \mathbb{R}^+ \setminus \Lambda$. □

Classical results (see e.g. [12], Theorem 6.26) show that the eigenfunctions of the Laplacian operator form a basis of $L^2(\Omega)$. In particular, we may choose a basis $\{\varphi_l\}_{l\geq 1} \subset H_0^1(\Omega)$ which is orthonormal with respect to the $L^2(\Omega)$-norm and satisfies

$$\begin{cases} -\Delta\varphi_l - \lambda_l\varphi_l = 0 & \text{in } \Omega \\ \varphi_l = 0 & \text{on } \partial\Omega \end{cases} \qquad \text{for } l = 1, 2, \dots \tag{4}$$

If $z \in \mathbb{C} \setminus \Lambda$, we can expand the solution map and the right hand side as

$$\mathcal{S}(z) = \sum_l u_l(z)\varphi_l \quad \text{and} \quad f = \sum_l f_l\varphi_l$$

and plug them into equation (1). This yields

$$\mathcal{S}(z) = \sum_l \frac{f_l}{\lambda_l - z}\varphi_l \tag{5}$$

From here we can prove a bound on the norm of the solution map.

Theorem 2.3. *If $z \in \mathbb{C} \setminus \Lambda$, the (unique) solution of (2) satisfies the a priori bound*

$$\|\mathcal{S}(z)\|_{V,w} \leq \frac{\sqrt{\alpha(z) + |\Re(z)| + w^2}}{\alpha(z)}\|f\|_{L^2(\Omega)} \tag{6}$$

where

$$\alpha(z) = \min_{\lambda\in\Lambda} |z - \lambda| \tag{7}$$

Proof. The bound is an improvement of the one shown in [2], whose proof is taken as reference for the following derivation.

First observe that the eigenvalue-eigenfunction relation (4) implies the orthogonality of the eigenfunction basis with respect to the $H_0^1(\Omega)$ scalar product:

$$\langle \nabla\varphi_l, \nabla\varphi_k\rangle_{L^2(\Omega)} = \langle -\Delta\varphi_l, \varphi_k\rangle_{L^2(\Omega)} = \lambda_l\langle\varphi_l, \varphi_k\rangle_{L^2(\Omega)} = \lambda_l\delta_{lk}$$

Thus, by exploiting (5), we have

$$\begin{aligned} \|\mathcal{S}(z)\|^2_{V,w} &= \|\nabla\mathcal{S}(z)\|^2_{L^2(\Omega)} + w^2\|\mathcal{S}(z)\|^2_{L^2(\Omega)} \\ &= \sum_l \frac{|f_l|^2}{|\lambda_l - z|^2}\left(\|\nabla\varphi_l\|^2_{L^2(\Omega)} + w^2\|\varphi_l\|^2_{L^2(\Omega)}\right) \\ &= \sum_l \frac{\lambda_l + w^2}{|\lambda_l - z|^2}|f_l|^2 \end{aligned}$$

Now we want to bound the term $\frac{\lambda_l + w^2}{|\lambda_l - z|^2}$ uniformly with respect to l: using the fact that $\Lambda \subset \mathbb{R}$, we have

$$\frac{\lambda_l + w^2}{|\lambda_l - z|^2} = \frac{\lambda_l - \Re(z) + \Re(z) + w^2}{|\lambda_l - z|^2} \leq \frac{|\lambda_l - \Re(z)| + |\Re(z)| + w^2}{|\lambda_l - z|^2} \leq \frac{|\lambda_l - z| + |\Re(z)| + w^2}{|\lambda_l - z|^2}$$

Let $g(x) = \frac{x + |\Re(z)| + w^2}{x^2} = \frac{1}{x} + \frac{|\Re(z)| + w^2}{x^2}$ for $x \in \mathbb{R}^+$. Since $g(x)$ is strictly decreasing over $\mathbb{R}^+$,

$$\arg\sup_{x\in B} g(x) = \inf B \qquad \text{for any } B \subset \mathbb{R}^+$$

If $B = \{|\lambda_l - z|\}_{l \geq 1}$, the minimum of the set coincides with $\alpha(z)$, defined as in (7). Hence

$$\|\mathcal{S}(z)\|_{V,w}^2 \leq g(\alpha(z)) \sum_l |f_l|^2 = \frac{\alpha(z) + |\Re(z)| + w^2}{\alpha(z)^2} \|f\|_{L^2(\Omega)}^2$$

□

A numerical check of bound (6) is shown in Section 6.1.

2.2 Regularity of the solution map $\mathcal{S}(z)$

In this section we prove that the solution map $\mathcal{S} : \mathbb{C} \setminus \Lambda \to V$ is holomorphic in its domain, and meromorphic in $\mathbb{C}$ (see Definition 2.1).

Theorem 2.4. *The solution map $\mathcal{S} : \mathbb{C} \setminus \Lambda \to V$ is continuous with respect to the (weighted) V-norm.*

Proof. For any $z \in \mathbb{C} \setminus \Lambda$, consider an arbitrarily small $h \in \mathbb{C}$. In particular, we assume that $|h| < \alpha(z)$, with α defined as in (7).

Since $S(z+h)$ and $\mathcal{S}(z)$ are solutions of (2) with parameter $z+h$ and z respectively, we have

$$a_{z+h}(S(z+h), v) = \langle f, v \rangle = a_z(\mathcal{S}(z), v) \qquad \forall v \in V$$

If we define $w_h(z) := S(z+h) - S(h)$, this implies

$$\langle \nabla w_h(z), \nabla v \rangle_{L^2(\Omega)} - (z+h)\langle w_h(z), v \rangle_{L^2(\Omega)} = h \langle \mathcal{S}(z), v \rangle_{L^2(\Omega)} \qquad \forall v \in V$$

Since $z + h \notin \Lambda$, we can apply Theorem 2.3:

$$\|w_h(z)\|_{V,w} \leq \frac{\sqrt{\alpha(z+h) + |\Re(z+h)| + w^2}}{\alpha(z+h)} |h| \|\mathcal{S}(z)\|_{L^2(\Omega)} \tag{8}$$

where $\alpha(\cdot) = \min_{\lambda \in \Lambda} |\cdot - \lambda|$.

Let $g(\cdot) = \frac{\sqrt{\alpha(\cdot) + |\Re(\cdot)| + w^2}}{\alpha(\cdot)}$. Since α is a continuous function, g is continuous as well. In particular

$$\lim_{h \to 0} g(z+h) = g(z) < \infty$$

since $z \notin \Lambda$.

Moreover, since $f \in H^{-1}(\Omega)$, we can apply Theorem 2.3 to see that $\|\mathcal{S}(z)\|_{V,w} < \infty$.

Hence (8) implies that $\lim_{h \to 0} \|\mathcal{S}(z+h) - \mathcal{S}(z)\|_{V,w} = \lim_{h \to 0} \|w_h(z)\|_{V,w} = 0$. □

Theorem 2.5. *The solution map $\mathcal{S} : \mathbb{C} \setminus \Lambda \to V$ admits complex derivative $\mathcal{S}_1 : \mathbb{C} \setminus \Lambda \to V$, which is the solution of*

$$a_z(\mathcal{S}_1(z), v) = \langle \mathcal{S}(z), v \rangle_{L^2(\Omega)} \qquad \forall v \in V \tag{9}$$

Proof. For any $z \in \mathbb{C} \setminus \Lambda$, consider an arbitrarily small $h \in \mathbb{C}$. In particular, we assume that $|h| < \alpha(z)$, with α defined as in (7).

Similarly to the proof of Theorem 2.4, we have

$$a_{z+h}(S(z+h), v) = \langle f, v \rangle = a_z(\mathcal{S}(z), v) \qquad \forall v \in V$$

If we define $\widetilde{w}_h(z) := \frac{S(z+h) - S(h)}{h}$, this implies

$$\langle \nabla \widetilde{w}_h(z), \nabla v \rangle_{L^2(\Omega)} - z \langle \widetilde{w}_h(z), v \rangle_{L^2(\Omega)} = \langle \mathcal{S}(z+h), v \rangle_{L^2(\Omega)} \qquad \forall v \in V$$

By taking the limit for $h \to 0$, we get

$$a_z\left(\lim_{h \to 0} \widetilde{w}_h(z), v\right) = \langle \mathcal{S}(z), v \rangle_{L^2(\Omega)} \qquad \forall v \in V$$

where we have used the continuity of the $L^2(\Omega)$ scalar product, of a_z and of $\mathcal{S}$ (see Theorems 2.1 and 2.4).

Since $\lim_{h\to 0} \widetilde{w}_h(z)$ coincides with the definition of the complex derivative $\mathcal{S}_1(z)$, we have proven (9). Since $\mathcal{S}(z) \in V \subset H^{-1}(\Omega)$, Theorems 2.1 and 2.2 imply that $\mathcal{S}_1(z)$ exists unique. □

Remark 2.2. *A consequence of Theorem 2.5 is that $\mathcal{S} \in \mathcal{H}(\mathbb{C} \setminus \Lambda; V)$. A little computation (similar to the one carried out in the proof above) shows that the α-th Taylor coefficient of $\mathcal{S}$ in $z \in \mathbb{C} \setminus \Lambda$ admits a recursive formulation:*

$$a_z(\mathcal{S}_\alpha(z), v) = \langle \mathcal{S}_{\alpha-1}(z), v\rangle_{L^2(\Omega)} \quad \forall v \in V \qquad \text{for } \alpha = 1, 2, \ldots$$

where $\mathcal{S}_0(z) := \mathcal{S}(z)$.

Since the multiplicity of each eigenvalue of the Laplacian operator is finite (see e.g. [12], Theorem 6.26), expression (5) can be rewritten as

$$\mathcal{S}(z) = \sum_k \frac{c_k}{\bar{\lambda}_k - z} \tag{10}$$

where $\{\bar{\lambda}_k\}_{k\geq 1}$ are the distinct elements of Λ, with multiplicities $\{m_k\}_{k\geq 1}$, and

$$c_k := \sum_{j=1}^{m_k} f_{k,j}\varphi_{k,j} = \sum_{j=1}^{m_k} \langle f, \varphi_{k,j}\rangle_{L^2(\Omega)}\varphi_{k,j} \qquad \text{for } k = 1, 2, \ldots$$

with $\{\varphi_{k,j}\}_{j=1}^{m_k}$ being the ($L^2(\Omega)$-orthonormal) eigenfunctions corresponding to $\bar{\lambda}_k$.

Hence each eigenvalue is a simple pole of $\mathcal{S}(z)$, and we can deduce that $\mathcal{S}$ is *meromorphic* in $\mathbb{C}$, according to the following definition (see e.g. [2]):

Definition 2.1. *A function $\mathcal{T} : \mathbb{C} \supset U \to V$ is meromorphic in U, and we write $\mathcal{T} \in \mathcal{M}(U; V)$, if:*

(a) there exists a discrete subset Λ of U such that $\mathcal{T} \in \mathcal{H}(U \setminus \Lambda; V)$;

(b) for each $\nu \in \Lambda$, there exists $k \in \mathbb{N}$ such that $(z - \nu)^k\mathcal{T}(z)$ admits holomorphic extension in ν.

3 Extension to scattering problems

Another family of problems similar to the one described in the previous section is wave scattering, which is formulated as follows. Given a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ (for $d = 2, 3$), whose boundary is partitioned into Γ_D and Γ_R, for any $k \in \mathbb{C}$ we consider the problem (see e.g. [14])

$$\begin{cases} -\Delta u - k^2 u = f & \text{in } \Omega \\ u = 0 & \text{on } \Gamma_D \\ \partial_n u = iku & \text{on } \Gamma_R \end{cases} \tag{11}$$

where f is a sufficiently smooth function on Ω, e.g. $f \in H^{-1}(\Omega)$.

A weak formulation of problem (11) is the following:

$$\text{find } u \in V_{\Gamma_D} \text{ s.t. } b_k(u,v) = \langle f, v\rangle \quad \forall v \in V_{\Gamma_D} \tag{12}$$

where

$$b_k(u,v) := \langle \nabla u, \nabla v\rangle_{L^2(\Omega)} - k^2 \langle u, v\rangle_{L^2(\Omega)} - ik\langle u, v\rangle_{L^2(\Gamma_D)} \quad \text{for } u, v \in V_{\Gamma_D}$$

and the space V_{Γ_D} is defined as $V_{\Gamma_D} = \{v \in H^1(\Omega), v|_{\Gamma_D} = 0\}$.

In this section we want to approach the following question: is the map which associates the wavenumber $k \in \mathbb{C}$ to the solution of problem (12) meromorphic?

3.1 Well-posedness

The main result regarding the well-posedness of problem (12) is the following.

Theorem 3.1. *Problem (12) is well-posed for any $k \in \mathbb{C}$, with $\Im(k) \geq 0$.*

Proof. For $k = 0$ we may apply the standard analysis for Laplace problems with mixed boundary conditions to prove well-posedness.

Let $k \neq 0$. Similarly to the proof of Theorem 2.2, we first check that the Fredholm alternative can be applied to prove the well-posedness of problem (12).

By applying the trace and Cauchy-Schwartz inequality, we can prove that b_k is continuous:

$$\begin{aligned} |b_k(u,v)| &\leq \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} + |k^2| \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} + |k| \|u\|_{L^2(\Gamma_R)} \|v\|_{L^2(\Gamma_R)} \\ &\leq \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} + |k^2| \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} + C_{tr}^2 |k| \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \\ &\leq \Big(\|\nabla u\|^2_{L^2(\Omega)} + \left(|k^2| + C_{tr}^2 |k|\right) \|u\|^2_{L^2(\Omega)} \Big)^{1/2} \Big(\|\nabla v\|^2_{L^2(\Omega)} + \left(|k^2| + C_{tr}^2 |k|\right) \|v\|^2_{L^2(\Omega)} \Big)^{1/2} \\ &\leq \max\left\{1, \frac{|k^2| + C_{tr}^2 |k|}{w^2}\right\} \|u\|_{V_{\Gamma_D},w} \|v\|_{V_{\Gamma_D},w} \qquad \forall u, v \in V_{\Gamma_D} \end{aligned}$$

The sequilinear form b_k satisfies a Gårding inequality:

$$\begin{aligned} \Re(b_k(v,v)) &= \|\nabla v\|^2_{L^2(\Omega)} - \Re(k^2) \|v\|^2_{L^2(\Omega)} + \Im(k) \|v\|^2_{L^2(\Gamma_R)} \\ &= \|v\|^2_{V_{\Gamma_D},w} - \left(w^2 + \Re(k^2)\right) \|v\|^2_{L^2(\Omega)} + \Im(k) \|v\|^2_{L^2(\Gamma_R)} \\ &\geq \gamma \|v\|^2_{V_{\Gamma_D},w} - \theta \|v\|^2_{L^2(\Omega)} \end{aligned}$$

with $\theta = w^2 + \Re(k^2)$ and $\gamma = 1$.

Hence we can apply the Fredhold alternative (see e.g. [3], Proposition 1.8). Assume that there exists a non-zero solution to the homogeneous problem

$$\text{find } u_0 \in V_{\Gamma_D} \text{ s.t. } b_k(u_0, v) = 0 \quad \forall v \in V_{\Gamma_D} \tag{13}$$

By plugging $v = u_0$ into (13) and taking the real and imaginary part, we obtain

$$\begin{cases} 0 = \Re(b_k(u_0,u_0)) = \|\nabla u_0\|^2_{L^2(\Omega)} - \Re(k^2)\|u_0\|^2_{L^2(\Omega)} + \Im(k)\|u_0\|^2_{L^2(\Gamma_R)} \\ 0 = \Im(b_k(u_0,u_0)) = -\Im(k^2)\|u_0\|^2_{L^2(\Omega)} - \Re(k)\|u_0\|^2_{L^2(\Gamma_R)} \end{cases}$$

Since $\Re(k^2) = \Re(k)^2 - \Im(k)^2$ and $\Im(k^2) = 2\Re(k)\Im(k)$, these conditions can be written equivalently as

$$\begin{cases} \|\nabla u_0\|^2_{L^2(\Omega)} + \left(\Im(k)^2 - \Re(k)^2\right)\|u_0\|^2_{L^2(\Omega)} + \Im(k)\|u_0\|^2_{L^2(\Gamma_R)} = 0 \\ 2\Re(k)\Im(k)\|u_0\|^2_{L^2(\Omega)} + \Re(k)\|u_0\|^2_{L^2(\Gamma_R)} = 0 \end{cases} \tag{14}$$

Now we have three cases:

(a) If $\Re(k) = 0$, the first equation in (14) becomes

$$\|\nabla u_0\|^2_{L^2(\Omega)} + \Im(k)^2\|u_0\|^2_{L^2(\Omega)} + \Im(k)\|u_0\|^2_{L^2(\Gamma_R)} = 0$$

Since we have assumed $k \neq 0$, we have that $\Im(k) > 0$ and

$$\|\nabla u_0\|_{L^2(\Omega)} = \|u_0\|_{L^2(\Omega)} = \|u_0\|_{L^2(\Gamma_R)} = 0$$

which implies $u_0 = 0$.

