DOI: 10.61782/fa.2025.0967

FORUM ACUSTICUM EURONOISE 2025

GENERALIZED PLANE WAVE QUASI-TREFFTZ SPACES FOR WAVE PROPAGATION IN INHOMOGENEOUS MEDIA

Ilaria Fontana[1] **Lise-Marie Imbert-Gérard**[1*]
[1] Department of Mathematics, University of Arizona, USA

ABSTRACT

Partial Differential Equations (PDEs) models for wave propagation in inhomogeneous media are relevant for many applications. We will discuss numerical methods tailored for tackling problems governed by these variable-coefficient PDEs. Trefftz methods rely, in broad terms, on the idea of approximating solutions to PDEs via Galerkin methods using basis functions that are exact solutions of the PDE, making explicit use of information about the ambient medium. However, wave propagation in inhomogeneous media is modeled by PDEs with variable coefficients, and in general no exact solutions are available. Quasi-Trefftz methods have been introduced, in the case of the Helmholtz equation, to address this problem: they rely instead on high-order approximate solutions constructed locally. We will discuss basis of Generalized Plane Waves, a particular kind of quasi-Trefftz functions, and how their construction can be related to the construction of polynomial quasi-Trefftz bases.

Keywords: wave propagation in inhomogeneous media, variable-coefficient partial differential equation, Discontinuous Galerkin method, quasi-Trefftz method, discrete spaces of quasi-Trefftz functions.

1. INTRODUCTION

From radio-wave propagation in the ionosphere, to underwater acoustics, to heating and probing fusion plasmas, wave propagation in inhomogeneous media has many applications. Generalized Plane Waves (GPWs) were first introduced in [1] for a time-harmonic problem of reflectometry in plasma physics, modeled by the so-called O-mode equation, namely:

$$-\Delta u - \frac{\omega^2}{c^2}\left(1 - \frac{\omega_p^2}{\omega^2}\right) u = 0. \tag{1}$$

In this Helmholtz equation with a variable wavenumber, ω is the frequency; c is the speed of light in vacuum; and ω_p is the plasma frequency. The latter is proportional to the square of the plasma density, which depends on the position in space. For low density, $\omega_p^2 < \omega$, waves can propagate, but at high density, $\omega_p^2 > \omega$, the medium is evanescent. The density such that $\omega_p^2 = \omega^2$ is called the cut-off density, and for a given plasma density, changing the frequency ω changes the position of the cut-off in the domain.

Providing a high-order numerical method for simulations of waves bouncing on a cut-off was the original motivation to develop GPW-based quasi-Trefftz methods, taking advantage of a particular type of efficient high-order methods –the so-called Trefftz methods– while handling variable-coefficient PDEs. A brief overview of these methods is presented in Section 1.1. Generally speaking, the specificity of these methods is to rely on problem-dependent discrete spaces to represent the unknown: these spaces are designed to retain good approximation properties (see Section 5.3) but with as few degrees of freedom as possible. Previous works [2, 3] have focused on GPW functions for specific problems. The present work is the first effort to present an abstract framework to study GPW spaces, in parallel to the abstract framework for polynomial quasi-Trefftz presented in [4]. Even though the GPW ansatz was introduced for the scalar wave propagation problem, one may seek a new ansatz of quasi-Trefftz functions for other types of problems, which the abstract framework developed here can handle.

***Corresponding author**: lmig@arizona.edu.

1.1 Trefftz and quasi-Trefftz methods

Discontinuous Galerkin (DG) methods are numerical techniques for solving PDE-driven problems, relying on a variational framework and on the definition of discrete spaces composed of discontinuous piecewise polynomials. Trefftz DG methods rely on discrete spaces using functions that are locally tailored to the specific equation under consideration. Specifically, the test and trial functions are exact solutions of the governing PDE on each mesh element. The properties of this high-order approach have been extensively studied particularly in the context of wave problems [5]. Notably, it has been shown that these techniques achieve a given level of accuracy with a reduced number of degrees of freedom compared to standard polynomial DG approaches. However, the main limitation of Trefftz methods is that when the PDE contains variable coefficients or a non-zero right-hand side, local exact solutions are generally unavailable.

To address this challenge, quasi-Trefftz methods rely on approximate (rather than exact) local solutions in the sense of a Taylor series truncation. More precisely, given a PDE defined, for a linear differential operator $\mathcal{L}$, by

$$\mathcal{L}u = 0, \tag{2}$$

the so-called *quasi-Trefftz property* for a function φ is

$$[T_q \circ \mathcal{L}][\varphi] = 0, \tag{3}$$

where T_q denotes the Taylor operator up to a suitable degree q and with center $\mathbf{x}_0$, and $T_q \circ \mathcal{L}$ is referred to as the *quasi-Trefftz operator*. Discrete spaces of functions with this property are referred to as *quasi-Trefftz spaces*.

Various families of quasi-Trefftz functions have been proposed [6]. A first choice is to construct a subspace of the polynomial space $\mathbb{P}^p$. Recent studies have demonstrated the convergence and stability of polynomial quasi-Trefftz methods for some scalar problems, for example, for the diffusion-advection-reaction equation [7] or for the Schrodinger equation [8]. Different types of GPW quasi-Trefftz functions have been studied in [1–3]. The original idea behind GPWs was to start from classical plane waves, that had been efficiently leveraged in Trefftz methods for problems modeling wave propagation in homogeneous media, and to allow for further flexibility via higher-order terms, either in the phase or in the amplitude. The guiding thread was to retain the oscillating behavior, since it provided high-order approximation properties with less degrees of freedom than classical polynomial spaces. The main goal of this paper is to investigate the general properties of quasi-Trefftz spaces of phase-based GPWs, i.e., functions e^P where P is a polynomial.

1.2 Quasi-Trefftz spaces

From now on, if γ is the order of the differential operator $\mathcal{L}$, we will fix $q = p - \gamma$ in the Taylor operator of the quasi-Trefftz property (3), since, as discussed in [6], this choice is optimal for both polynomials and GPWs.

For polynomial quasi-Trefftz spaces, the operator $T_{p-\gamma} \circ \mathcal{L}$ between $\mathbb{P}$ and $\mathbb{P}^{p-\gamma}$ is linear; the dimension of the kernel of $T_{p-\gamma} \circ \mathcal{L}$ is infinite; the dimension of the kernel of the operator

$$\mathcal{T}_{pol} := T_{p-\gamma} \circ \mathcal{L}|_{\mathbb{P}^p} \tag{4}$$

is finite. In this context, the quasi-Trefftz space is defined as

$$\ker \mathcal{T}_{pol} = \{\varphi \in \mathbb{P}^p, \mathcal{T}_{pol}[\varphi] = 0\}. \tag{5}$$

A block-triangular structure of the operator $\mathcal{T}_{pol}$ can be leveraged to construct a quasi-Trefftz basis, and this approach is presented in [4].

Generalized Plane Waves are defined starting from a plane wave but have higher-order terms, either in their amplitude or in their phase. Here, we focus on phase-based GPWs. In order to leverage GPWs, quasi-Trefftz methods rely on the existence of convenient GPW bases.

For phase-based GPW quasi-Trefftz spaces, the situation differs from the polynomial case. First, the operator between $\mathbb{P}$ and $\mathbb{P}^{p-\gamma}$,

$$T_{p-\gamma} \circ \mathcal{L} \circ \exp : P \to [T_q \circ \mathcal{L}][e^P] \tag{6}$$

is not linear as in general $\mathcal{L}e^P = (\mathcal{N}(P))e^P$, where $\mathcal{N}$ is some nonlinear operator. Here, the quasi-Trefftz operator is

$$\mathcal{T}_{gen} := T_{p-\gamma} \circ \mathcal{N}|_{\mathbb{P}^p}. \tag{7}$$

Second, the dimension of $\{\varphi = e^P, P \in \mathbb{P}, \mathcal{T}_{gen}[P] = 0\}$ is infinite, even if $\mathbb{P}$ is replaced by $\mathbb{P}^p$. As a consequence, defining the quasi-Trefftz space as

$$\mathbb{QT}_p := \{\varphi = e^P, P \in \mathbb{P}^p, \mathcal{T}_{gen}[P] = 0\}, \tag{8}$$

its dimension is infinite and $\mathcal{T}_{gen}$ is nonlinear. The goal is then to construct a set of linearly independent GPWs with desired approximation properties, despite the challenge brought by nonlinearity.

1.3 Scope

This article focuses on a general approach to achieve this goal, providing an abstract framework that can be applied to other choices of ansatz for quasi-Trefftz functions in the future.

The procedure to construct a GPW quasi-Trefftz function relies on a fundamental assumption. Given a fixed point $\mathbf{x}_0$, the differential operator of order $\gamma \in \mathbb{N}$

$$\mathcal{L}_c := \sum_{m=0}^{\gamma} \sum_{|\mathbf{j}|=m} c_{\mathbf{j}}\left(\mathbf{x}\right) \partial_{\mathbf{x}}^{\mathbf{j}}, \text{ and } \mathcal{L}_* := \sum_{|\mathbf{j}|=\gamma} c_{\mathbf{j}}\left(\mathbf{x}_0\right) \partial_{\mathbf{x}}^{\mathbf{j}} \tag{9}$$

this assumption states that $\mathcal{L}_*$ is surjective between spaces of homogeneous polynomials, namely between $\tilde{\mathbb{P}}^{n+\gamma}$ and $\tilde{\mathbb{P}}^{n}$ for any $n \in \mathbb{N}$.

Thanks to the decomposition of polynomial spaces into direct sums of homogeneous polynomial spaces, the quasi-Trefftz operator $\mathcal{T}_{gen}$ defined in (7) can conveniently be decomposed as the sum of:

- the terms that do not depend on the (polynomial) independent variable, $\mathcal{T}_{gen}^{ind}$,
- $\mathcal{T}_{gen} - \mathcal{T}_{gen}^{ind}$, in turn split into
  - $\mathcal{L}_*$ defined in (9),
  - the remainder that will be denoted $\mathcal{R}$.