(b) If $\Re(k) \neq 0$ and $\Im(k) > 0$, the second equation in (14) becomes

$$2\Im(k)\|u_0\|^2_{L^2(\Omega)} + \|u_0\|^2_{L^2(\Gamma_R)} = 0$$

which implies

$$\|u_0\|_{L^2(\Omega)} = \|u_0\|_{L^2(\Gamma_R)} = 0$$

and $u_0 = 0$.

(c) If $\Re(k) \neq 0$ and $\Im(k) = 0$, the second equation in (14) becomes

$$\|u_0\|^2_{L^2(\Gamma_R)} = 0$$

which implies $u_0 \in H^1_0(\Omega)$. Thus u_0 satisfies

$$\langle \nabla u_0, \nabla v\rangle_{L^2(\Omega)} - k^2\langle u_0, v\rangle_{L^2(\Omega)} = 0 \quad \forall v \in V_{\Gamma_D}$$

Now consider a bounded domain $\Omega^* \supset \Omega$ such that $\Gamma_D \subset \partial\Omega^*$ and $\Gamma_R \subset \Omega^*$. Let $u_0^* \in L^2(\Omega^*)$ denote the extension of u_0 by zero on $\Omega^* \setminus \Omega$.

Observe that $\partial_n u_0 = iku_0 = 0$ on Γ_R, which implies that $u_0 \in \left\{v \in H^1_0(\Omega), \partial_n v = 0 \text{ on } \Gamma_R\right\}$. Hence u_0^* belongs to $\left\{v \in H^1(\Omega^*), v|_{\Gamma_D} = 0\right\}$, since the necessary compatibility conditions on Γ_R are satisfied.

Also, we have

$$\langle \nabla u_0^*, \nabla v\rangle_{L^2(\Omega^*)} - k^2\langle u_0^*, v\rangle_{L^2(\Omega^*)} = 0 \qquad \forall v \in H^1(\Omega^*), v|_{\Gamma_D} = 0$$

Elliptic regularity (see e.g. [17], Theorem 8.27) implies that $u_0^* \in H^2(Q)$ for any compact set $Q \subset \Omega^*$, in particular in an open Ω^*-neighborhood of Γ_R. From the unique continuation principle (see e.g. [18]) we can conclude that $u_0^* = 0$ on Ω^* and $u_0 = 0$ on Ω.

Since this is a contradiction, the Fredholm alternative allows us to conclude that problem (11) is well-posed. □

At first glance, the presence of Dirichlet conditions in problem (11) may lead to believe that the problem is ill-posed at some *real* (resonant) frequencies. However, Theorem 3.1 proves that the phenomenon of resonance does not appear for any real choice of k.

The well-posedness of the Helmholtz problem cannot be easily extended to the case $\Im(k) < 0$. The first portion of the proof above can be still applied, and the Fredholm alternative still holds. However the analysis of the homogeneous problem (13) is very non-trivial for $\Im(k) < 0$.

We show in the next section the analysis of a very simplified problem, for which there exists a discrete set of values of k which make the problem ill-posed.

3.2 A scattering problem on a square

Given $k \in \mathbb{C}$ and $L > 0$, consider the homogeneous version of problem (11) on a square domain:

$$\begin{cases} -\Delta u - k^2 u = 0 & \text{in } \Omega := (0,L)^2 \\ u = 0 & \text{on } \Gamma_D := (0,L) \times \{0, L\} \\ \partial_n u = iku & \text{on } \Gamma_R := \{0,L\} \times (0,L) \end{cases} \tag{15}$$

We want to find a non-trivial solution of the form $u(x,y) = X(x)Y(y)$. From the Helmholtz equation we get

$$-\frac{X''(x)}{X(x)} - \frac{Y''(y)}{Y(y)} - k^2 = 0$$

which implies that there exist $\mu^2, \nu^2 \in \mathbb{C}$ such that

$$-\frac{X''(x)}{X(x)} = \mu^2 \qquad \text{and} \qquad -\frac{Y''(y)}{Y(y)} = \nu^2$$

with $\mu^2 + \nu^2 = k^2$.

If we also consider the boundary conditions, we have

$$\begin{cases} -X'' - \mu^2 X = 0 & \text{in } (0,L) \\ -X'(0) = ikX(0),\ X'(L) = ikX(L) \end{cases} \qquad \text{and} \qquad \begin{cases} -Y'' - \nu^2 Y = 0 & \text{in } (0,L) \\ Y(0) = Y(L) = 0 \end{cases}$$

A non-zero Y can be obtained only for specific values of ν, corresponding to the eigenvalues of the one-dimensional Laplacian on $(0,L)$ (with Dirichlet boundary conditions), i.e. for $n := \frac{L\nu}{\pi} \in \mathbb{Z}^*$, where $\mathbb{Z}^* := \mathbb{Z} \setminus \{0\}$.

Hence we must have

$$k^2 = \mu^2 + \left(\frac{\pi n}{L}\right)^2 \qquad \text{for some } n \in \mathbb{Z}^*$$

which is equivalent to

$$\left\{\frac{L}{\pi}\sqrt{k^2 - \mu^2}\right\} \subset \mathbb{Z}^* \tag{16}$$

The analysis for X is less trivial. A simple computation shows that a non-zero X can be obtained only if

$$(k-\mu)^2 e^{iL\mu} = (k+\mu)^2 e^{-iL\mu}$$

i.e. if

$$\left(e^{iL\mu} - e^{-iL\mu}\right) k^2 - 2\mu \left(e^{iL\mu} + e^{-iL\mu}\right) k + \mu^2 \left(e^{iL\mu} - e^{-iL\mu}\right) = 0$$

If $\frac{L\mu}{\pi} \in \mathbb{Z}$, the equation becomes $\mu k = 0$, which only yields trivial solutions to problem (15). Let $\frac{L\mu}{\pi} \notin \mathbb{Z}$. The equation can be rewritten as

$$k^2 - 2\mu \frac{e^{iL\mu} + e^{-iL\mu}}{e^{iL\mu} - e^{-iL\mu}} k + \mu^2 = 0$$

whose two solutions are:

$$k^{\pm}(\mu) = \mu \frac{e^{iL\mu/2} \pm e^{-iL\mu/2}}{e^{iL\mu/2} \mp e^{-iL\mu/2}}$$

Plugging $k = k^+(\mu)$ into (16) yields (the derivation for $k = k^-(\mu)$ can be carried out similarly):

$$\left\{ \frac{L}{\pi} \sqrt{\frac{4\mu^2}{\left(e^{iL\mu/2} - e^{-iL\mu/2}\right)^2}} \right\} \subset \mathbb{Z}^*$$

i.e.

$$h^+(\mu) := \frac{2L\mu}{\pi\left(e^{iL\mu/2} - e^{-iL\mu/2}\right)} = -\frac{2i}{\pi} \frac{L\mu/2}{\sin\left(L\mu/2\right)} \in \mathbb{Z}^* \tag{17}$$

We determine qualitatively the number of solutions of (17) from the contour plot of h^+, in particular by looking at intersections between contour lines of the form

$$\zeta_0^+ := \left\{\mu \in \mathbb{C} \mid \Im(h^+(\mu)) = 0\right\} \quad \text{and} \quad \gamma_n^+ := \left\{\mu \in \mathbb{C} \mid \Re(h^+(\mu)) = n\right\} \tag{18}$$

for $n \in \mathbb{Z}^*$.

Let $n \in \mathbb{Z}^*$ be fixed. From contour plots of the real and imaginary parts of h^+ (see Figure 1), it appears that γ_n^+ is the countable union of disjoint simple curves, each intersecting ζ_0^+ only once. See Section 6.2 for a numerical example regarding one such intersection.

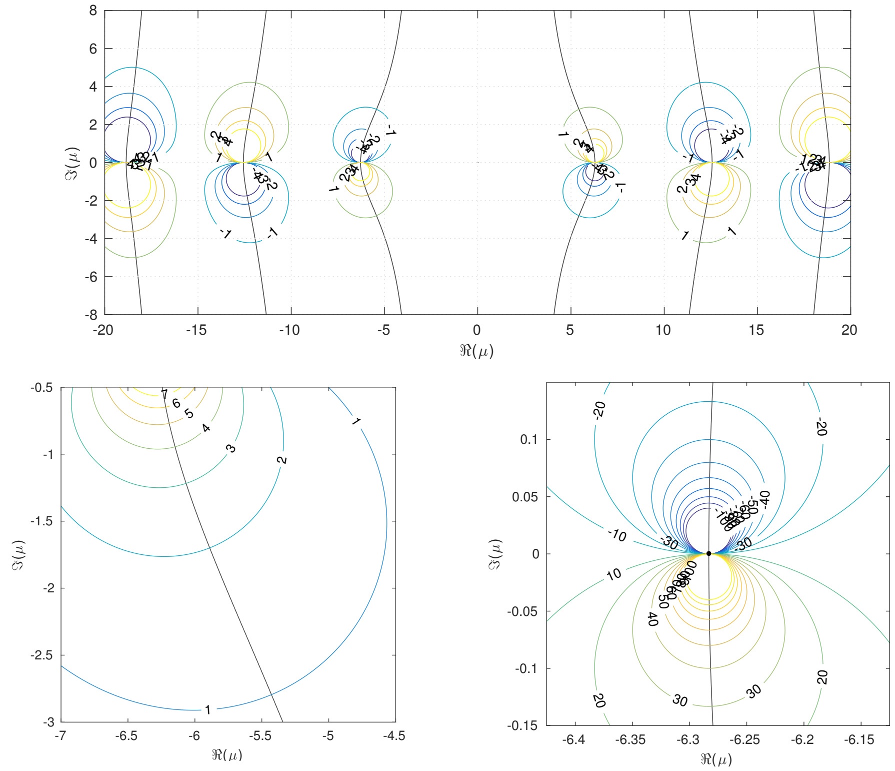

Figure 1: Plot of ζ_0^+ (black lines) and of $\gamma_n^+, n \in \mathbb{Z}^*$ (colored lines with labels), in the case $L = 1$. Any intersection between ζ_0^+ and some γ_n^+ belongs to M^+. However, poles of h^+ (i.e. any intersection between ζ_0^+ and the real axis $\Im(\mu) = 0$) must be excluded.

This could be made rigorous with a non-negligible amount of computation (which is skipped in this report): for instance one could parametrize ζ_0^+ near a pole of h^+ (e.g. by the implicit function theorem) and prove the monotonicity of $\Re(h^+)$ when restricted to the parametrized curve.

Thus we have somewhat proven that there exists a discrete set M^+ of values μ which satisfy equation (17). If M^- is defined from a similar analysis of k^-, then the *discrete* set $K := k^+(M^+) \cup k^-(M^-)$ includes all the possible values of k such that (15) has non-trivial solutions.

Still, we are far away from proving that the solution map is meromorphic. For example, we would require some approximation theory results in order to espress the general solution of the scattering problem (11) on a square as a (countable) superposition of non-trivial solutions of the homogeneous problem.

For this reason, in the next sections we limit our dissertation to the case of homogeneous Dirichlet boundary conditions, i.e. problem (1), with the possible extensions discussed in Remark 2.1.

4 Padé approximants in a functional environment

Since the solution map of the Helmholtz problem with homogeneous Dirichlet boundary conditions is well defined for any $z \in \mathbb{C} \setminus \Lambda$, it is quite easy to obtain a numerical approximation of $\mathcal{S}(\bar{z})$ for a single point $\bar{z} \in \mathbb{C} \setminus \Lambda$. For instance, it is possible to apply a Finite Elements approach (see e.g. [19]) to obtain an approximation of $\mathcal{S}(\bar{z})$ in some finite-dimensional subspace of V.

A completely different issue (which is of main interest in this project) is the approximation of $\mathcal{S}$ over a compact set $A \subset \mathbb{C}$, i.e. to estimate the behavior of the solution of (1) with respect to z.

Approximations based on grid evaluations are accurate, but too computationally expensive. Conversely, approaches which rely on Taylor expansions are not ideal because of the lack of regularity of $\mathcal{S}$, leading to approximations which are accurate only locally.

In the next sections we describe some approaches bases on Padé approximants, which exploit the properties described in Section 2.

4.1 An optimization-based definition of the Padé approximant of a meromorphic function

In this section we want to define the Padé approximant of a meromorphic function $\mathcal{T} \in \mathcal{M}(U; W)$, with $U \subset \mathbb{C}$. We indicate with $\Lambda \subset U$ the (discrete) set of poles of $\mathcal{T}$. Moreover, we assume that W is a Hilbert space[2] over $\mathbb{C}$, whose scalar product is $\langle \cdot, \cdot \rangle_W$.

We introduce the following notation: given a function ψ which is holomorphic in z_0, we indicate with $(\psi)_n(z_0)$ (or with $\psi_n(z_0)$ if this does not cause confusion) the n-th coefficient of its Taylor series centered in z_0, i.e.

$$\psi(z) = \sum_{n=0}^{\infty} \psi_n(z_0)\ (z - z_0)^n \qquad \forall z \in B_\rho(z_0) := \{z \in \mathbb{C}, |z - z_0| < \rho\}$$

for some $\rho > 0$.

Now we can define a Padé approximant of $\mathcal{T}$ (we defer to Section 4.2 a proof that such an approximant exists):

Definition 4.1. *Let $z_0 \in U$, $\rho > 0$ and $N, M, E \in \mathbb{N}$ be fixed. A $[M/N]$-Padé approximant (which depends parametrically on E and ρ) of $\mathcal{T}$ at point z_0 is a rational function*

$$\mathcal{T}^{[M/N]}(z) = \frac{P^{[M/N]}(z)}{Q^{[M/N]}(z)} \in \mathcal{M}(U; W)$$

such that (see [2]-[5]):

(a) with respect to z, $P^{[M/N]}$ is a polynomial with degree (at most) M, whose coefficients are elements of W. We indicate the set of polynomials which satisfy this property with $\mathbb{P}_M(\mathbb{C}; W)$.

(b) $Q^{[M/N]}$ is a polynomial with degree (at most) N, which is normalized in the following sense:

$$\sum_{n=0}^{N} \left| (Q^{[M/N]})_n(z_0) \right|^2 = 1$$

We indicate the set of polynomials which satisfy this property with $\mathbb{P}_N^{z_0}(\mathbb{C})$.

(c) $(P^{[M/N]}, Q^{[M/N]})$ achieves the minimum of j_E over $\mathbb{P}_M(\mathbb{C}; W) \times \mathbb{P}_N^{z_0}(\mathbb{C})$, where we define

$$j_E(P, Q) := \left(\sum_{\alpha=0}^{E} \|(Q\mathcal{T} - P)_\alpha(z_0)\|_W^2 \rho^{2\alpha} \right)^{1/2} \tag{19}$$

[2] The standard definition of $[M/N]$-Padé approximants (see e.g. [4]) assumes $W = \mathbb{C}$ and is carried out quite differently, being fundamentally based on the resolution of a linear system of equations of size $M + N + 1$.

The particular form of j_E is quite useful in proving convergence results for the Padé approximant, and can be interpreted as a measure of the distance between the truncated Taylor series (centered in z_0) of $\mathcal{T}$ and of $\frac{P}{Q}$.

In particular, we can show that j_E is related to some kind of approximation error on the circumference $\partial B_\rho(z_0)$.

Proposition 4.1. *Let j_E be defined as in (19) and let $\gamma = \partial B_\rho(z_0)$. We have*

$$j_E(P,Q) = \left(\frac{1}{2\pi i} \int_\gamma \frac{1}{z-z_0} \left\| \sum_{\alpha=0}^{E} (Q\mathcal{T} - P)_\alpha(z_0)\ (z-z_0)^\alpha \right\|_W^2 \mathrm{d}z \right)^{1/2}$$

Proof. First observe that, for any $\alpha, \beta \in \mathbb{N}$, we have

$$\frac{1}{2\pi i} \int_\gamma \frac{(z-z_0)^\alpha (z-z_0)^{*\beta}}{z-z_0} \mathrm{d}z = \int_0^1 \rho^\alpha e^{2\pi i \alpha\theta} \rho^\beta e^{-2\pi i \beta\theta} \mathrm{d}\theta = \rho^{2\alpha} \delta_{\alpha\beta}$$

Let $\psi = Q\mathcal{T} - P$. Definition (19) can be expressed as

$$\begin{aligned}
j_E(P,Q)^2 =& \sum_{\alpha=0}^{E} \sum_{\beta=0}^{E} \Big\langle \psi_\alpha(z_0), \psi_\beta(z_0) \Big\rangle_W \rho^{2\alpha} \delta_{\alpha\beta} \\
=& \sum_{\alpha=0}^{E} \sum_{\beta=0}^{E} \Big\langle \psi_\alpha(z_0), \psi_\beta(z_0) \Big\rangle_W \frac{1}{2\pi i} \int_\gamma \frac{(z-z_0)^\alpha (z-z_0)^{*\beta}}{z-z_0} \mathrm{d}z \\
=& \frac{1}{2\pi i} \int_\gamma \frac{1}{z-z_0} \Big\langle \sum_{\alpha=0}^{E} \psi_\alpha(z_0)\ (z-z_0)^\alpha, \sum_{\beta=0}^{E} \psi_\beta(z_0)\ (z-z_0)^\beta \Big\rangle_W \mathrm{d}z \\
=& \frac{1}{2\pi i} \int_\gamma \frac{1}{z-z_0} \left\| \sum_{\alpha=0}^{E} \psi_\alpha(z_0)\ (z-z_0)^\alpha \right\|_W^2 \mathrm{d}z
\end{aligned}$$

□

Many properties of Padé approximants, given by Definition 4.1, are described in [2]. A particularly important result is the following theorem, which proves the convergence of the approximant:

Theorem 4.1. *Let $z_0 \in U \setminus \Lambda$ and $N \in \mathbb{N}$ be fixed. Let $R \geq \rho > 0$ be positive real numbers for which there exist $h \in \mathcal{H}(B_R(z_0); W)$ and $g \in \mathbb{P}_N(\mathbb{C})$ such that*

$$\mathcal{T}(z) = \frac{h(z)}{g(z)} \qquad \text{for } z \in B_R(z_0)$$

(Equivalently we can ask that the cumulative order of the poles of $\mathcal{T}$ in $B_R(z_0)$ is (at most) N.) Moreover, let $M, E \in \mathbb{N}$ with $E \geq M + N$.