1.4 Polynomial versus GPWs

The first description of an abstract framework to describe quasi-Trefftz spaces was first presented in [4] for the polynomial case. In the same spirit, this work introduces an abstract framework entailing the essential properties of the quasi-Trefftz operator for the GPW case. To provide perspective, the two cases are compared in the following way:

- regarding the quasi-Trefftz operator, it is linear in the polynomial case and nonlinear in the GPW case, due to the choice of ansatz;
- regarding the operator splitting, due to the composition of Taylor truncation and a differential operator,
  - both in the polynomial and in the GPW operators, the $\mathcal{L}_*$ terms play a pivotal role,
  - the remainder $\mathcal{R}$ is linear in the polynomial case and nonlinear in the GPW case,
  - there is, in the GPW case, a structure corresponding to the triangular structure in the polynomial case;
- regarding the construction of quasi-Trefftz functions, in both cases, the algorithm relies on an iteration over spaces of homogeneous polynomials of increasing degree, with the resolution of a linear system for the operator $\mathcal{L}_*$ at each step, corresponding only in the polynomial case to a forward substitution procedure;
- regarding the number of degrees of freedom in the construction of an individual quasi-Trefftz function, it is the same in both cases, namely $\dim \ker \mathcal{L}_*|_{\mathbb{P}^p}$;
- regarding the range of application, in both cases, the sufficient condition on the differential operator $\mathcal{L}$ is the surjectivity of $\mathcal{L}_*$ between spaces of homogeneous polynomials.

In the polynomial framework, quasi-Trefftz functions are elements of the kernel of a linear operator. In that case, the questions of existence and uniqueness of quasi-Trefftz functions can be answered directly by linear algebra arguments: there always exists quasi-Trefftz functions since the kernel at least contains the polynomial zero; there exists a unique quasi-Trefftz function if the kernel is trivial and the dimension of the space is the dimension of the kernel. By contrast, for a nonlinear operator the questions of existence and uniqueness are open and this work provides a general framework to answer them for GPWs: a constructive algorithm proves their existence and the quest for approximation properties leads to prove that the dimension of the space is actually infinite, but convenient finite bases with desired properties can be constructed.

2. NOTATION

In this paper, blackboard bold letters (e.g. $\mathbb{A}$, $\mathbb{B}$) indicate spaces, while callygraphic letters (e.g. $\mathcal{S}$, $\mathcal{R}$) represent operators between spaces. Moreover, we use the notation $[\![n_1, n_2]\!]$, $n_1 < n_2$, to denote the set of all integer numbers from n_1 to n_2. The set of positive integers is denoted $\mathbb{N}$ and the set of non-negative integers is denoted $\mathbb{N}_+$. For any $p \in \mathbb{N}_+$, $\mathbb{P}^p$ denotes the space of polynomials of degree at most equal to p, and $\tilde{\mathbb{P}}^p$ denotes the space of homogeneous polynomials of degree p.

3. ABSTRACT FRAMEWORK

Given a graded space $\mathbb{B} = \bigoplus_{n=0}^{s} \mathbb{B}_n$ with $s+1$ layers for some $s \in \mathbb{N}$, each element $y \in \mathbb{B}$ can be uniquely decomposed as the sum of layer elements:

$$y = \sum_{n=0}^{s} y_n \text{ where } \forall n \in [\![0, s]\!],\ y_n \in \mathbb{B}_n. \tag{10}$$

We now introduce the projection operators

$$\begin{aligned} \forall n \in [\![0, s]\!],\ \mathcal{P}_{\mathbb{B}_n} : \mathbb{B} &\to \mathbb{B}_n, \\ y &\mapsto y_n. \end{aligned} \tag{11}$$

Theorem 1. *Let $\mathbb{A} = \mathbb{F} \oplus (\bigoplus_{n=0}^{s} \mathbb{A}_n)$ and $\mathbb{B} = \bigoplus_{n=0}^{s} \mathbb{B}_n$ be two graded spaces, $s \in \mathbb{N}$, and let $\mathcal{T} \colon \mathbb{A} \to \mathbb{B}$ be a nonlinear operator. Suppose that $\mathcal{T}$ can be decomposed as the sum of two operators $\mathcal{T} = \mathcal{T}_* + \mathcal{R}$ such that*