A Padé approximant $\mathcal{T}^{[M/N]}$ of $\mathcal{T}$ (given by Definition 4.1 with these choices of M, N, E and ρ) satisfies the following properties:

(a) $\lim_{M\to\infty} \|\mathcal{T}^{[M/N]}(z) - \mathcal{T}(z)\|_W = 0$ uniformly on all compact subsets of $B_R(z_0) \setminus \Lambda$.

(b) for any compact subset $A \subset (B_\rho(z_0) \setminus \Lambda)$, there exists M^* such that

$$\sup_{z\in A} \left\| \mathcal{T}^{[M/N]}(z) - \mathcal{T}(z) \right\|_W \leq C \left(\frac{\rho}{R} \right)^{M+1} \qquad \text{for any } M \geq M^* \tag{20}$$

where C depends only on A, ρ, R, N and $\mathcal{T}$.

Proof. See [2], Theorem 5.1. □

We can draw the following (more or less predictable) conclusions:

(a) A $[M/N]$-Padé approximant centered in z_0 is able to provide an approximation of the N poles of $\mathcal{T}$ which are closer to z_0 (counting each pole as many times as its order).

(b) The region of convergence of $\mathcal{T}^{[M/N]}$ is an open circle whose radius is the distance between z_0 and the $(N+1)$-th closest pole of $\mathcal{T}$ (again, counting each pole as many times as its order).

(c) There appears to be a trade-off in the choice of ρ: a larger ρ increases the size of the region where the approximant converges exponentially; conversely, a smaller ρ increases the speed of (exponential) convergence.

4.2 Basic algorithm for the computation of Padé approximants

In this section we want to describe an algorithm for the computation of a Padé approximant according to Definition 4.1. As a first (instructive) step, we want to prove the existence of such an approximation.

Theorem 4.2. *Let $z_0 \in U$, $\rho > 0$ and $N, M, E \in \mathbb{N}$ be fixed, with $E \geq M+1$. A $[M/N]$-Padé approximation which satisfies Definition 4.1 exists.*

Proof. We want to prove that j_E admits global minumum over $\mathbb{P}_M(\mathbb{C}; W) \times \mathbb{P}_N^{z_0}(\mathbb{C})$.

Using the fact that P has degree M, we can expand j_E as

$$\begin{aligned} j_E(P,Q)^2 &= \sum_{\alpha=0}^{M} \|(Q\mathcal{T} - P)_\alpha(z_0)\|_W^2 \rho^{2\alpha} + \sum_{\alpha=M+1}^{E} \|(Q\mathcal{T} - P)_\alpha(z_0)\|_W^2 \rho^{2\alpha} \\ &= \sum_{\alpha=0}^{M} \|(Q\mathcal{T} - P)_\alpha(z_0)\|_W^2 \rho^{2\alpha} + \sum_{\alpha=M+1}^{E} \|(Q\mathcal{T})_\alpha(z_0)\|_W^2 \rho^{2\alpha} \end{aligned}$$

Now, let Q be fixed. There exists a choice of P for which the first $M+1$ terms of the sum are zero:

$$P_Q(z) := \sum_{\alpha=0}^{M} (Q\mathcal{T})_\alpha(z_0)\ (z - z_0)^\alpha$$

Since each term of the sum is non-negative, and only the first $M+1$ terms of the sum depend on P, it is trivial to see that this choice is optimal.

Hence the problem of minimizing j_E can be restricted to the minimization over Q of

$$\bar{j}_E(Q) := j_E(P_Q, Q) = \left(\sum_{\alpha=M+1}^{E} \|(Q\mathcal{T})_\alpha(z_0)\|_W^2 \rho^{2\alpha} \right)^{1/2} \tag{21}$$

Since $\bar{j}_E$ is continuous and $\mathbb{P}_N^{z_0}(\mathbb{C})$ is compact (since it is homeomorphic to the surface of the unit sphere in $\mathbb{C}^{N+1}$), $\bar{j}_E$ admits global minimum on $\mathbb{P}_N^{z_0}(\mathbb{C})$. □

Taking inspiration from the previous proof, we can devise the following algorithm:

Algorithm 1.

(a) Fix $z_0 \in U, \rho > 0$ and $M, N, E \in \mathbb{N}$, with $E \geq M + N$.

(b) Find the denominator $Q^{[M/N]}$ as the element of $\mathbb{P}_N^{z_0}(\mathbb{C})$ which minimizes $\bar{j}_E(Q)$, defined as in (21).

(c) Find the numerator $P^{[M/N]}$ as $P^{[M/N]}(z) := \sum_{\alpha=0}^{M} (Q^{[M/N]}\mathcal{T})_\alpha(z_0)\ (z - z_0)^\alpha$.

Of particular interest is Step (b), where we must solve a constrained optimization problem. In solving it, we can exploit a particular structure which follows from definition (21).

Let $q_\alpha := Q_\alpha(z_0)$ for $\alpha = 0, 1, \ldots$ (in particular, $q_\alpha = 0$ for $\alpha \geq N+1$). Since $(Q\mathcal{T})_\alpha(z_0) = \sum_{n=0}^{\alpha} q_n \mathcal{T}_{\alpha-n}(z_0)$, we have

$$\begin{aligned}
\bar{j}_E(Q)^2 &= \sum_{\alpha=M+1}^{E} \Big\langle (Q\mathcal{T})_\alpha(z_0), (Q\mathcal{T})_\alpha(z_0) \Big\rangle_W \rho^{2\alpha} \\
&= \sum_{\alpha=M+1}^{E} \Big\langle \sum_{j=0}^{\alpha} q_j \mathcal{T}_{\alpha-j}(z_0), \sum_{i=0}^{\alpha} q_i \mathcal{T}_{\alpha-i}(z_0) \Big\rangle_W \rho^{2\alpha} \\
&= \sum_{\alpha=M+1}^{E} \sum_{i,j=0}^{\alpha} q_i^* q_j \Big\langle \mathcal{T}_{\alpha-j}(z_0), \mathcal{T}_{\alpha-i}(z_0) \Big\rangle_W \rho^{2\alpha}
\end{aligned}$$

which is a quadratic form with respect to $\mathbf{q} := (q_n)_{n=0}^N$.

As such there exists a Hermitian matrix $G_E^{[M/N]} \in \mathbb{C}^{(N+1)\times(N+1)}$ such that $\bar{j}_E(Q) = \mathbf{q}^* G_E^{[M/N]} \mathbf{q}$. Its expression is the following:

$$\left(G_E^{[M/N]}\right)_{ij} = \sum_{\alpha=\max\{i,j,M+1\}}^{E} \Big\langle \mathcal{T}_{\alpha-j}(z_0), \mathcal{T}_{\alpha-i}(z_0) \Big\rangle_W \rho^{2\alpha} \qquad \text{for } i,j = 0,\ldots,N \tag{22}$$

Definition (21) implies that $G_E^{[M/N]}$ is positive-semidefinite. Moreover, observe that the two conditions $Q \in \mathbb{P}_N^{z_0}(\mathbb{C})$ and $\|\mathbf{q}\|_2 = 1$ are equivalent.

Thus Step (b) corresponds to the problem of identifying the smallest eigenvector of $G_E^{[M/N]}$, which more generally can be expressed as a Quadratically Constrained Quadratic Program (QCQP).

An interesting interpretation of (22) is the following: $G_E^{[M/N]}$ is obtained through a weighted sum of sub-matrices extracted from a Gram matrix associated to $\mathcal{T}$ through the W-scalar product, whose elements are of the form:

$$(G)_{ij} = \langle \mathcal{T}_i(z_0), \mathcal{T}_j(z_0) \rangle_W \qquad \text{for } i,j = 0,\ldots,N$$

See Figure 2 for a graphical representation of this process.

In particular, the choice of ρ impacts the algorithm only by changing the weights of such a sum: smaller values of ρ give more importance to sub-matrices located in the top-left portion of G, whereas bigger values of ρ enhance bottom-right entries.

Some numerical experiments in the particular case of the Helmholtz equation (1), which are shown in Section 6.3, lead to believe that the choice of the parameter ρ has little effect on

$$G = \begin{bmatrix}
\langle \mathcal{T}, \mathcal{T} \rangle & \langle \mathcal{T}, \mathcal{T}_1 \rangle & \langle \mathcal{T}, \mathcal{T}_2 \rangle & \cdots & & \\
\langle \mathcal{T}_1, \mathcal{T} \rangle & \langle \mathcal{T}_1, \mathcal{T}_1 \rangle & \langle \mathcal{T}_1, \mathcal{T}_2 \rangle & \langle \mathcal{T}_1, \mathcal{T}_3 \rangle & \cdots & \\
\langle \mathcal{T}_2, \mathcal{T} \rangle & \langle \mathcal{T}_2, \mathcal{T}_1 \rangle & \langle \mathcal{T}_2, \mathcal{T}_2 \rangle & \langle \mathcal{T}_2, \mathcal{T}_3 \rangle & \langle \mathcal{T}_2, \mathcal{T}_4 \rangle & \cdots \\
\vdots & \langle \mathcal{T}_3, \mathcal{T}_1 \rangle & \langle \mathcal{T}_3, \mathcal{T}_2 \rangle & \langle \mathcal{T}_3, \mathcal{T}_3 \rangle & \langle \mathcal{T}_3, \mathcal{T}_4 \rangle & \cdots \\
& \vdots & \langle \mathcal{T}_4, \mathcal{T}_2 \rangle & \langle \mathcal{T}_4, \mathcal{T}_3 \rangle & \langle \mathcal{T}_4, \mathcal{T}_4 \rangle & \cdots \\
& & \vdots & \vdots & \vdots &
\end{bmatrix}$$

$$G_E^{[M/N]} = \ldots + \rho^6 \begin{bmatrix}
\langle \mathcal{T}_3, \mathcal{T}_3 \rangle & \langle \mathcal{T}_2, \mathcal{T}_3 \rangle & \langle \mathcal{T}_1, \mathcal{T}_3 \rangle \\
\langle \mathcal{T}_3, \mathcal{T}_2 \rangle & \langle \mathcal{T}_2, \mathcal{T}_2 \rangle & \langle \mathcal{T}_1, \mathcal{T}_2 \rangle \\
\langle \mathcal{T}_3, \mathcal{T}_1 \rangle & \langle \mathcal{T}_2, \mathcal{T}_1 \rangle & \langle \mathcal{T}_1, \mathcal{T}_1 \rangle
\end{bmatrix} + \ldots$$

Figure 2: Gram matrix (top) associated to $\mathcal{T}$ through $\langle\cdot,\cdot\rangle$. We omit the argument (z_0) of the Taylor coefficients $\mathcal{T}_\alpha$, for $\alpha = 0, 1, \ldots$. In blue the sub-matrix extracted for $N = 2, \alpha = 3$, which provides a contribution to $G_E^{[M/N]}$ (bottom) with weight ρ^6. Observe that a transposition with respect to the secondary diagonal is carried out before computing the sum.

the convergence of the Padé approximant. As a consequence of this observation, we propose a modification to Step (b) in Algorithm 1.

4.3 Modification of the algorithm through the marginalization of ρ

For extreme values of ρ ($\rho \to 0$ or $\rho \to \infty$) one of the terms of the sum appearing in (21) dominates the others. Hence we may choose to simplify Definition 4.1 by considering just the leading term:

Definition 4.2. *Let $z_0 \in U$ and $N, M, E \in \mathbb{N}$ be fixed. A fast $[M/N]$-Padé approximant (which depends parametrically on E) of $\mathcal{T}$ at point z_0 is a rational function*

$$\mathcal{T}^{[M/N]}(z) = \frac{P^{[M/N]}(z)}{Q^{[M/N]}(z)} \in \mathcal{M}(U; W)$$

such that:

(a) $P^{[M/N]} \in \mathbb{P}_M(\mathbb{C}; W)$.

(b) $Q^{[M/N]} \in \mathbb{P}_N^{z_0}(\mathbb{C})$.

(c) The Taylor series (centered in z_0) of $(Q^{[M/N]}\mathcal{T} - P^{[M/N]})$ is such that

$$\left(Q^{[M/N]}\mathcal{T} - P^{[M/N]}\right)_\alpha (z_0) = 0 \qquad \textit{for } \alpha = 0, \dots, M$$

(d) $Q^{[M/N]}$ achieves the minimum of $\widetilde{j}_E$ over $\mathbb{P}_N^{z_0}(\mathbb{C})$, where we define

$$\widetilde{j}_E(Q) := \|(Q\mathcal{T})_E(z_0)\|_W \tag{23}$$

The corresponding algorithm is:

Algorithm 2.

(a) Fix $z_0 \in U$ and $M, N, E \in \mathbb{N}$, with $E \geq M, N$.

(b) Find the denominator $Q^{[M/N]}$ as the element of $\mathbb{P}_N^{z_0}(\mathbb{C})$ which minimizes $\widetilde{j}_E$, defined as in (23).

(c) Find the numerator $P^{[M/N]}$ as $P^{[M/N]}(z) := \sum_{\alpha=0}^{M} (Q^{[M/N]}\mathcal{T})_\alpha(z_0)\ (z - z_0)^\alpha$.

As in the previous algorithm, Step (b) can be carried out by finding the smallest eigenvector of a positive-semidefinite Hermitian matrix. However, in this case, only a single $(N+1) \times (N+1)$-window of the Gram matrix G has to be considered:

$$\left(G_E^{[M/N]}\right)_{ij} = \left\langle \mathcal{T}_{E-j}(z_0), \mathcal{T}_{E-i}(z_0) \right\rangle_W \qquad \text{for } i, j = 0, \dots, N \tag{24}$$

The resulting rational function $\mathcal{T}^{[M/N]} = \frac{P^{[M/N]}}{Q^{[M/N]}}$ is an approximation of the Padé approximant described in Definition 4.1.

At first glance, this algorithm appears to be an improvement of the old one: if the amount of known information about $\mathcal{T}$ (i.e. $E + 1$, the number of known derivatives of $\mathcal{T}$) is kept fixed, Algorithm 2 can lead to an approximant with higher degree (in terms of M). This should lead to a better result, since a higher value of M corresponds to a better accuracy in approximating the Taylor series of $\mathcal{T}$.

As a negative side, it is reasonable to expect a decrease in the accuracy of the approximation of the poles of $\mathcal{T}$, due to the reduced amount of information which is considered in Step (b).

Interestingly, a numerical comparison of the two algorithms, which is shown in Section 6.4, leads to conclude that, for fixed M, they have a similar accuracy in the case where all the poles of the target (meromorphic) function are simple.

Still, a formal proof of the convergence rate for Algorithm 2 appears to be quite complicated. We limit our dissertation to some preliminary results.

4.4 Some results for the modified algorithm

The crucial step of Algorithm 2 consists in the computation of the smallest eigenvector of the Hermitian matrix defined as in (24).

Now, any function $\mathcal{T}$ belonging to $\mathcal{M}(B_R(z_0); W)$, which has exactly N simple poles $\{\lambda_n\}_{n=1}^N \subset B_R(z_0)$, admits a representation of the form

$$\mathcal{T}(z) = \sum_{n=1}^{N} \frac{h_n(z)}{\lambda_n - z}$$

with $h_n \in \mathcal{H}(B_R(z_0); W)$ such that $h_n(\lambda_n) \neq 0$ (for $n = 1, \ldots, N$).

Using this expression, the Taylor coefficient $\mathcal{T}_\alpha(z_0)$ assumes the form

$$\begin{aligned}\mathcal{T}_\alpha(z_0) =& \sum_{n=1}^{N} \sum_{\beta=0}^{\alpha} (h_n)_\beta(z_0) \left((\lambda_n - \cdot)^{-1}\right)_{\alpha-\beta}(z_0) \\ =& \sum_{n=1}^{N} \sum_{\beta=0}^{\alpha} (h_n)_\beta(z_0)\ (\lambda_n - z_0)^{-1-\alpha+\beta} \\ =& \sum_{n=1}^{N} (\lambda_n - z_0)^{-1-\alpha} \sum_{\beta=0}^{\alpha} (h_n)_\beta(z_0)\ (\lambda_n - z_0)^{\beta}\end{aligned}$$

For ease of notation we define

$$\psi_{n,z_0}^{\alpha} := \sum_{\beta=0}^{\alpha} (h_n)_\beta(z_0)\ (\lambda_n - z_0)^{\beta} \qquad \text{for } n = 1, \ldots, N,\ \alpha \in \mathbb{N} \tag{25}$$

An entry of the Gram sub-matrix (24) becomes

$$\begin{aligned}\left(G_E^{[M/N]}\right)_{ij} &= \sum_{n,n'=1}^{N} (\lambda_n - z_0)^{-1-E+j} \left((\lambda_{n'} - z_0)^*\right)^{-1-E+i} \langle \psi_{n,z_0}^{E-j}, \psi_{n',z_0}^{E-i} \rangle_W \\ &= \sum_{n,n'=1}^{N} (\lambda_n - z_0)^{-1-E+j} \left((\lambda_{n'} - z_0)^*\right)^{-1-E+i} \langle \psi_{n,z_0}^{E} - \varepsilon_{n,j}^{E}, \psi_{n',z_0}^{E} - \varepsilon_{n',i}^{E} \rangle_W\end{aligned}$$

where $\varepsilon_{n,i}^{\alpha}$ is defined as

$$\varepsilon_{n,i}^{\alpha} := \psi_{n,z_0}^{\alpha} - \psi_{n,z_0}^{\alpha-i} = \sum_{\beta=\alpha-i+1}^{\alpha} (h_n)_\beta(z_0)\ (\lambda_n - z_0)^{\beta}$$

In particular, the Gram matrix (24) can be decomposed as $G_E^{[M/N]} = H_E^{[M/N]} + \delta H_E^{[M/N]}$, where

$$\left(H_E^{[M/N]}\right)_{ij} := \sum_{n,n'=1}^{N} (\lambda_n - z_0)^{-1-E+j} \left((\lambda_{n'} - z_0)^*\right)^{-1-E+i} \langle \psi_{n,z_0}^{E}, \psi_{n',z_0}^{E} \rangle_W \tag{26}$$

and

$$\begin{aligned}\left(\delta H_E^{[M/N]}\right)_{ij} := \sum_{n,n'=1}^{N} (\lambda_n - z_0)^{-1-E+j} &\left((\lambda_{n'} - z_0)^*\right)^{-1-E+i} \\ &\left(\langle \varepsilon_{n,j}^{E}, \varepsilon_{n',i}^{E} \rangle_W - \langle \varepsilon_{n,j}^{E}, \psi_{n',z_0}^{E} \rangle_W - \langle \psi_{n,z_0}^{E}, \varepsilon_{n',i}^{E} \rangle_W\right)\end{aligned} \tag{27}$$

We can prove the following property:

Proposition 4.2. *Let $\widehat{Q} \in \mathbb{P}_N^{z_0}(\mathbb{C})$ be a (normalized) polynomial whose zeros are $\lambda_1, \ldots, \lambda_N$, and let $\widehat{q}_n := (\widehat{Q})_n(z_0)$ for $n = 0, \ldots, N$. The vector $\widehat{\mathbf{q}} := (\widehat{q}_n)_{n=0}^N$ is an eigenvector of $H_E^{[M/N]}$, and its corresponding eigenvalue is minimal.*

Proof. First we show that $H_E^{[M/N]}$ is positive-semidefinite, so that the minimum eigenvalue is non-negative.

Let $\mathbf{q} \in \mathbb{C}^{N+1}$. We can compute the quadratic form

$$\mathbf{q}^* H_E^{[M/N]} \mathbf{q} = \sum_{i,j=0}^{N} \sum_{n,n'=1}^{N} (\lambda_n - z_0)^{-1-E+j} \left((\lambda_{n'} - z_0)^*\right)^{-1-E+i} q_i^* q_j \langle \psi_{n,z_0}^E, \psi_{n',z_0}^E \rangle_W =$$

$$= \sum_{n,n'=1}^{N} \left((\lambda_n - z_0)(\lambda_{n'} - z_0)^*\right)^{-1-E} \langle \psi_{n,z_0}^E, \psi_{n',z_0}^E \rangle_W \left(\sum_{i=0}^{N} q_i (\lambda_{n'} - z_0)^i\right)^* \left(\sum_{j=0}^{N} q_j (\lambda_n - z_0)^j\right)$$

Let $Q(z) := \sum_{i=0}^N q_i (z - z_0)^i$. Bringing the sums inside the inner product yields

$$\mathbf{q}^* H_E^{[M/N]} \mathbf{q} = \left\langle \sum_{n=1}^{N} (\lambda_n - z_0)^{-1-E} \psi_{n,z_0}^E Q(\lambda_n), \sum_{n'=1}^{N} (\lambda_{n'} - z_0)^{-1-E} \psi_{n',z_0}^E Q(\lambda_{n'}) \right\rangle_W$$

$$= \left\| \sum_{n=1}^{N} \frac{\psi_{n,z_0}^E}{(\lambda_n - z_0)^{1+E}} Q(\lambda_n) \right\|_W^2$$

which is non-negative.

Since $\widehat{Q}(\lambda_n) = 0$ for $n = 1, \ldots, N$, we have that $\widehat{\mathbf{q}}^* H_E^{[M/N]} \widehat{\mathbf{q}} = 0$. Hence $\widehat{\mathbf{q}}$ is an eigenvector of $H_E^{[M/N]}$ with eigenvalue 0 (which is minimal). □

Now, observe that ψ_{n,z_0}^α, defined as in (25), corresponds to the truncated Taylor series (centered in z_0) of h_n, evaluated at $z = \lambda_n$. Since h_n is holomorphic in $B_R(z_0)$, ψ_{n,z_0}^α converges uniformly to $h_n(\lambda_n)$ as $\alpha \to \infty$. Thus $\{\psi_{n,z_0}^\alpha\}_{\alpha \geq 0}$ is a Cauchy sequence in W, and $\|\varepsilon_{n,i}^\alpha\|_W$ tends to 0 as $\alpha \to \infty$ for any fixed i and n.

As E increases, $\|\psi_{n,z_0}^E\|_W$ stays bounded and converges to the positive value $\|h_n(\lambda_n)\|_W$. Hence we can use the Cauchy-Schwartz inequality to deduce that

$$\lim_{E \to \infty} \frac{\|\delta H_E^{[M/N]}\|}{\|H_E^{[M/N]}\|} = 0 \qquad \text{(e.g. in the Frobenius norm)}$$

i.e. that the relative size of the perturbation tends to zero as E increases.

Then it is reasonable to assume that, for $E \to \infty$, the subspace generated by the smallest eigenvector of $G_E^{[M/N]}$ converges to the span of $\widehat{\mathbf{q}}$, defined as in Proposition 4.2. An exact proof of this assumption, which may be based e.g. on the "$\sin \theta$" theorem (see e.g. [9]), is omitted.

4.5 Conditioning of the Gram matrix and single-pole approximations

The results obtained in the previous section can be interpreted as follows: by increasing E, it is possible to remove some of the noise caused by the numerators h_n ($n = 1, \ldots, N$), and ease the identification of the poles of $\mathcal{T}$.

Still, numerical experiments show that the Gram sub-matrix $G_E^{[M/N]}$ is quite ill-conditioned for big values of E (see e.g. Section 6.7), with most of its eigenvalues converging to 0. This is quite troublesome for the accuracy and the efficiency of most numerical solvers for eigenvalue problems.

Indeed, we can verify that this is true in general (for a holomorphic function with simple poles):

Proposition 4.3. *Assume that*

$$|\lambda_1 - z_0| < |\lambda_n - z_0| \quad \textit{for } n = 2, \dots, N$$

For $E \to \infty$, a multiple of the Gram sub-matrix $|\lambda_1 - z_0|^{2+2E} G_E^{[M/N]}$ converges (e.g. in the Frobenius norm) to a rank 1 matrix, whose range is the span of $(1, \lambda_1 - z_0, \dots, (\lambda_1 - z_0)^N)^*$.

Proof. Let $\omega_n := \lambda_n - z_0$ for $n = 1, \dots, N$. From (26) we have

$$\left(H_E^{[M/N]}\right)_{ij} = \omega_1^{-1-E+j} \left(\omega_1^*\right)^{-1-E+i} \sum_{n,n'=1}^{N} \left(\frac{\omega_1}{\omega_n}\right)^{1+E-j} \left(\frac{\omega_1^*}{\omega_{n'}^*}\right)^{1+E-i} \langle \psi_{n,z_0}^E, \psi_{n',z_0}^E \rangle_W$$

If all the poles are simple, we have that $\|\psi_{n,z_0}^E\|_W$ is bounded for any n and E (since h_n is holomorphic). Hence we can neglect all the terms of the sum except the first one:

$$\lim_{E\to\infty} \left(\left(H_E^{[M/N]}\right)_{ij} - \omega_1^{-1-E+j} \left(\omega_1^*\right)^{-1-E+i} \|\psi_{1,z_0}^E\|_W^2 \right) = 0$$

for $i, j = 0, \dots, N$.

Let $\mathbf{w}_1 := (1, \omega_1, \dots, \omega_1^N)^*$. We have

$$\lim_{E\to\infty} \left\| |\omega_1|^{2+2E} H_E^{[M/N]} - \|\psi_{1,z_0}^E\|_W^2 \ \mathbf{w}_1 \mathbf{w}_1^* \right\| = 0 \qquad \text{(e.g. in the Frobenius norm)}$$

Since $G_E^{[M/N]}$ converges to $H_E^{[M/N]}$ for $E \to \infty$, the proof follows from the triangular inequality. □

From the previous property, we can derive a quite simple algorithm for the approximation of a single pole of $\mathcal{T}$, in particular the one closest to z_0, without solving any eigenvalue problem:

Algorithm 3.

(a) Fix $z_0 \in U$ and $E \in \mathbb{N} \setminus \{0\}$.

(b) Compute $\mathcal{T}_{E-1}(z_0)$ and $\mathcal{T}_E(z_0)$.

(c) Approximate the closest pole of $\mathcal{T}$ as

$$\lambda^{[E]} = z_0 + \frac{(G_E^{[E|1]})_{01}}{(G_E^{[E|1]})_{00}} = z_0 + \frac{\langle \mathcal{T}_{E-1}(z_0), \mathcal{T}_E(z_0) \rangle_W}{\langle \mathcal{T}_E(z_0), \mathcal{T}_E(z_0) \rangle_W}$$

If the minimum of $|\cdot - z_0|$ over Λ is unique, Proposition 4.3 proves that $\lambda^{[E]}$ converges to an element of Λ as $E \to \infty$. See Section 6.5 for an example of the application of this algorithm.

5 Multi-point Padé approximants in a functional environment

In this section we want to extend Algorithm 2 to the case of multi-point rational approximants.

In the case of a complex-valued function ϕ, we can define a multi-point $[M/N]$-Padé approximant as follows. Given $Z_K = \{z_1, \ldots, z_K\} \subset \mathbb{C}$ sample points with average $\bar{z} = \frac{1}{K}\sum_{k=1}^{K} z_k$, a multi-point $[M/N]$-Padé approximant of ϕ based on Z_K is a rational function $\phi^{[M/N]}$ such that:

(a) $\phi^{[M/N]}(z) = \frac{P^{[M/N]}(z)}{Q^{[M/N]}(z)}$, with $P^{[M/N]} \in \mathbb{P}_M(\mathbb{C})$ and $Q^{[M/N]} \in \mathbb{P}_N^{\bar{z}}(\mathbb{C})$.

(b) The Taylor series of $\phi^{[M/N]}$ centered in z_k matches the one of ϕ, up to (at least) order $\left\lfloor \frac{M+N+1}{K} \right\rfloor - 1$, for $k = 1, \ldots, K$.

We refer to [4] for a more general discussion of the properties of such an approximant.

Our aim is to apply the general idea of the multi-point Padé approximant to obtain a similar definition for W-valued meromorphic functions (with W defined as in Section 4.1).

5.1 A definition of multi-point Padé approximants for meromorphic functions

We choose to define the multi-point Padé approximant of a W-valued meromorphic function $\mathcal{T}$ by generalizing Definition 4.2:

Definition 5.1. *Let $Z_K = \{z_1, \ldots, z_K\} \subset U$ be a set of distinct sample points, and $W_K = \{w_1, \ldots, w_K\}$ be positive weights such that $\sum_{k=1}^{K} w_k = 1$. Let $\bar{z} = \sum_{k=1}^{K} w_k z_k$ be the weighted average of the sample points. Also, let $N, M, E \in \mathbb{N}$ be fixed.*

A multi-point $[M/N]$-Padé approximant (which depends parametrically on E) of $\mathcal{T}$ based on (Z_K, W_K) is a rational function

$$\mathcal{T}^{[M/N]}(z) = \frac{P^{[M/N]}(z)}{Q^{[M/N]}(z)} \in \mathcal{M}(U; W)$$

such that:

(a) $P^{[M/N]} \in \mathbb{P}_M(\mathbb{C}; W)$.

(b) $Q^{[M/N]} \in \mathbb{P}_N^{\bar{z}}(\mathbb{C})$.

(c) for any $k = 1, \ldots, K$, the Taylor series (centered in z_k) of $(Q^{[M/N]}\mathcal{T} - P^{[M/N]})$ is such that

$$\left(Q^{[M/N]}\mathcal{T} - P^{[M/N]}\right)_{\alpha}(z_k) = 0 \qquad \text{for } \alpha = 0, \ldots, \left\lfloor \frac{M+1}{K} \right\rfloor - 1 \tag{28}$$

(d) $Q^{[M/N]}$ achieves the minimum of $\widehat{j}_E$ over $\mathbb{P}_N^{\bar{z}}(\mathbb{C})$, where we define

$$\widehat{j}_E(Q)^2 = \sum_{k=1}^{K} w_k \|(Q\mathcal{T})_E(z_k)\|_W^2 \tag{29}$$

We may additionally require that $M+1$ is a multiple of K, so that (c) identifies $P^{[M/N]}$ uniquely, given $Q^{[M/N]}$.

In order to find a multi-point Padé approximant, we may apply the following algorithm:

Algorithm 4.

(a) Fix Z_K, W_K, N, M, E as in Definition 4.2, with $E \geq \max\left\{N, \left\lfloor \frac{M+1}{K} \right\rfloor - 1\right\}$.

(b) Find the denominator $Q^{[M/N]}$ as the element of $\mathbb{P}_N^{\bar{z}}(\mathbb{C})$ which minimizes $\widehat{j}_E$, defined as in (29).

(c) Find the numerator $P^{[M/N]}$ as an element of $\mathbb{P}_M(\mathbb{C}; W)$ which satisfies (28).

Observe that Step (c) consists in computing the Hermite interpolant (see e.g. [20], Chapter 8.5) of $Q^{[M/N]}\mathcal{T}$ in the sample points Z_K, using $\lfloor \frac{M+1}{K} \rfloor$ derivatives in each point.

As in the previous cases, let us discuss in detail how Step (b) is carried out. First we prove the following result.

Proposition 5.1. *Let $Q \in \mathbb{P}_N(\mathbb{C})$. For any two values $z', z'' \in \mathbb{C}$, there exists a linear transformation $T_{z'\to z''} : \mathbb{C}^{N+1} \to \mathbb{C}^{N+1}$ which associates $(Q_n(z'))_{n=0}^N$ to $(Q_n(z''))_{n=0}^N$. The representative matrix of $T_{z'\to z''}$ (using the canonical basis of $\mathbb{C}^{N+1}$) is provided in the proof.*

Proof. If $z' = z''$, trivially $T_{z'\to z''} = I$.

Let $z' \neq z''$. By definition, $Q(z) = \sum_{j=0}^N Q_j(z')\ (z-z')^j$.