1. *(a) $\mathcal{T}_* \colon \mathbb{A} \to \mathbb{B}$ is linear;*
   (b) for all $n \in [\![0, s]\!]$, $\mathcal{T}_*(\mathbb{A}_n) \subseteq \mathbb{B}_n$;
   (c) $\mathcal{T}_*$ is surjective;
   (d) $\mathcal{S}_* \colon \mathbb{B} \to \mathbb{A}$ is a (linear) right inverse of $\mathcal{T}_*$;
   (e) for all $n \in [\![0, s]\!]$, $\mathcal{S}_*(\mathbb{B}_n) \subseteq \mathbb{A}_n$;
2. *(a) $\mathcal{R} \colon \mathbb{A} \to \mathbb{B}$ is nonlinear;*
   (b) for all $n \in [\![0, s-1]\!]$
   $$\mathcal{R}\left(\bigoplus_{k=n}^{s} \mathbb{A}_k\right) \subseteq \bigoplus_{k=n+1}^{s} \mathbb{B}_k;$$
   (c) $\mathcal{R}(\mathbb{A}_s) = \{0_{\mathbb{B}}\}$;
   (d) for all $n \in [\![1, s]\!]$ and for all $x \in \mathbb{A}$ with zero $\mathbb{F}$-component,
   $$\mathcal{P}_{\mathbb{B}_n}\left[\mathcal{R}\left(\sum_{k=0}^{s} x_k\right)\right] = \mathcal{P}_{\mathbb{B}_n}\left[\mathcal{R}\left(\sum_{k=0}^{n-1} x_k\right)\right]$$
   where $x = 0_{\mathbb{F}} + \sum_{k=0}^{s} x_k$, with $x_k \in \mathbb{A}_k$ for all $k \in [\![0, s]\!]$, and $\mathcal{P}_{\mathbb{B}_n} : \mathbb{B} \to \mathbb{B}_n$ is the projection operator defined as in (11).

Then, $\mathcal{T}$ is surjective, and a right inverse is the operator $\mathcal{S} \colon \mathbb{B} \to \mathbb{A}$ defined by

$$\mathcal{S} = \mathcal{S}_* \circ \sum_{k=0}^{s} (-\mathcal{R} \circ \mathcal{S}_*)^k \tag{12}$$

Proof. Given $y \in \mathbb{B}$, and for any $x \in \mathbb{A}$ with zero $\mathbb{F}$-component, write

$$x = \sum_{k=0}^{s} x_k \quad \text{and} \quad y = \sum_{k=0}^{s} y_k \tag{13}$$

where $(x_k, y_k) \in \mathbb{A}_k \times \mathbb{B}_k$ for all $k \in [\![0, s]\!]$. Then

$$\begin{aligned} \mathcal{T}(x) &= \mathcal{T}_*(x) + \mathcal{R}(x) = \mathcal{T}_*\left(\sum_{k=0}^{s} x_k\right) + \mathcal{R}\left(\sum_{k=0}^{s} x_k\right) \\ &= \sum_{k=0}^{s} \mathcal{T}_*(x_k) + \sum_{n=0}^{s} \mathcal{P}_{\mathbb{B}_n}\left[\mathcal{R}\left(\sum_{k=0}^{s} x_k\right)\right] \\ &= \sum_{k=0}^{s} \mathcal{T}_*(x_k) + \sum_{n=1}^{s} \mathcal{P}_{\mathbb{B}_n}\left[\mathcal{R}\left(\sum_{k=0}^{n-1} x_k\right)\right] \end{aligned} \tag{14}$$

As a consequence, $\mathcal{T}(x) = y$ if and only if

$$\sum_{k=0}^{s} \mathcal{T}_*(x_k) + \sum_{k=1}^{s} \mathcal{P}_{\mathbb{B}_k}\left[\mathcal{R}\left(\sum_{l=0}^{k-1} x_l\right)\right] = \sum_{k=0}^{s} y_k \tag{15}$$

if and only if

$$\begin{cases} \mathcal{T}_*(x_0) = y_0 \\ \mathcal{T}_*(x_k) = y_k - \mathcal{P}_{\mathbb{B}_k}\left[\mathcal{R}\left(\sum_{l=0}^{k-1} x_l\right)\right] \quad \forall k \in [\![1, s]\!] \end{cases} \tag{16}$$

For each $k \in [\![0, s]\!]$, if the right-hand side is known, a solution $x_k \in \mathbb{A}_k$ exists under hypotheses 1c and 1b. Hence, a suitable solution $x \in \mathbb{A}$ to the system can be constructed by solving for x_k for increasing values of k.

The components of a possible solution $x \in \mathbb{A}$ are:

$$\begin{cases} z = 0_{\mathbb{F}} \\ x_0 = \mathcal{S}_*(y_0) \\ x_k = \mathcal{S}_*\left(y_k - \mathcal{P}_{\mathbb{B}_k}\left[\mathcal{R}\left(\sum_{l=0}^{k-1} x_l\right)\right]\right) \quad \forall k \in [\![1, s]\!] \end{cases} \tag{17}$$