For $i, j = 0, \ldots, N$, define

$$(T_{z'\to z''})_{ij} := \left((\,\cdot - z')^j\right)_i (z'') = \begin{cases} \binom{j}{i}\ (z''-z')^{j-i} & \text{if } i \leq j \\ 0 & \text{if } i > j \end{cases}$$

We have

$$Q_i(z'') = \sum_{j=i}^N Q_j(z') \binom{j}{i}\ (z''-z')^{j-i} \qquad \text{for } i = 0, \ldots, N$$

or equivalently

$$\left(Q_n(z'')\right)_{n=0}^N = T_{z'\to z''} \left(Q_n(z')\right)_{n=0}^N$$

□

Let $Q \in \mathbb{P}_N(\mathbb{C})$. For any $z \in U$, let $q_n^{(z)} := Q_n(z)$ for $n = 0, \ldots, N$, and $\mathbf{q}^{(z)} := \left(q_n^{(z)}\right)_{n=0}^N$. We can expand $\widehat{\jmath}_E$ as

$$\begin{aligned} \widehat{\jmath}_E(Q) =& \sum_{k=1}^K w_k \sum_{i,j=0}^N \left(q_i^{(z_k)}\right)^* q_j^{(z_k)} \left\langle \mathcal{T}_{E-j}(z_k), \mathcal{T}_{E-i}(z_k)\right\rangle_W \\ =& \sum_{k=1}^K w_k \sum_{i,j=0}^N \left(q_i^{(z_k)}\right)^* q_j^{(z_k)} \left(G_E^{[M/N]}\big|_{z_k}\right)_{ij} \end{aligned}$$

where

$$\left(G_E^{[M/N]}\big|_{z_k}\right)_{ij} = \left\langle \mathcal{T}_{E-j}(z_k), \mathcal{T}_{E-i}(z_k)\right\rangle_W \quad \text{for } i,j = 0, \ldots, N \text{ and } k = 1, \ldots, K \tag{30}$$

Proposition 5.1 ensures that there exist K matrices $\{T_{\bar{z}\to z_k}\}_{k=1}^K \subset \mathbb{C}^{(N+1)\times(N+1)}$ such that

$$\mathbf{q}^{(z_k)} = T_{\bar{z}\to z_k}\mathbf{q}^{(\bar{z})}$$

Hence

$$\begin{aligned} \widehat{\jmath}_E(Q) =& \sum_{k=1}^K w_k \sum_{i,j=0}^N \left(T_{\bar{z}\to z_k}\mathbf{q}^{(\bar{z})}\right)_i^* \left(T_{\bar{z}\to z_k}\mathbf{q}^{(\bar{z})}\right)_j \left(G_E^{[M/N]}\big|_{z_k}\right)_{ij} \\ =& \sum_{k=1}^K w_k\ \mathbf{q}^{(\bar{z})*} T_{\bar{z}\to z_k}^* G_E^{[M/N]}\big|_{z_k} T_{\bar{z}\to z_k}\mathbf{q}^{(\bar{z})} \\ =&\ \mathbf{q}^{(\bar{z})*} \left(\sum_{k=1}^K w_k T_{\bar{z}\to z_k}^* G_E^{[M/N]}\big|_{z_k} T_{\bar{z}\to z_k}\right) \mathbf{q}^{(\bar{z})} \end{aligned}$$

and Step (b) can be treated in the same way as in Algorithm 2: it is enoguh to find the smallest eigenvector of the positive-semidefinite Hermitian matrix

$$\sum_{k=1}^K w_k T_{\bar{z}\to z_k}^* G_E^{[M/N]}\big|_{z_k} T_{\bar{z}\to z_k}$$

5.2 On the convergence rate of multi-point Padé approximants

At the time of the writing, there are no (known) theoretical results regarding the accuracy of the multi-point Padé approximant given by Definition 5.1. We may guess a possible convergence rate for such an approximant by looking at known results for Hermite interpolation (see e.g. [20]).

Let $\psi \in \mathcal{H}(\mathbb{C};\mathbb{C})$, $E, K \in \mathbb{N}$ and $Z_K = \{z_1, \ldots, z_K\} \subset \mathbb{C}$. Set $M = (E+1)K - 1$ and let $\widehat{\psi} \in \mathbb{P}_M(\mathbb{C})$ be the Hermite interpolant of ψ based on $E+1$ derivatives in each point z_k.

Let $B \subset \mathbb{C}$ be the smallest ball including Z_K. For any $z \in B$, there exists $\xi \in B$ such that

$$\widehat{\psi}(z) - \psi(z) = \psi_{M+1}(\xi) \prod_{k=1}^{K} (z - z_k)^{E+1}$$

which implies

$$\left|\widehat{\psi}(z) - \psi(z)\right| \leq \max_{\xi \in B} |\psi_{M+1}(\xi)| \prod_{k=1}^{K} |z - z_k|^{E+1} \tag{31}$$

We can interpret (31) as the natural generalization of the error for truncated Taylor series

$$\left|\sum_{n=0}^{N} \psi_n(z_0)(z - z_0)^n - \psi(z)\right| \leq \max_{\xi \in B_{|z-z_0|}(z_0)} |\psi_{N+1}(\xi)| |z - z_0|^{N+1}$$

Hence we may hope that a generalization of the error estimate (20) could hold for the multi-point Padé approximant by replacing

$$\left(\frac{\rho}{R}\right)^{M+1} \simeq \left(\frac{\sup_{z \in A} |z - z_0|}{\inf_{z \in \partial B_R(z_0)} |z - z_0|}\right)^{M+1}$$

with

$$\prod_{k=1}^{K} \left(\frac{\sup_{z \in A} |z - z_k|}{\inf_{z \in \partial B_R(\bar{z})} |z - z_k|}\right)^{E+1} = \prod_{k=1}^{K} \left(\frac{\sup_{z \in A} |z - z_k|}{R - |z_k - \bar{z}|}\right)^{E+1}$$

This yields

$$\sup_{z \in A} \left\|\mathcal{T}^{[M/N]}(z) - \mathcal{T}(z)\right\|_W \leq C \prod_{k=1}^{K} \left(\frac{\sup_{z \in A} |z - z_k|}{R - |z_k - \bar{z}|}\right)^{E+1} \tag{32}$$

with C independent of E and K.

If we assume the correctness of such a bound, it is important to observe that the region of (exponential) convergence is *strictly* included in $B_R(\bar{z})$. Indeed, it can be found as

$$\left\{z \in U, \prod_{k=1}^{K} |z - z_k| < \prod_{k=1}^{K} (R - |z_k - \bar{z}|)\right\} \tag{33}$$

which does not include $\partial B_{R-\varepsilon}(\bar{z})$ for $\varepsilon > 0$ small enough (except for the trivial case $K = 1$).

We refer to Section 6.6 for a numerical validation of this bound.

6 Numerical examples

We report some numerical experiments on the problems and the algorithms described in the previous sections. In particular, we use the convention $\mathbf{x} = (x, y)^T$.

6.1 Check of the theoretical bound on the solution map of the Helmholtz equation

We consider the Helmholtz problem (1) with homogeneous Dirichlet boundary conditions, with $\Omega = [0, \pi]^2$. Let $\mathbf{k} = \sqrt{12}\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)^T \in \mathbb{R}^2$. The forcing term f is such that the (unique) solution of (1) with $z = 12$ is

$$\widetilde{u}(\mathbf{x}) = \frac{16}{\pi^4} xy(\pi - x)(\pi - y)e^{-i\mathbf{k}^T\mathbf{x}}$$

i.e. the product between a quadratic bubble vanishing on $\partial\Omega$ and a plane wave with wavenumber $\mathbf{k}$ (see Figure 3).

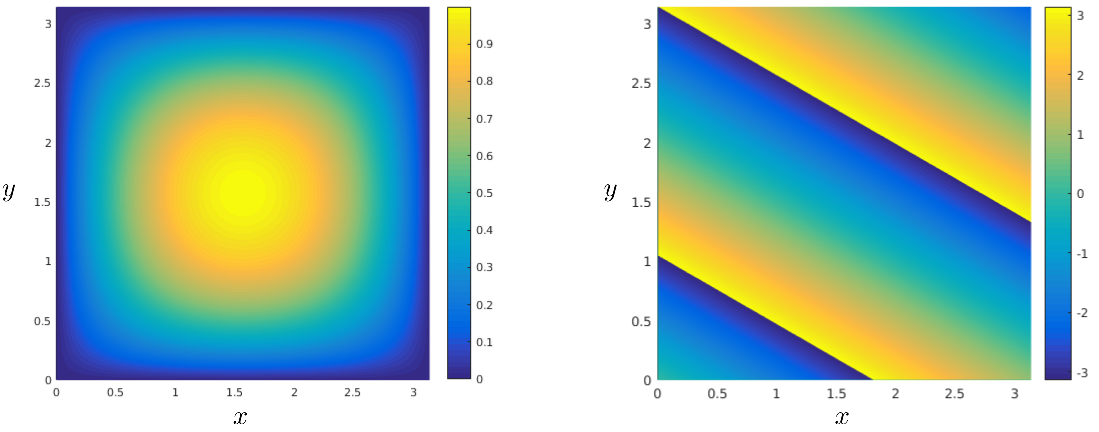

Figure 3: Magnitude (left) and phase (right) of the solution of problem (1) with the parameters described above and $z = 12$.

Let $w = 3$. We want to verify bound (6) for $z \in [4, 14]$. The solution map $\mathcal{S}(z)$ is approximated over 200 sample points with $\mathbb{P}_3$-Finite Elements using FreeFem++ (see Code 1).

The results are shown in Figure 4, and agree with Theorem 2.3. Moreover, some of the poles of $\mathcal{S}$ (5, 8, 10 and 13) can be identified.

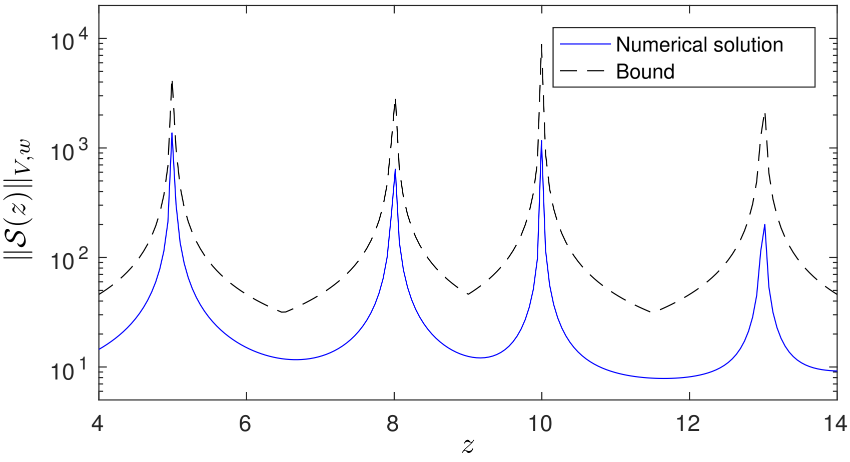

Figure 4: Norm of the solution of the Helmholtz problem with respect to z (blue). In black the bound given by Theorem 2.3.

6.2 An eigenvalue of a scattering problem in a square

We consider problem (11) with the domain and the boundary conditions described in Section 3.2 for $L = 1$. The forcing term f is such that the (unique) solution of (11) with $k = k_0 = \sqrt{12}$ is $\widetilde{u}(\mathbf{x}) = 4y(1-y)e^{-ik_0 x(1-x)}$ (see Figure 5).

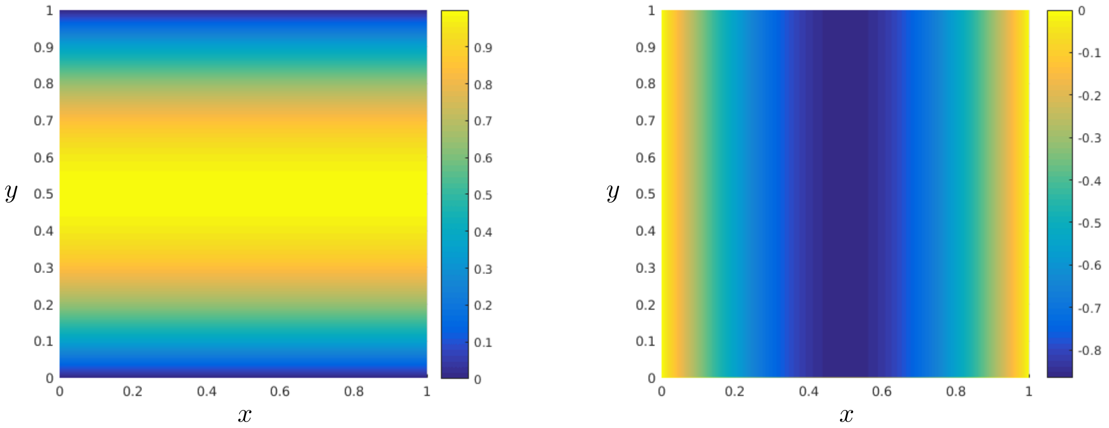

Figure 5: Magnitude (left) and phase (right) of the numerical solution of problem (11) with the parameters described above and $k = \sqrt{12}$.

In Section 3.2 we have shown that this problem can be ill-posed only for k belonging to a discrete set K. Let $\mathcal{S}_{sc} : (\mathbb{C} \setminus K) \to V_{\Gamma_D}$ denote the map which associates a value of k with the corresponding (unique) solution of problem (11), with the domain and forcing term described above.

In order to identify one element of K, we want to find an intersection between ζ_0^- and γ_1^-, which we define similarly to (18) in the case $k = k^-(\mu)$.

Through an approximate representation of these curves, shown in Figure 6 (left), we can deduce that $\widehat{\mu} = -2.388 - 1.493i$ is close to an element of $\zeta_0^- \cap \gamma_1^-$. The corresponding approximation for an element of K is $\widehat{k} = k^-(\widehat{\mu}) = -3.7723 - 0.9454i$.

The solution of (11) is approximated with $\mathbb{P}_3$-Finite Elements using FreeFem++ (see Code 2), for some values of k near $\widehat{k}$. The norm of the numerical solution (for $w = k_0$) is shown in Figure 6 (right). The plot appears to confirm the presence of a pole of $\mathcal{S}_{sc}$ near $\widehat{k}$.

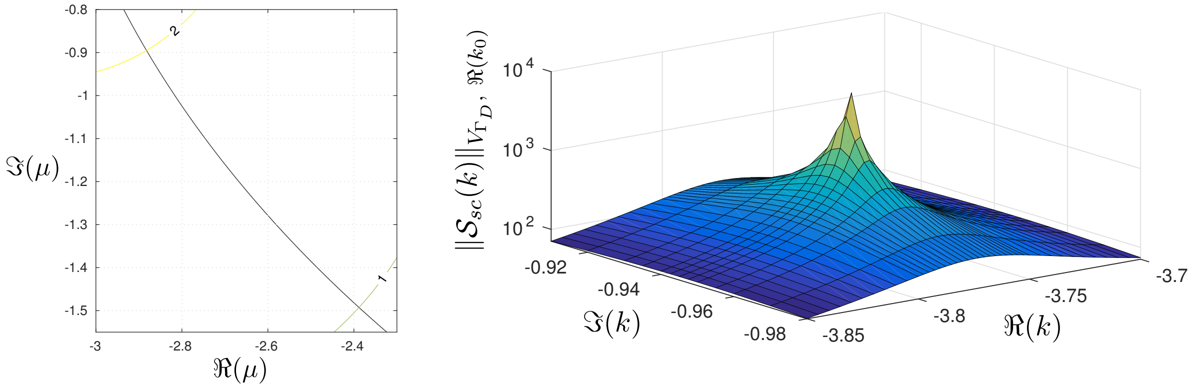

Figure 6: On the left, plot of ζ_0^- (black), γ_1^- (green) and γ_2^- (yellow). The value $\widehat{\mu}$ is an approximation of the intersection between ζ_0^- and γ_1^-. On the right, the (weighted) V_{Γ_D}-norm of $\mathcal{S}_{sc}$ for values of k belonging to a grid centered in $\widehat{k}$.

6.3 The role of ρ in the approximation of the solution map of the Helmholtz equation

As in Section 6.1, we consider the Helmholtz problem (1) with homogeneous Dirichlet boundary conditions, with $\Omega = [0,\pi]^2$. Again, the forcing term f is such that the (unique) solution of (1) with $z = 12$ is

$$\widetilde{u}(\mathbf{x}) = \frac{16}{\pi^4} xy(\pi - x)(\pi - y)e^{-i\mathbf{k}^T\mathbf{x}}$$

where $\mathbf{k} = \sqrt{12}\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)^T$.

Let $z_0 = 9.5 + 0.5i$, $w^2 = \Re(z_0)$, and $z_{tar} = 11.5$. We define (see Figure 7):

- r, the distance between z_0 and the closest pole of $\mathcal{S}$ ($r = |z_0 - 10|$).
- ρ_0, the distance between z_0 and z_{tar}.
- R, the distance between z_0 and the third closest pole of $\mathcal{S}$ ($R = |z_0 - 13|$).

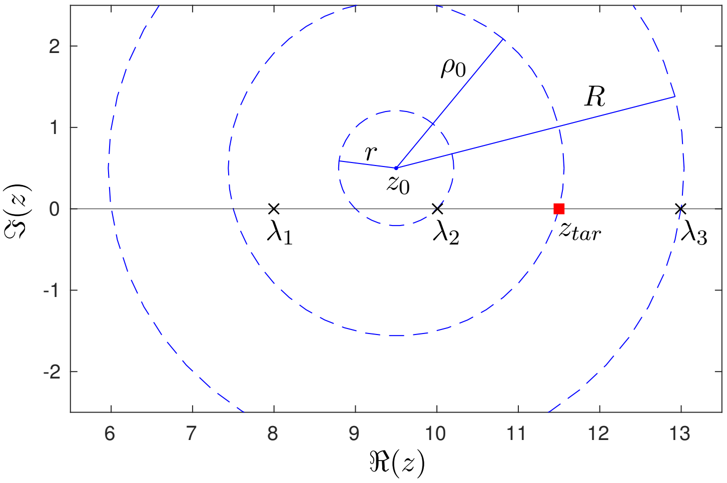

Figure 7: Representation of some of the quantities of interest in the complex plane. $\{\lambda_n\}_{n=1}^3$ is a subset of Λ, which contains all the poles of $\mathcal{S}$.

We want to approximate $\mathcal{S}$ with a $[M/N]$-Padé approximant centered in z_0 (see Definition 4.1) by applying Algorithm 1 with $N = 2$ and $E = M + N$ (see Code 3). In particular, the external library `LAPACK` (see e.g. [21]) is used to solve the eigenvalue problems involving $G_E^{[M/N]}$. Our aim is to compare the results for different choices of ρ. From Theorem 4.1, we expect $\rho = \rho_0$ to lead to the optimal rate of convergence. Also, convergence is not guaranteed for $\rho \geq R$.

The results are shown in Figure 8 for some values of M. The norm of the error for $z = z_{tar}$ does not appear to depend significantly on ρ: for a fixed M, the maximum difference between norms is approximately 10^{-10}. Moreover, the approximant converges to the real function also for $\rho = 2\rho_0 > R$: the error in $z = z_{tar}$ appears to be bounded by $\left(\frac{|z_0 - z_{tar}|}{R}\right)^M$ and not by $\left(\frac{\rho}{R}\right)^M$. The results of Section 4.5 can explain why the condition number of $G_E^{[M/N]}$ increases with ρ: this is due to the fact that more weight is given to worse-conditioned Gram sub-matrices.

6.4 Comparison of the accuracy of two algorithms for the approximation of the solution map of the Helmholtz equation

We consider the same problem and parameters introduced in Section 6.3. We want to compare the standard $[M/N]$-Padé approximant (see Definition 4.1) and its fast version (see Definition 4.2), centering the approximant in z_0 in both cases.

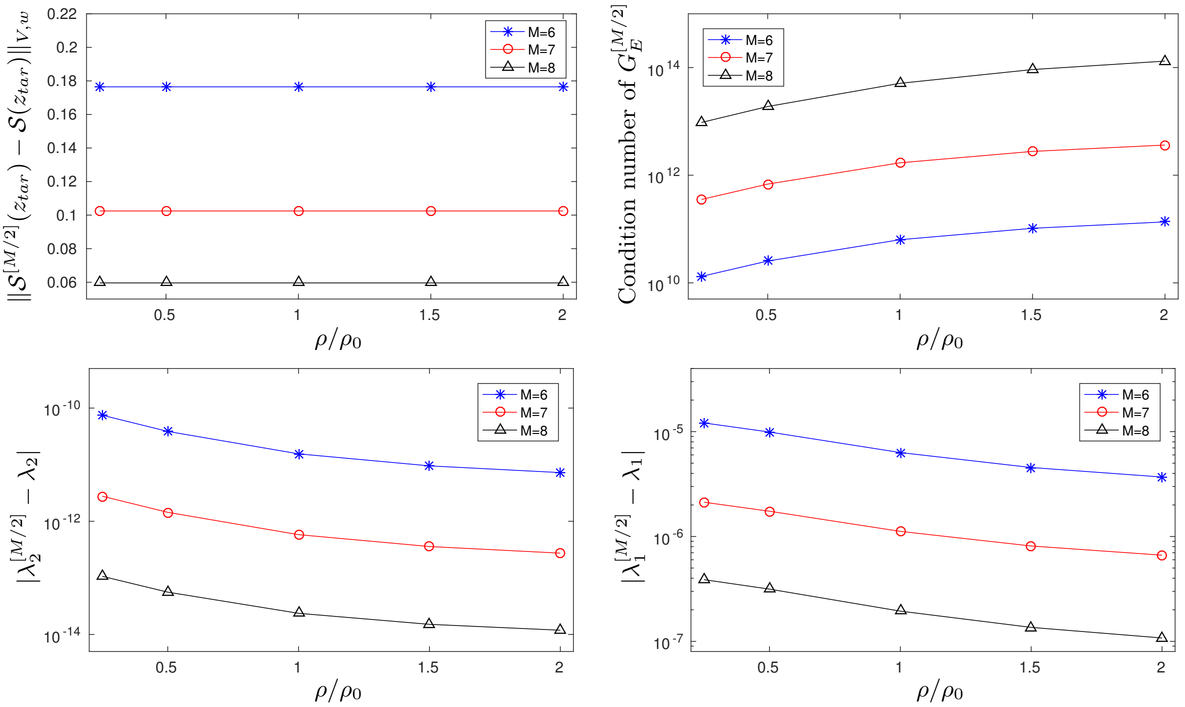

Figure 8: Comparison between the results obtained with a $[M/2]$-Padé approximant with different values of ρ: norm of the error in z_{tar} (top left), condition number of the matrix $G_E^{[M/2]}$ (top right), error in the approximation of λ_2 (bottom left) and of λ_1 (bottom right).

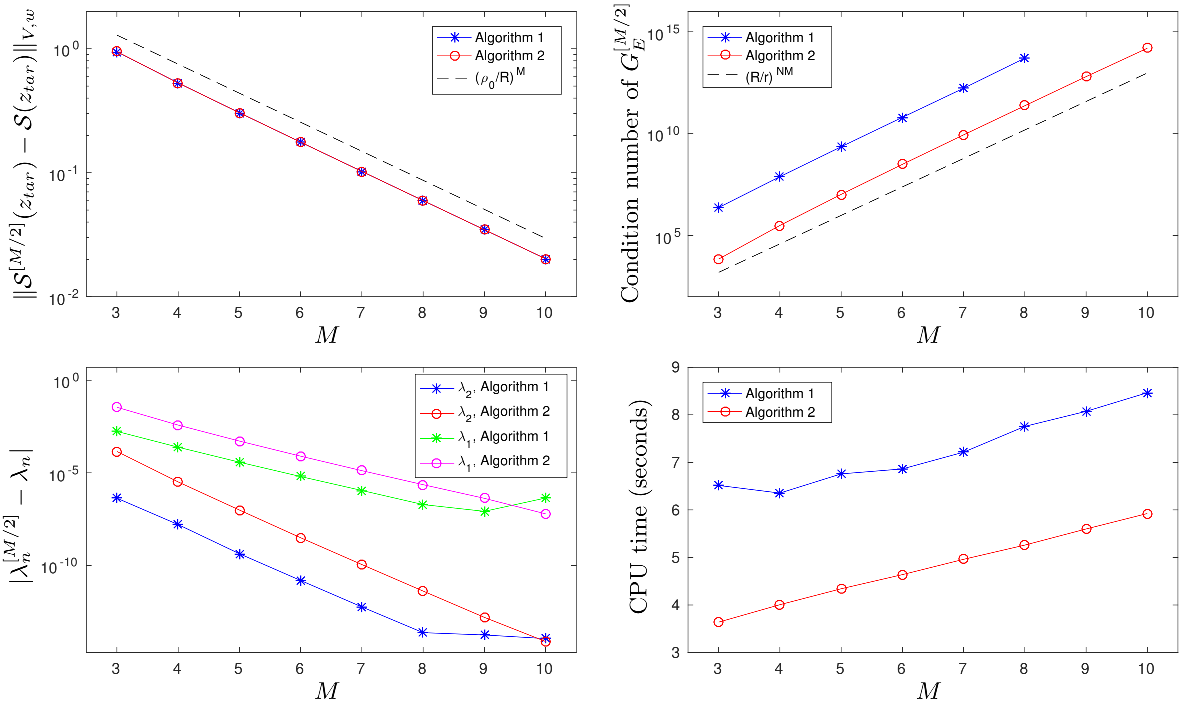

Figure 9: Convergence results for standard and fast $[M/2]$-Padé approximants: norm of the error in z_{tar} (top left), condition number of the matrix $G_E^{[M/2]}$ (top right), error in the approximation of λ_n for $n = 1, 2$ (bottom left) and computation time (bottom right).

We fix $N = 2$. For the computation of the standard approximant we use Algorithm 1 with $E = M + N$ and $\rho = \rho_0$ (see Code 3), whereas in computing the fast approximant we use Algorithm 2 with $E = M$ (see Code 4). The external library `LAPACK` is used to solve the eigenvalue problems involving $G_E^{[M/N]}$. In particular, observe that, for a fixed M, the computation of the standard Padé approximant requires the calculation of two more derivatives of the solution map.

The results are shown in Figure 9 for some values of M. The convergence rate of the norm of the error for $z = z_{tar}$ does not appear to depend significantly on the choice of the algorithm: in both cases the optimal convergence rate $\left(\frac{\rho_0}{R}\right)^M$ appears to be achieved.

For a fixed M, the condition number of $G_E^{[M/N]}$ appears much smaller for the fast approximant. Indeed, to compute the standard Padé approximant, the Gram sub-matrix of order M is averaged with two other worse-conditioned sub-matrices, whereas only the sub-matrix of order M is necessary for the fast approximant. Moreover, the growth rate of the condition number appears to be $\left(\frac{R}{r}\right)^{NM}$.

For $M \in \{9, 10\}$, the condition number of $G_E^{[M/N]}$ for the standard approximant is higher than the reciprocal of the machine epsilon: in this case the round-off errors are responsible for the lack of accuracy in the approximation of the poles of $\mathcal{S}$.

As expected, for a fixed $M \leq 8$, the approximation of the poles provided by the standard approximant is better, since it uses more information than the fast approximant.

However, if we compare the results with respect to the same E, i.e. by keeping fixed the number of known derivatives of the solution map, the fast approximant performs much better than the standard one in terms of the norm of the error, slightly better in terms of the approximation of the poles, and slightly worse in terms of the condition number.

Additionally, we can see that the CPU time needed for the computation of the fast Padé approximant is much smaller than the one needed for the computation of the standard approximant.

6.5 Single-pole approximation for the solution map of the Helmholtz equation

We consider the same problem and parameters introduced in Section 6.3. We want to apply Algorithm 3 to approximate the pole of $\mathcal{S}$ which is closer to z_0, i.e. $\lambda_2 = 10$.

In the FreeFem++ implementation (see Code 5), $\mathbb{P}_3$-Finite Elements are used to find an approximation of the exact solution. The results for several values of E are shown in Figure 10.

As expected, the convergence rate appears to be $\left(\frac{|z_0-\lambda_2|}{|z_0-\lambda_1|}\right)^{2E}$, with $\lambda_1 = 8$ being the second closest pole of $\mathcal{S}$.

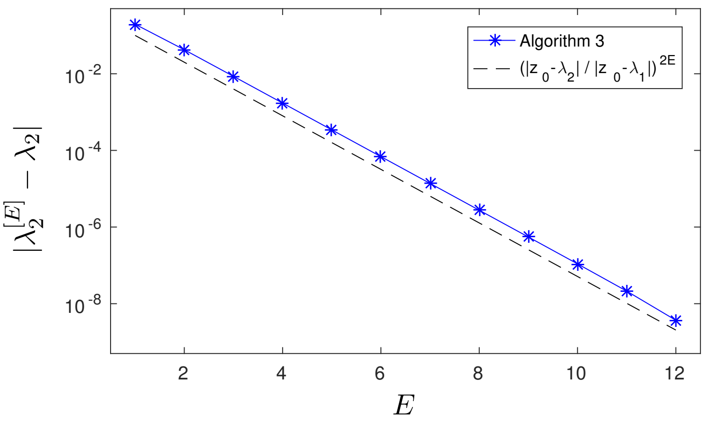

Figure 10: Error in the approximation of λ_2 for different values of E.

6.6 Multi-point Padé approximants for the solution map of the Helmholtz equation

We consider the same problem introduced in Section 6.3. Let $z_1 = 9.5 + 0.5i$, $z_2 = 8 + 0.5i$, and $z_3 = 11 + 0.5i$ be sample points (see Figure 11), with corresponding weights $w_1 = 0.5$ and $w_2 = w_3 = 0.25$. The corresponding weighted average (see Definition 5.1) is $\bar{z} = z_1$.

We define $w^2 = \Re(\bar{z})$, and R as the distance between $\bar{z}$ and the third closest pole of $\mathcal{S}$ ($R = |\bar{z} - \lambda_3|$). The target point is $z_{tar} = 11.5$.

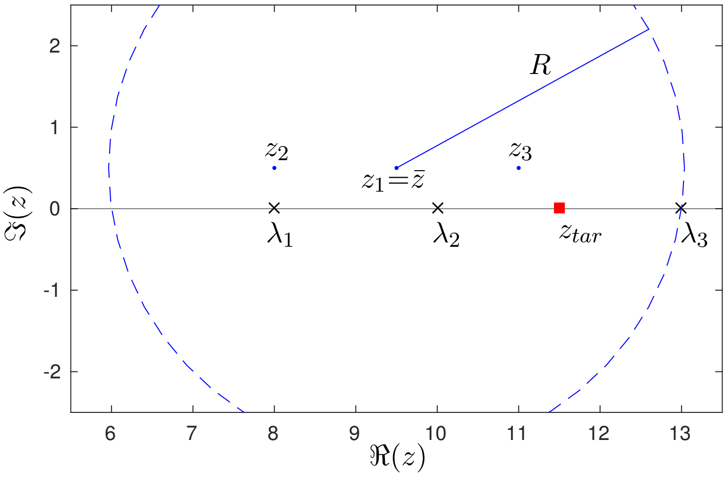

Figure 11: Representation of some of the quantities of interest in the complex plane. $\{\lambda_n\}_{n=1}^3$ is a subset of the set Λ, which contains all the (simple) poles of $\mathcal{S}$.

We want to approximate $\mathcal{S}$ with a multi-point $[M/N]$-Padé approximant based on $(\{z_1, z_2, z_3\}, \{w_1, w_2, w_3\})$ (see Definition 5.1) by applying Algorithm 4 (see Code 6), with $N = 2$ and $E = \frac{M+1}{3} - 1$.

The results are shown in Figure 12 for some values of M. The convergence of the error norm for $z = z_{tar}$ appears to follow the predicted rate (32). However, the magnitude of the error norm is higher than the one obtained with the standard (single-point) Padé approximant for similar values of M.

The condition number of the Gram sub-matrix $G_E^{[M/N]}$ is much lower than for the standard approximant, since only $\frac{M+1}{3}$ derivatives are considered at each sample point instead of $M + 1$. Similarly to the single-point Padé approximant, it appears that the condition number grows at the rate of approximately $\left(\frac{R}{|\bar{z}-\lambda_2|}\right)^{NE}$.

The error in the approximation of the poles appears to converge to zero, but is higher than the error achieved with the standard approximant for similar values of M.

The CPU time needed for the algorithm is quite high, and appears to increase at a super-linear rate. This can be explained by the fact that the computation of the numerator $P^{[M/N]}$, which is carried out through the use of a divided differences table (see e.g. [20]), requires $\mathcal{O}(M^2)$ operations involving $\mathbb{P}_3$ functions on the computational mesh.

6.7 Convergence results

We want to compare the convergence results for the fast $[M/N]$-Padé approximant (see Definition 4.2) and for the multi-point $[M/N]$-Padé approximant (see Definition 5.1).

We consider the problem introduced in Section 6.3. For the parameters of the approximants, see Sections 6.4 and 6.6. The results are shown in Figure 13 for several values of N and E.

In particular, observe that we plot the error norm against E, i.e. the number of known derivatives of $\mathcal{S}$ in each sample point. We remind that E coincides with M (the degree of the numerator) for fast Padé approximants, whereas $M = 3E + 2$ for multi-point approximants.

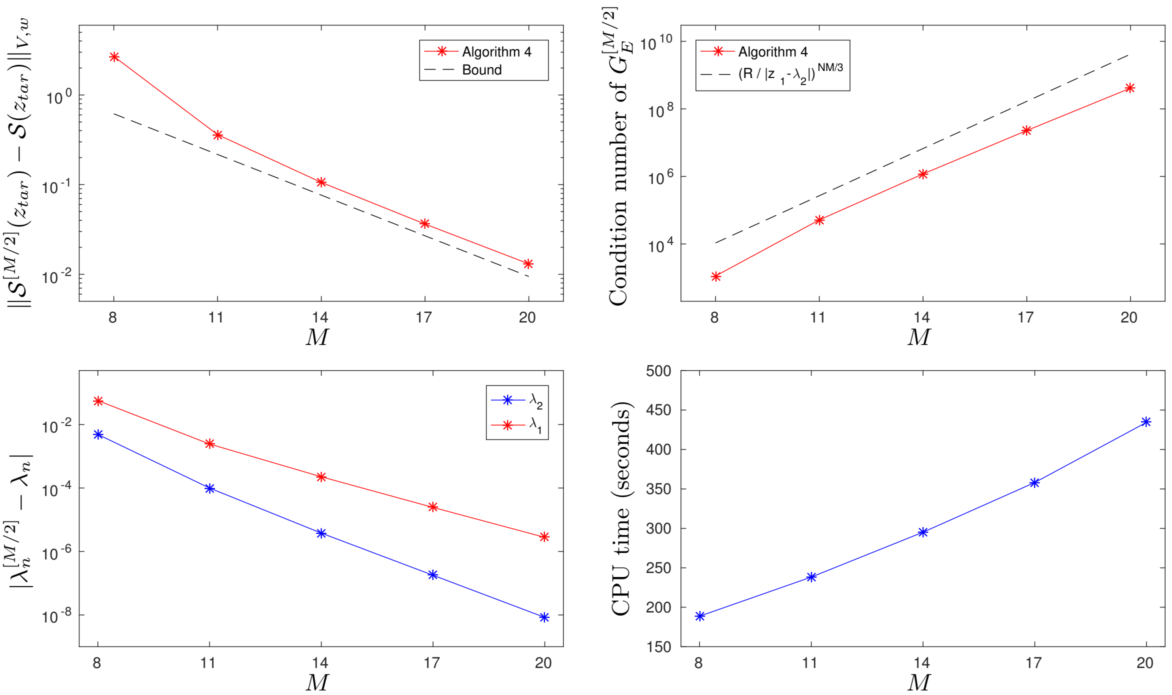

Figure 12: Convergence results for the multi-point $[M/2]$-Padé approximant: norm of the error in z_{tar} (top left), condition number of the matrix $G_E^{[M/2]}$ (top right), error in the approximation of λ_n, for $n = 1, 2$ (bottom left) and computation time (bottom right).