Summing all the components, we get

$$\begin{aligned} x &= \mathcal{S}_*(y_0) + \sum_{k=1}^{s} \mathcal{S}_*\left(y_k - \mathcal{P}_{\mathbb{B}_k}\left[\mathcal{R}\left(\sum_{l=0}^{k-1} x_l\right)\right]\right) \\ &= \mathcal{S}_*\left(y - \sum_{k=1}^{s} \mathcal{P}_{\mathbb{B}_k}\left[\mathcal{R}\left(\sum_{l=0}^{s} x_l\right)\right]\right) \\ &= \mathcal{S}_*\left(y - \sum_{k=0}^{s} \mathcal{P}_{\mathbb{B}_k}[R(x)]\right) \end{aligned} \tag{18}$$

So $x = \mathcal{S}_*(y - R(x))$. Iterating this l times

$$x = \mathcal{S}_* \left(\sum_{k=0}^{l} (-\mathcal{R} \circ \mathcal{S}_*)^k (y) \right) - \mathcal{S}_* (-\mathcal{R} \circ \mathcal{S}_*)^l (\mathcal{R}(x)) \tag{19}$$

Moreover, hypothesis 1e together with 2b and 2c imply that the operator $\mathcal{R} \circ \mathcal{S}_* : \mathbb{B} \to \mathbb{B}$ is nilponent, in particular $(\mathcal{R} \circ \mathcal{S}_*)^{s+1} = 0|_{\mathbb{B}\to\mathbb{B}}$. As a consequence,

$$\begin{aligned} x &= \mathcal{S}_* \left(\sum_{k=0}^{s+1} (-\mathcal{R} \circ \mathcal{S}_*)^k (y) \right) - \mathcal{S}_* (-\mathcal{R} \circ \mathcal{S}_*)^{s+1} (\mathcal{R}(x)) \\ &= \mathcal{S}_* \left(\sum_{k=0}^{s} (-\mathcal{R} \circ \mathcal{S}_*)^k (y) \right) \end{aligned} \tag{20}$$

which concludes the proof. □

The following theorem extends the result of the previous one by presenting an explicit algorithm to construct elements in the counterimage of $y \in \mathbb{B}$, inspired by the forward substitution technique of the polynomial case.

Theorem 2. *Let $\mathbb{A} = \mathbb{F} \oplus (\bigoplus_{n=0}^{s} \mathbb{A}_n)$ and $\mathbb{B} = \bigoplus_{n=0}^{s} \mathbb{B}_n$ be two graded spaces, $s \in \mathbb{N}$, and let $\mathcal{T} : \mathbb{A} \to \mathbb{B}$ be a nonlinear operator. Suppose that $\mathcal{T}$ can be decomposed as the sum of two operators $\mathcal{T} = \mathcal{T}_* + \mathcal{R}$ such that properties 1a–1c and 2a–2c of Theorem 1 hold. Additionally, assume that $\mathcal{T}_*$ and $\mathcal{R}$ satisfies the following conditions:*

1. *$\mathcal{T}_*(\mathbb{F}) = \{0_\mathbb{B}\}$*
2. *for all $n \in [\![0, s]\!]$, $\mathbb{A}_n = \mathbb{U}_n \oplus \mathbb{V}_n$ and $\mathcal{T}_*|_{\mathbb{V}_n}$ is invertible;*
3. *for all $n \in [\![0, s]\!]$, $\mathcal{S}_n : \mathbb{B}_n \to \mathbb{A}_n$ is the (linear) right inverse of $\mathcal{T}_*|_{\mathbb{A}_n}$ defined by $\mathcal{S}_n = [\mathcal{T}_*|_{\mathbb{V}_n}]^{-1}$;*
4. *for all $x \in \mathbb{A}$, $\mathcal{P}_{\mathbb{B}_0}[\mathcal{R}(x)] = \mathcal{P}_{\mathbb{B}_0}[\mathcal{R}(z)]$, and for all $n \in [\![1, s]\!]$,*

   $$\mathcal{P}_{\mathbb{B}_n}[\mathcal{R}(x)] = \mathcal{P}_{\mathbb{B}_n}\left[\mathcal{R}\left(z + \sum_{k=0}^{n-1} x_k\right)\right]$$

   where $x = z + \sum_{k=0}^{s} x_k$, with $z \in \mathbb{F}$ and $x_k \in \mathbb{A}_k$ for all $k \in [\![0, s]\!]$.

Then, given $y \in \mathbb{B}$, the following algorithm explicitly computes the components of a generic element $x \in \mathbb{A}$ such that $\mathcal{T}(x) = y$.