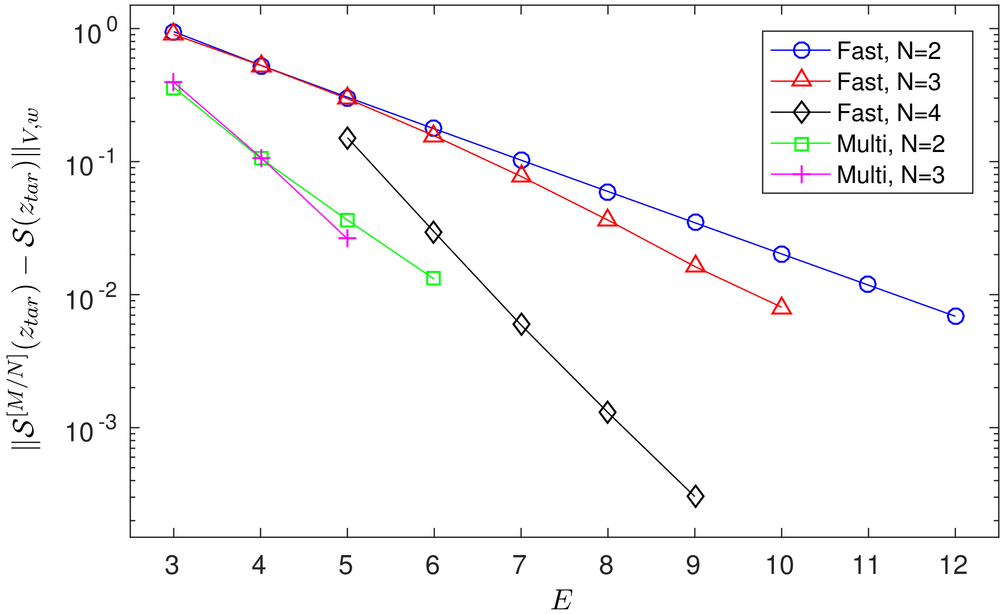

Figure 13: Convergence results for fast $[M/N]$-Padé approximants and multi-point $[M/N]$-Padé approximants.

6.8 Comparison of the solution norm

We compare the approximations of the norm of the solution map $\mathcal{S}$ obtained through several algorithms.

We consider the problem introduced in Section 6.3. For the parameters of the approximants, see Sections 6.4 and 6.6. The results are shown in Figure 14 for 100 sample points in the interval $[6, 13)$.

In particular, the plot appears to confirm that the multi-point approximant has a smaller convergence region than the single-point approximant. Indeed, the intersection between the predicted convergence region (33) and the real axis is approximately the interval $(6.823, 12.18)$, whereas for the standard approximant the real convergence interval is $(6, 13)$.

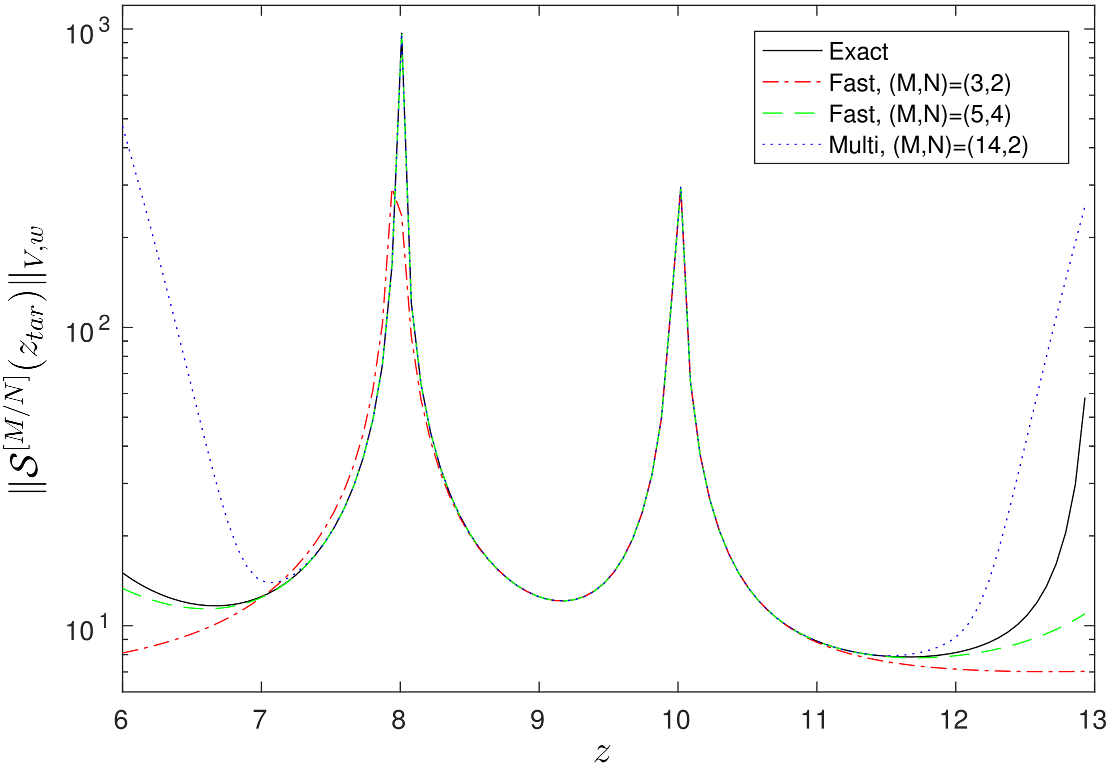

Figure 14: Frequency response of the Helmholtz problem with homogeneous Dirichlet conditions. Direct evaluation of $\mathcal{S}$ (black), fast [3/2]-Padé approximant (red), fast [5/4]-Padé approximant (green), and multi-point [14/2]-Padé approximant (blue).

7 Conclusion and outlook

In this report, we have proven that the solution map of the Helmholtz problem, endowed with Dirichlet boundary conditions, is meromorphic in $\mathbb{C}$. Its simple poles correspond to the eigenvalues of the Laplacian operator with homogeneous Dirichlet boundary conditions, and form a discrete set with no finite accumulation point.

Moreover, we have shown that the Bohr-Sommerfield radiation condition on a portion of the boundary makes the Helmholtz problem well-posed for any wavenumber with non-negative imaginary part. We have verified that the poles of the solution map still form a discrete set for a specific example. Further studies may ascertain whether this is true in general, and if the solution map is still meromorphic.

We have described an algorithm for the computation of a rational approximant of any Hilbert space-valued meromorphic function, whose rate of convergence is exponential. We have shown that a simplified (and more efficient) version of the algorithm performs better than the original one, when applied to the Helmholtz problem, and in general to functions whose poles have order one. However, a formal proof of the convergence rate of this algorithm is still missing.

Moreover, we provided a definition (and a corresponding algorithm) for multi-point Padé approximants, for which we have also guessed a convergence rate. Despite being confirmed by a numerical experiment, the convergence rate (and even the convergence of the approximant itself) remains to be proven.

Many of the properties described in the report were confirmed through numerical experiments. We have shown the presence of a trade-off between accuracy of the approximation, size of the convergence region, condition number of the matrix appearing in the eigenvalue problem, and computation time.

In particular, the condition number of the Gram sub-matrix $G_E^{[M/N]}$ appears to be the most important issue, since it provides a strict upper bound for the numerator degree M. The envisioned solutions involve using multi-point approximants, approximating the poles through different means, or exploiting special properties of the target function (e.g. the fact that the poles are real).

8 FreeFem++ code segments

```
load "Element_P3"

/// parameters
real k0 = sqrt(12), d1 = cos(pi / 3), d2 = sin(pi / 3), w = 3;

/// forcing term
func wave     = exp(- 1i * k0 * (x * d1 + y * d2));
real bnorm    = 16 / pi^4;
func bubble   = bnorm * x * y * (x - pi) * (y - pi);
func bubblex  = bnorm * (2 * x * y^2 - 2 * pi * x * y - pi * y^2 + pi^2 * y);
func bubbley  = bnorm * (2 * x^2 * y - 2 * pi * x * y - pi * x^2 + pi^2 * x);
func bubblexx = bnorm * 2 * (y^2 - pi * y);
func bubbleyy = bnorm * 2 * (x^2 - pi * x);
func uex = wave * bubble;
func f = wave * (2i * k0 * (d1 * bubblex + d2 * bubbley) - (bubblexx + bubbleyy));

/// Finite Elements space
mesh Th = square(80, 80, [pi * x, pi * y]);
fespace Vh(Th, P3);
Vh<complex> uh, vh;

/// sequilinear form
complex zz = 9 + .5i;
varf a(u,v) = int2d(Th, qforder = 8)(dx(u) * conj(dx(v)) + dy(u) * conj(dy(v)))
            - int2d(Th, qforder = 8)(zz * u * conj(v))
            + on(1, 2, 3, 4, u = 0);

/// solve problem
solve Helmholtz(uh, vh) = a - int2d(Th, qforder = 8)(f * conj(vh));
plot(uh, wait = 1, fill = 1, value = 1, cmm = "Numerical solution for z = " + zz);

/// compute solution norm
real solNorm = sqrt(int2d(Th, qforder = 8)(abs(dx(uh))^2  + abs(dy(uh))^2)
                  + w^2 * int2d(Th, qforder = 8)(abs(uh)^2));
cout << "Solution norm: " << solNorm << endl;
```

Code 1. Numerical resolution of the Helmholtz equation with homogeneous boundary conditions. The parameters are the ones described in Section 6.1. Moreover, the weighted V-norm of the solution is computed.

```
load "Element_P3"

/// parameters
real w = sqrt(12), k0 = sqrt(12);

/// forcing term
func wave     = exp(- 1i * k0 * x * (1 - x));
func bubble   = 4 * y * (1 - y);
func wavex    = - 1i * k0 * (1 - 2 * x) * wave;
func bubbley  = 4 * (1 - 2 * y);
func wavexx   = (- k0^2 * (1 - 2 * x)^2 + 2i * k0) * wave;
func bubbleyy = - 8;
func uex = wave * bubble;
func f = - wavexx * bubble - bubbleyy * wave - k0^2 * uex;

/// Finite Elements space
mesh Th = square(80, 80);
fespace Vh(Th, P3);
Vh<complex> uh, vh;

/// sequilinear form
complex k = - 3.7723 - 0.9454i;
varf a(u,v) = int2d(Th, qforder = 8)(dx(u) * conj(dx(v)) + dy(u) * conj(dy(v)))
            - int2d(Th, qforder = 8)(k^2 * u * conj(v))
            - int1d(Th, 2, 4, qforder = 8)(1i * k * u * conj(v))
            + on(1, 3, u = 0);

```

```
/// solve problem
solve Helmholtz(uh, vh) = a - int2d(Th, qforder = 8)(f * conj(vh));
plot(uh, wait = 1, fill = 1, value = 1, cmm = "Numerical solution for k = " + k);

/// compute solution norm
real solNorm = sqrt(int2d(Th, qforder = 8)(abs(dx(uh))^2  + abs(dy(uh))^2)
                  + w^2 * int2d(Th, qforder = 8)(abs(uh)^2));
cout << "Solution norm: " << solNorm << endl;
```

Code 2. Numerical resolution of the scattering problem on a square. The parameters are the ones described in Section 6.2. Moreover, the weighted V_{Γ_D}-norm of the solution is computed.

```
load "Element_P3"
load "lapack" /// solver for eigenvalue problems

/// parameters
real k0 = sqrt(12), d1 = cos(pi / 3), d2 = sin(pi / 3);

/// forcing term
func wave     = exp(- 1i * k0 * (x * d1 + y * d2));
real bnorm    = 16 / pi^4;
func bubble   = bnorm * x * y * (x - pi) * (y - pi);
func bubblex  = bnorm * (2 * x * y^2 - 2 * pi * x * y - pi * y^2 + pi^2 * y);
func bubbley  = bnorm * (2 * x^2 * y - 2 * pi * x * y - pi * x^2 + pi^2 * x);
func bubblexx = bnorm * 2 * (y^2 - pi * y);
func bubbleyy = bnorm * 2 * (x^2 - pi * x);
func uex = wave * bubble;
func f = wave * (2i * k0 * (d1 * bubblex + d2 * bubbley) - (bubblexx + bubbleyy));

/// Finite Elements space
mesh Th = square(80, 80, [pi * x, pi * y]);
fespace Vh(Th, P3);
Vh<complex> uh, vh;

/// sequilinear form
complex zz;
varf a(u,v) = int2d(Th, qforder = 8)(dx(u) * conj(dx(v)) + dy(u) * conj(dy(v)))
            - int2d(Th, qforder = 8)(zz * u * conj(v))
            + on(1, 2, 3, 4, u = 0);

/// approximant parameters
int N = 2, M = 6, E = M + N;
complex z0 = 9.5 + .5i, ztar = 11.5;
real rho = abs(z0 - ztar), w = sqrt(real(z0));

/// find Taylor series of S(z)
Vh<complex>[int] T(E + 1);
zz = z0;
Vh<complex> rhs = f;
for(int i = 0; i <= E; i++)
{
  solve Helmoltz(uh, vh) = a - int2d(Th, qforder = 8)(rhs * conj(vh));
  T[i] = uh;
  rhs = uh;
}

/// build matrix
complex[int, int] A(N + 1, N + 1); A = 0.;
for(int alpha = max(M + 1, N); alpha <= E; alpha++)
{
  real weight = rho ^ (2 * alpha);
  for(int i = alpha - N; i <= alpha; i++)
    for(int j = i; j <= alpha; j++)
      A(alpha - j, alpha - i) +=
            weight * (int2d(Th, qforder = 8)(dx(T[i]) * conj(dx(T[j]))
                                           + dy(T[i]) * conj(dy(T[j])))
                    + w^2 * int2d(Th, qforder = 8)(T[i] * conj(T[j])));
}
for(int i = 0; i <= N; i++)
  for(int j = 0; j <= i - 1; j++)
  A(i, j) = conj(A(j, i));

/// find smallest eigenvector
complex[int] ev(N + 1);
complex[int, int] eV(N + 1, N + 1);
int l = zgeev(A, ev, eV);
int imin = 0;
for(int i = 1; i <= N; i++)
 if(real(ev[i]) < real(ev[imin]))
  imin = i;
complex[int] QN = eV(:, imin);

/// build numerator
Vh<complex>[int] PM(M + 1);
for(int i = 0; i <= M; i++)
{
 PM[i] = 0;
 for(int j = 0; j <= min(i, N); j++)
  PM[i] = PM[i] + QN[j] * T[i - j];
}

/// evaluate Pade' approximant
Vh<complex> upade = PM[0];
for(int i = 1; i <= M; i++)
 upade = upade + PM[i] * (ztar - z0)^i;
complex d = QN[0];
for(int i = 1; i <= N; i++)
 d = d + QN[i] * (ztar - z0)^i;
upade = upade / d;

/// numerical solution in ztar
zz = ztar;
solve Helmoltz(uh, vh) = a - int2d(Th, qforder = 8)(f * conj(vh));

/// compute error norm
Vh<complex> err = uh - upade;
real errNorm = sqrt(int2d(Th, qforder = 8)(abs(dx(err))^2  + abs(dy(err))^2)
      + w^2 * int2d(Th, qforder = 8)(abs(err)^2));
cout << "Error norm: " << errNorm << endl;

plot(uh, wait = 1, fill = 1, value = 1, cmm = "Numerical sol for z = " + ztar);
plot(upade, wait = 1, fill = 1, value = 1, cmm = "Pade' approx for z = " + ztar);
plot(err, wait = 1, fill = 1, value = 1, cmm = "Error");
```

Code 3. Implementation of Algorithm 1. The parameters are the ones described in Section 6.3, with $M = 6$ and $\rho = \rho_0$.

```
load "Element_P3"
load "lapack" /// solver for eigenvalue problems

/// parameters
real k0 = sqrt(12), d1 = cos(pi / 3), d2 = sin(pi / 3);

/// forcing term
func wave     = exp(- 1i * k0 * (x * d1 + y * d2));
real bnorm    = 16 / pi^4;
func bubble   = bnorm * x * y * (x - pi) * (y - pi);
func bubblex  = bnorm * (2 * x * y^2 - 2 * pi * x * y - pi * y^2 + pi^2 * y);
func bubbley  = bnorm * (2 * x^2 * y - 2 * pi * x * y - pi * x^2 + pi^2 * x);
func bubblexx = bnorm * 2 * (y^2 - pi * y);
func bubbleyy = bnorm * 2 * (x^2 - pi * x);
func uex = wave * bubble;
func f = wave * (2i * k0 * (d1 * bubblex + d2 * bubbley) - (bubblexx + bubbleyy));

/// Finite Elements space
mesh Th = square(80, 80, [pi * x, pi * y]);
fespace Vh(Th, P3);
Vh<complex> uh, vh;

/// sequilinear form
```

```
complex zz;
varf a(u,v) = int2d(Th, qforder = 8)(dx(u) * conj(dx(v)) + dy(u) * conj(dy(v)))
            - int2d(Th, qforder = 8)(zz * u * conj(v))
            + on(1, 2, 3, 4, u = 0);

/// approximant parameters
int N = 2, M = 6, E = M;
complex z0 = 9.5 + .5i, ztar = 11.5;
real w = sqrt(real(z0));

/// find Taylor series of S(z)
Vh<complex>[int] T(E + 1);
zz = z0;
Vh<complex> rhs = f;
for(int i = 0; i <= E; i++)
{
  solve Helmoltz(uh, vh) = a - int2d(Th, qforder = 8)(rhs * conj(vh));
  T[i] = uh;
  rhs = uh;
}

/// build matrix
complex[int, int] A(N + 1, N + 1); A = 0.;
for(int i = E - N; i <= E; i++)
  for(int j = i; j <= E; j++)
    A(E - j, E - i) += int2d(Th, qforder = 8)(dx(T[i]) * conj(dx(T[j]))
                                            + dy(T[i]) * conj(dy(T[j])))
                     + w^2 * int2d(Th, qforder = 8)(T[i] * conj(T[j]));
for(int i = 0; i <= N; i++)
  for(int j = 0; j <= i - 1; j++)
    A(i, j) = conj(A(j, i));

/// find smallest eigenvector
complex[int] ev(N + 1);
complex[int, int] eV(N + 1, N + 1);
int l = zgeev(A, ev, eV);
int imin = 0;
for(int i = 1; i <= N; i++)
  if(real(ev[i]) < real(ev[imin]))
    imin = i;
complex[int] QN = eV(:, imin);

/// build numerator
Vh<complex>[int] PM(M + 1);
for(int i = 0; i <= M; i++)
{
  PM[i] = 0;
  for(int j = 0; j <= min(i, N); j++)
    PM[i] = PM[i] + QN[j] * T[i - j];
}

/// evaluate Pade' approximant
Vh<complex> upade = PM[0];
for(int i = 1; i <= M; i++)
  upade = upade + PM[i] * (ztar - z0)^i;
complex d = QN[0];
for(int i = 1; i <= N; i++)
  d = d + QN[i] * (ztar - z0)^i;
upade = upade / d;

/// numerical solution in ztar
zz = ztar;
solve Helmoltz(uh, vh) = a - int2d(Th, qforder = 8)(f * conj(vh));

/// compute error norm
Vh<complex> err = uh - upade;
real errNorm = sqrt(int2d(Th, qforder = 8)(abs(dx(err))^2  + abs(dy(err))^2)
            + w^2 * int2d(Th, qforder = 8)(abs(err)^2));
cout << "Error norm: " << errNorm << endl;

plot(uh, wait = 1, fill = 1, value = 1, cmm = "Numerical sol for z = " + ztar);
plot(upade, wait = 1, fill = 1, value = 1, cmm = "Pade’ approx for z = " + ztar);
plot(err, wait = 1, fill = 1, value = 1, cmm = "Error");
```

Code 4. Implementation of Algorithm 2. The parameters are the ones described in Section 6.4, with $M = 6$.

```
load "Element_P3"

/// parameters
real k0 = sqrt(12), d1 = cos(pi / 3), d2 = sin(pi / 3);
complex zz;

/// forcing term
func wave     = exp(- 1i * k0 * (x * d1 + y * d2));
real bnorm    = 16 / pi^4;
func bubble   = bnorm * x * y * (x - pi) * (y - pi);
func bubblex  = bnorm * (2 * x * y^2 - 2 * pi * x * y - pi * y^2 + pi^2 * y);
func bubbley  = bnorm * (2 * x^2 * y - 2 * pi * x * y - pi * x^2 + pi^2 * x);
func bubblexx = bnorm * 2 * (y^2 - pi * y);
func bubbleyy = bnorm * 2 * (x^2 - pi * x);
func uex = wave * bubble;
func f = wave * (2i * k0 * (d1 * bubblex + d2 * bubbley) - (bubblexx + bubbleyy));

/// Finite Elements space
mesh Th = square(80, 80, [pi * x, pi * y]);
fespace Vh(Th, P3);
Vh<complex> uh, vh;

/// sequilinear form
varf a(u,v) = int2d(Th, qforder = 8)(dx(u) * conj(dx(v)) + dy(u) * conj(dy(v)))
            - int2d(Th, qforder = 8)(zz * u * conj(v))
            + on(1, 2, 3, 4, u = 0);

/// approximation parameters
int N = 2, E = 6;
complex z0 = 9.5 + .5i;
real w = sqrt(real(z0));

/// find Taylor series of S(z)
Vh<complex> Told, Tnew = f;
zz = z0;
for(int i = 0; i <= E; i++)
{
  Told = Tnew;
  solve Helmoltz(uh, vh) = a - int2d(Th, qforder = 8)(Told * conj(vh));
  Tnew = uh;
}

/// compute approximation
complex num = int2d(Th, qforder = 8)(dx(Told) * conj(dx(Tnew))
                                   + dy(Told) * conj(dy(Tnew)))
            + w^2 * int2d(Th, qforder = 8)(Told * conj(Tnew));
complex den = int2d(Th, qforder = 8)(abs(dx(Tnew))^2
                                   + abs(dy(Tnew))^2)
            + w^2 * int2d(Th, qforder = 8)(abs(Tnew)^2);
complex pole = z0 + num / den;
cout << "Approximation of the closest pole: " << pole << endl;
```

Code 5. Implementation of Algorithm 3. The parameters are the ones described in Section 6.5, with $E = 6$.

```
load "Element_P3"
load "lapack" /// solver for eigenvalue problems

/// parameters
real k0 = sqrt(12), d1 = cos(pi / 3), d2 = sin(pi / 3);

/// forcing term
```

```
func wave     = exp(- 1i * k0 * (x * d1 + y * d2));
real bnorm    = 16 / pi^4;
func bubble   = bnorm * x * y * (x - pi) * (y - pi);
func bubblex  = bnorm * (2 * x * y^2 - 2 * pi * x * y - pi * y^2 + pi^2 * y);
func bubbley  = bnorm * (2 * x^2 * y - 2 * pi * x * y - pi * x^2 + pi^2 * x);
func bubblexx = bnorm * 2 * (y^2 - pi * y);
func bubbleyy = bnorm * 2 * (x^2 - pi * x);
func uex = wave * bubble;
func f = wave * (2i * k0 * (d1 * bubblex + d2 * bubbley) - (bubblexx + bubbleyy));

/// Finite Elements space
mesh Th = square(80, 80, [pi * x, pi * y]);
fespace Vh(Th, P3);
Vh<complex> uh, vh;

/// sequilinear form
complex zz;
varf a(u,v) = int2d(Th, qforder = 8)(dx(u) * conj(dx(v)) + dy(u) * conj(dy(v)))
            - int2d(Th, qforder = 8)(zz * u * conj(v))
            + on(1, 2, 3, 4, u = 0);

/// approximant parameters
int N = 2, M = 17;
complex[int] zsample = [9.5 + 0.5 * 1i, 8 + 0.5 * 1i, 11 + 0.5 * 1i];
real[int] weights = [.5, .25, .25];
complex z0 = zsample[0], ztar = 11.5;
real w = sqrt(real(z0));
int E = ceil((M + 1) / zsample.n) - 1, DTsize = zsample.n * (E + 1);

/// helper functions
func int shift(int i, int j, int c) { return c * i + j; }
func real fact(int n)
{
  real f = 1.;
  for(int i = 2; i <= n; i++) f *= i;
  return f;
}
func real binom(int n, int k){ return fact(n) / fact(k) / fact(n - k); }

/// find Taylor series of S(z)
Vh<complex>[int] DTs(DTsize);
Vh<complex> rhs;
for(int s = 0; s < zsample.n; s++) // sample points
{
  zz = zsample[s];
  rhs = f;
  for(int d = 0; d <= E; d++) // derivatives (i.e. columns)
  {
    solve Helmoltz(uh, vh) = a - int2d(Th, qforder = 8)(rhs * conj(vh));
    DTs[shift(s, d, E + 1)] = uh;
    rhs = uh;
  }
}

/// effective sample points (each sample is repeated E+1 times)
complex[int] zsmpleff(DTsize);
for(int i = 0; i < zsample.n; i++)
  zsmpleff((i * (E + 1)):((i + 1) * (E + 1) - 1)) = zsample[i];

/// build matrix
complex[int, int] A(N + 1, N + 1);
A = 0.;
/// add contribution in each sample point
for(int s = 0; s < zsample.n; s++)
{
  complex[int, int] Aloc(N + 1, N + 1);

  for(int i = E - N; i <= E; i++)
    for(int j = i; j <= E; j++)
      Aloc(E - j, E - i) = int2d(Th, qforder = 8)(
                 dx(DTs[shift(s, i, E + 1)]) * conj(dx(DTs[shift(s, j, E + 1)]))
```

```
               + dy(DTs[shift(s, i, E + 1)]) * conj(dy(DTs[shift(s, j, E + 1)])))
               + w^2 * int2d(Th, qforder = 8)(
                 DTs[shift(s, i, E + 1)] * conj(DTs[shift(s, j, E + 1)]));
  for(int i = E - N; i <= E; i++)
    for(int j = i + 1; j <= E; j++)
      Aloc(E - i, E - j) = conj(Aloc(E - j, E - i));

  complex[int, int] AlocShifted(N + 1, N + 1);
  if(zsample[s] == z0)
    AlocShifted = Aloc;
  else{
    AlocShifted = 0;
    for(int i = 0; i <= N; i++)
      for(int j = 0; j <= N; j++)
        for(int l = 0; l <= i; l++)
          for(int k = 0; k <= j; k++)
            AlocShifted(i, j) += binom(i, l) * conj(zsample[s] - z0)^(i - l)
                  * Aloc(l, k) * binom(j, k) * (zsample[s] - z0)^(j - k);
  }

  A += weights[s] * AlocShifted;
}

/// find smallest eigenvector
complex[int] ev(N + 1);
complex[int, int] eV(N + 1, N + 1);
int l = zgeev(A, ev, eV);
int imin = 0;
for(int i = 1; i <= N; i++)
  if(real(ev[i]) < real(ev[imin]))
    imin = i;
complex[int] QN = eV(:, imin);

/// table containing coefficients of derivatives of Q
complex[int, int] Qtable(N + 1, N + 1);
Qtable(0, :) = QN;
for(int i = 1; i <= N; i++) // Q derivative order
  for(int j = 0; j <= N - i; j++) // poly degree
    Qtable(i, j) = Qtable(i - 1, j + 1) * (j + 1);
/// function for computation of Taylor coeffs of Q
func complex Qtaylor(complex zz, int order)
{
  if(order > N) return 0;
  complex res = Qtable(order, 0);
  for(int i = 1; i <= N - order; i++)
    res += Qtable(order, i) * (zz - z0) ^ i;
  return res / fact(order);
}

/// find Hermite series for S*Q through divided differences table
Vh<complex>[int] difftable(DTsize ^ 2);
Vh<complex> aux;

/// fill initial triangles with Taylor coefficients
for(int s = 0; s < zsample.n; s++) // sample points
{
  zz = zsample[s];
  for(int c = 0; c <= E; c++) // derivatives (i.e. columns)
  {
    aux = 0;
    for(int alpha = 0; alpha <= min(c, N); alpha++)
      aux = aux + Qtaylor(zz, alpha) * DTs[shift(s, c - alpha, E + 1)];
    for(int r = s * (E + 1); r < (s + 1) * (E + 1) - c; r++) // copy on equal rows
      difftable[shift(r, c, DTsize)] = aux;
  }
}
/// fill complementary triangles with finite differences
for(int s = 0; s < zsample.n - 1; s++)
  for(int c = 1; c <= E; c++) // columns
    for(int r = (s + 1) * (E + 1) - c; r < (s + 1) * (E + 1); r++) // rows
      difftable[shift(r, c, DTsize)] = (difftable[shift(r, c - 1, DTsize)]
                                       - difftable[shift(r + 1, c - 1, DTsize)])
                                       / (zsample[s] - zsample[s + 1]);
/// fill the rest with finite differences
for(int c = E + 1; c < DTsize; c++) // columns
  for(int r = 0; r < DTsize - c; r++) // rows
    difftable[shift(r, c, DTsize)] = (difftable[shift(r, c - 1, DTsize)]
                                     - difftable[shift(r + 1, c - 1, DTsize)])
                                     / (zsmpleff[r] - zsmpleff[c + r]);

/// build numerator
Vh<complex>[int] PM(M + 1);
for(int i = 0; i <= M; i++)
  PM[i] = difftable[i];

/// evaluate Pade' approximant
complex d;
complex newtonpoly = 1.;
Vh<complex> upade = 0;
zz = ztar;
for(int i = 0; i <= M; i++)
{
  upade = upade + PM[i] * newtonpoly;
  newtonpoly *= zz - zsmpleff[i];
}
d = Qtaylor(zz, 0);
upade = upade / d;

/// numerical solution in ztar
zz = ztar;
solve Helmoltz(uh, vh) = a - int2d(Th, qforder = 8)(f * conj(vh));

/// compute error norm
Vh<complex> err = uh - upade;
real errNorm = sqrt(int2d(Th, qforder = 8)(abs(dx(err))^2  + abs(dy(err))^2)
            + w^2 * int2d(Th, qforder = 8)(abs(err)^2));
cout << "Error norm: " << errNorm << endl;

plot(uh, wait = 1, fill = 1, value = 1, cmm = "Numerical sol for z = " + ztar);
plot(upade, wait = 1, fill = 1, value = 1, cmm = "Pade' approx for z = " + ztar);
plot(err, wait = 1, fill = 1, value = 1, cmm = "Error");
```

Code 6. Implementation of Algorithm 4. The parameters are the ones described in Section 6.6, with $M = 17$.

9 References

[1] F. Hecht, "FreeFem++. Third Edition, Version 3.50", Université Pierre et Marie Curie, Paris, `freefem.org/ff++/ftp/freefem++doc.pdf`.

[2] F. Bonizzoni, F. Nobile, I. Perugia, "Convergence analysis of Padé approximations for Helmholtz frequency response problems", MATHICSE Technical Report, Nr. 24, July 2016.

[3] S.A. Sauter, "*hp*-Finite Elements for Highly Indefinite Helmholtz Problems", Lecture Notes of the Zürich Summerschool, August 2016.

[4] G.A. Baker, P.R. Graves-Morris, "Padé approximants", Cambridge University Press, 1996.

[5] P. Guillaume, A. Huard, V. Robin, "Generalized Multivariate Padé Approximants", Journal of Approximation Theory, Nr. 95, pagg. 203-214, 1998.

[6] P.R. Graves-Morris, "Solution of integral equations using generalized inverse, function-valued Padé approximants, I", Journal of Computational and Applied Mathematics, Nr. 32, pagg. 117-124, 1990.

[7] P.R. Graves-Morris, R. Thukral, "Solution of integral equations using function-valued Padé approximants II", Numerical Algorithms 3, pagg. 223-234, 1992.

[8] M. Nica, "Eigenvalues and Eigenfunctions of the Laplacian", The Waterloo Mathematics Review, Vol. 1, 2011.

[9] G.H. Golub, C.F. Van Loan, "Matrix computations", Third edition, The Johns Hopkins University Press, Baltimore and London, 1996.

[10] R.D. Riess, "Error Estimates of Hermite Interpolation", BIT Numerical Mathemathics, Vol. 13, Issue 3, pagg. 338-343, 1973.

[11] I.M. Babuška, S.A.Sauter, "Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers?", SIAM Journal on Numerical Analysis 34, Nr. 6, pagg. 2392–2423, 1997.

[12] T. Kato, "Perturbation Theory for Linear Operators", The Waterloo Mathematics Review, Vol. 1, 2011.

[13] M. Tarek, "Domain Decomposition Methods for the Numerical Solution of Partial Differential Equations", Springer, 2008.

[14] P. Monk, "Finite Element Methods for Maxwell's Equations", Oxford Science Publications, 2003.

[15] M. Cessenat, "Mathematical Methods in Electromagnetism: Linear Theory and Applications", Series on Advances in Mathematics for Applied Sciences, Vol. 41, 1996.

[16] A.-S. Bonnet-Bendhia, L. Chesnel, S.A. Nazarov, "Non-scattering wavenumbers and far field invisibility for a finite set of incident/scattering directions", Cornell University Library, `arxiv.org/abs/1410.8382`, 2015.

[17] S. Salsa, "Partial Differential Equations in Action", First edition, Springer, 2007.

[18] M.H. Protter, "Unique continuation for elliptic equations", Trans. Amer. Math. Soc., Nr. 95, pagg. 81-91, 1960.

[19] A. Quarteroni, "Numerical Models for Differential Problems", Second edition, Springer, 2014.

[20] A. Quarteroni, R.Sacco, F.Saleri, "Numerical Mathematics", Second edition, Springer, 2007.

[21] E. Anderson, Z. Bai, C. Bishof et al., "LAPACK Users' Guide", Third edition, SIAM, Philadelphia, 1999.

[22] The Mathworks, Inc., Massachusset, USA, `mathworks.com/products/matlab`.