Algorithm 1: Construction of counterimages

Input: $z \in \mathbb{F}$ and $u = \sum_{k=0}^{s} u_k \in \bigoplus_{k=0}^{s} \mathbb{U}_k$
$x_0 \leftarrow u_0 + \mathcal{S}_0(y_0 - \mathcal{P}_{\mathbb{B}_0}[\mathcal{R}(z)] - \mathcal{T}_*(u_0))$;
for $k \in [\![1, s]\!]$ **do**
  $x_k \leftarrow u_k + \mathcal{S}_k \left(y_k - \mathcal{P}_{\mathbb{B}_k}\left[\mathcal{R}\left(z + \sum_{l=0}^{k-1} x_k\right)\right] - \mathcal{T}_*(u_k) \right)$;
end
Output: $x = z + \sum_{k=0}^{s} x_k$

Proof. We need to check that, if $x \in \mathbb{A}$ is the output of the algorithm, then $\mathcal{T}(x) = y$. Proceeding as in eq. (14), we have

$$\begin{aligned} \mathcal{T}(x) &= \mathcal{T}_*(x_0) + \sum_{k=1}^{s} \mathcal{T}_*(x_k) + \mathcal{P}_{\mathbb{B}_0}[\mathcal{R}(z)] \\ &\quad + \sum_{k=1}^{s} \mathcal{P}_{\mathbb{B}_k}\left[\mathcal{R}\left(z + \sum_{l=0}^{k-1} x_l\right)\right] \end{aligned} \tag{21}$$

We analyze the first two terms separately. In particular, we use the fact that $\mathcal{T}_* \circ \mathcal{S}_k = I_{\mathbb{B}_k}$ for all $k \in [\![0, s]\!]$.

$$\begin{aligned} \mathcal{T}_*(x_0) &= \mathcal{T}_*(u_0 + \mathcal{S}_0(y_0 - \mathcal{P}_{\mathbb{B}_0}[\mathcal{R}(z)] - \mathcal{T}_*(u_0)) \\ &= \mathcal{T}_*(u_0) + y_0 - \mathcal{P}_{\mathbb{B}_0}[\mathcal{R}(z)] - \mathcal{T}_*(u_0) \\ &= y_0 - \mathcal{P}_{\mathbb{B}_0}[\mathcal{R}(z)] \end{aligned} \tag{22}$$

and for all $k \in [\![1, s]\!]$

$$\begin{aligned} \mathcal{T}_*(x_k) &= \mathcal{T}_*(u_k) + y_k - \mathcal{P}_{\mathbb{B}_k}\left[\mathcal{R}\left(\sum_{l=0}^{k-1} x_l\right)\right] - \mathcal{T}_*(u_k) \\ &= y_k - \mathcal{P}_{\mathbb{B}_k}\left[\mathcal{R}\left(\sum_{l=0}^{k-1} x_l\right)\right] \end{aligned} \tag{23}$$

Finally, combining (21), (22), and (23) we conclude that $\mathcal{T}(x) = y$. □

4. CONSTRUCTION OF A GPW FUNCTION

This section aims to illustrate the previous results with practical examples to construct quasi-Trefftz functions. Choose $s = p-\gamma$, $\mathbb{F} = \mathbb{P}^{\gamma-1}$, $\mathbb{A}_n = \tilde{\mathbb{P}}^{n+\gamma}$, $\mathbb{B}_n = \tilde{\mathbb{P}}^n$ for all $n \in [\![0, p-\gamma]\!]$, so $\mathbb{A} = \mathbb{P}^p$ and $\mathbb{B} = \mathbb{P}^{p-\gamma}$, together with $\mathcal{T}_{gen} := T_{p-\gamma} \circ \mathcal{N}$, and $\mathcal{T}_* := \mathcal{L}_*$ defined in (9).

4.1 Surjectivity

A general discussion on operators $\mathcal{L}_*$ satisfying the theorem's hypotheses is proposed in [4]. As a particular case, consider operators $\mathcal{L}_*$ of order $\gamma = 2$ for which $\exists \mathbf{j} \in (\mathbb{N}_+)^d$ with $|\mathbf{j}| = \gamma$ and some component $j_k = \gamma$ such that $c_{\mathbf{j}}(\mathbf{x}_0) \neq 0$. They are surjective between spaces of polynomials. Moreover, regarding the inverse operator $\mathcal{S}_n$ for any $n \in \mathbb{N}_0$: consider

$$\mathbb{V}_n = \left\{ P \in \tilde{\mathbb{P}}^{n+\gamma}, P = \sum_{|\mathbf{j}|=n+\gamma, j_k \notin \{0,1\}} \alpha_{\mathbf{j}} \mathbf{X}^{\mathbf{j}} \right\}, \quad (24)$$

then $\mathcal{L}_*|_{\mathbb{V}_n}$ is invertible. The inverse of this invertible operator satisfies the hypothesis on $\mathcal{S}_n$ from the previous Theorem 2. This case includes both the Helmholtz and the convected Helmholtz operators.

4.2 Helmholtz Equation

Consider the Helmholtz equation $\mathcal{L}\varphi = 0$, where

$$\mathcal{L}\varphi := \Delta\varphi + \kappa^2\varphi, \quad (25)$$

and κ is a variable coefficient. Here, $\mathcal{L}e^P = \mathcal{N}(P)e^P$ with, for any $P \in \mathbb{P}^p$,

$$\mathcal{N}(P) := \Delta P + |\nabla P|^2 + \kappa^2, \quad (26)$$

$$T_{p-2} \circ \mathcal{N}(P) = \Delta P + T_{p-2}|\nabla P|^2 + T_{p-2}\,\kappa^2, \quad (27)$$

$$\mathcal{T}_{gen}(P) = 0 \;\Leftrightarrow\; \Delta P + T_{p-2}|\nabla P|^2 = -T_{p-2}\,\kappa^2, \quad (28)$$

and we accordingly define the operators

- $\mathcal{T}(P) := \Delta P + T_{p-2}|\nabla P|^2$,
- $\mathcal{T}_*(P) := \Delta P$, • $\mathcal{R}(P) := T_{p-2}|\nabla P|^2$.

We now check that $\mathcal{T}_*$ and $\mathcal{R}$ satisfy the desired properties. Notice that by their definition 1a and 2a of Theorem 1 automatically hold. Moreover,

- *[1b of Th.1 and 1. of Th.2]* For all $n \in [\![0, p-2]\!]$, $\mathcal{T}_*(\tilde{\mathbb{P}}^{n+2}) \subseteq \tilde{\mathbb{P}}^n$, and $\mathcal{T}_*(\mathbb{P}^1) = \{0_{\mathbb{B}}\}$.
- *[1c of Th.1 and 2.-3. of Th. 2 (which imply 1d-1e of Th. 1)]* See Subsection 4.1.
- *[2b and 2c of Th.1]* Given $n \in [\![0, p-2]\!]$ and

$$P \in \bigoplus_{k=n}^{p-2} \tilde{\mathbb{P}}^{k+2}, \text{ then } P = \sum_{k=n}^{p-2} P_{k+2} \quad (29)$$

where $P_{k+2} \in \tilde{\mathbb{P}}^{k+2}$ for all $k \in [\![n, p-2]\!]$, and

$$\begin{aligned} |\nabla P|^2 &= \left| \nabla \left(\sum_{k=n}^{p-2} P_{k+2} \right) \right|^2 \\ &= \sum_{m=2n+2}^{2p} \sum_{k=-1}^{m-1} \nabla P_{k+2} \cdot \nabla P_{m-k}. \end{aligned} \quad (30)$$

This expression uses the decomposition into the sum of homogeneous polynomials, indeed

$$\sum_{k=-1}^{m-1} \nabla P_{k+2} \cdot \nabla P_{m-k} \in \tilde{\mathbb{P}}^m \quad (31)$$

for all $m \in [\![0, p-2]\!]$. Moreover, the action of $\mathcal{R}$ can be written as

$$\mathcal{R}(P) = \sum_{m=2n+2}^{p-2} \sum_{k=-1}^{m-1} \nabla P_{k+2} \cdot \nabla P_{m-k} \quad (32)$$

if $n \in [\![0, \lfloor p/2 \rfloor - 2]\!]$, while

$$\mathcal{R}(P) = 0 \quad (33)$$

if $n \in [\![\lfloor p/2 \rfloor - 1, p-2]\!]$ since the lowest degree in (30) is $2n+2$. As a consequence, for all $n \in [\![0, \lfloor p/2 \rfloor - 2]\!]$,

$$\mathcal{R}\left(\bigoplus_{k=n}^{p-2} \tilde{\mathbb{P}}^{k+2} \right) \subseteq \bigoplus_{k=2n+2}^{p-2} \tilde{\mathbb{P}}^k \subset \bigoplus_{k=n+1}^{p-2} \tilde{\mathbb{P}}^k, \quad (34)$$

and for all $n \in [\![\lfloor p/2 \rfloor - 1, p-2]\!]$,

$$\mathcal{R}\left(\bigoplus_{k=n}^{p-2} \tilde{\mathbb{P}}^{k+2} \right) = \{0_{\mathbb{B}}\}. \quad (35)$$

- *[4. of Th.2 (which implies 2d of Th.1)]* We have

$$\mathcal{R}\left(\sum_{k=-2}^{p-2} P_{k+2} \right) = \sum_{m=0}^{p-2} \sum_{k=-1}^{m-1} \nabla P_{k+2} \cdot \nabla P_{m-k} \quad (36)$$

because $\nabla P_0 \cdot \nabla P_{k+2} = 0$ for all $k \in [\![-2, p-2]\!]$. As a consequence, for all $n \in [\![0, p-2]\!]$,

$$\begin{aligned}\mathcal{P}_{\tilde{\mathbb{P}}^n}&\left[\mathcal{R}\left(\sum_{k=-2}^{p-2} P_{k+2}\right)\right]\\ &= \sum_{k=-1}^{n-1} \nabla P_{k+2} \cdot \nabla P_{n-k} \qquad (37)\\ &= \mathcal{P}_{\tilde{\mathbb{P}}^n}\left[\mathcal{R}\left(\sum_{k=-2}^{n-1} P_{k+2}\right)\right]\end{aligned}$$

4.3 Convected Helmholtz Equation

We now consider the convected Helmholtz equation for the acoustic potential:

$$\begin{aligned}\nabla \cdot (\rho(\nabla\varphi - (\boldsymbol{M}\cdot\nabla\varphi)\boldsymbol{M} + \mathrm{i}\kappa\varphi\boldsymbol{M})\\ + \rho(\kappa^2\varphi + \mathrm{i}\kappa\boldsymbol{M}\cdot\nabla\varphi) = 0 \quad (38)\end{aligned}$$

where ρ is the fluid density, $\boldsymbol{M}$ is the rescaled vector-valued fluid velocity, and κ is the (constant) wave number. Using phase-based GPWs, a polynomial $P \in \mathbb{P}^p$ satisfies the quasi-Trefftz property if

$$\begin{aligned}\mathcal{T}(P) := T_{p-2}[&\nabla\rho\cdot\nabla P + \rho\Delta P - \rho(\nabla\boldsymbol{M}\nabla P)\cdot\boldsymbol{M}\\ &- (\nabla(\rho\boldsymbol{M}) - 2\rho\mathrm{i}\kappa)(\boldsymbol{M}\cdot\nabla P)\\ &- \rho(\nabla(\nabla P)\boldsymbol{M})\cdot\boldsymbol{M} + \rho|\nabla P|^2\\ &- \rho(\boldsymbol{M}\cdot\nabla P)^2]\\ = &- T_{p-2}[\nabla\cdot(\rho\mathrm{i}\kappa\boldsymbol{M}) + \rho\kappa^2]\end{aligned} \qquad (39)$$

Also in this case, $\mathcal{T}\colon \mathbb{P}^p \to \mathbb{P}^{p-2}$ is a nonlinear operator that can decomposed into the sum of two operators $\mathcal{T} = \mathcal{T}_* + \mathcal{R}$ defined as follows: if

$$T_{p-2}\,\rho = \sum_{k=0}^{p-2}\rho_k \qquad \text{and} \qquad T_{p-2}\,\boldsymbol{M} = \sum_{k=0}^{p-2}\boldsymbol{M}_k \quad (40)$$

are the truncated Taylor series of ρ and $\boldsymbol{M}$, then

$$\mathcal{T}_*P := \rho_0\Delta P - \rho_0(\nabla(\nabla P)\boldsymbol{M}_0)\cdot\boldsymbol{M}_0, \qquad (41)$$

and $\mathcal{R}(P) := \mathcal{T}(P) - \mathcal{T}_*P$. For this second case, it can be shown using an approach similar to that of the previous section that $\mathcal{T}_*$ and $\mathcal{R}$ satisfy the desired properties. The main difference in this case is the need to include the Taylor expansion components of the variable coefficients ρ and $\boldsymbol{M}$. For the sake of simplicity, we omit the details.

5. GPW SPACES: BASIS AND APPROXIMATION PROPERTIES

Beyond the existence of GPW quasi-Trefftz functions, further aspects of GPW spaces are still problem-dependent. Various of these aspects have been studied in the literature, see [2, 3, 6]. In particular, linear independence and approximation properties are closely related in the following way: the former is a by-product of the proof of the latter.

This section revisits pre-existing results in the light of the abstract framework from Section 3, highlighting the role played by the surjectivity of the quasi-Trefftz operator in proving approximation properties of quasi-Trefftz spaces.

5.1 Candidate for a basis

The dimension of the GPW quasi-Trefftz space $\mathbb{QT}_p$ defined in (8) is infinite. Yet a space of finite dimension, and more particularly its basis, are required in practice for TDG methods. The intuition behind the choice of phase-based GPW ansatz is to seek functions under the form

$$\varphi(\mathbf{x}) = \exp\left(\mathrm{i}\kappa\mathbf{d}\cdot(\mathbf{x}-\mathbf{x}_0) + H.O.T.\right) \qquad (42)$$

where $H.O.T.$ are higher order polynomial terms while $\mathbf{d} \in \mathbb{C}^d$, a unit vector, corresponds to the direction of propagation for classical plane waves. Hence, in the literature for the Helmholtz equation, families of GPWs were introduced by

- initializing the construction of all functions in a similar fashion by setting
  - the linear coefficients as $\mathrm{i}\kappa\mathbf{d}$;
  - all the other free coefficients to zero;
- then for each function picking a distinct $\mathbf{d}$.

For a set of distinct directions $\{\mathbf{d}_\ell, \mathbf{d}_\ell \neq \mathbf{d}_{\tilde{\ell}} \forall \ell \neq \tilde{\ell}\}_{\ell\in I}$ for some index set I, the corresponding GPW family, denoted $\{\varphi_\ell\}_{\ell\in I}$, is related to a standard plane wave basis, namely $\{\psi_\ell : \mathbf{x} \to \exp \mathrm{i}\kappa\mathbf{d}_\ell\cdot(\mathbf{x}-\mathbf{x}_0)\}_{\ell\in I}$, used in Trefftz methods. Even though such candidate families are not polynomial, the polynomial space of their Taylor truncations of degree p are of particular interest for their study.

5.2 Plane wave case

For reference, given a set of distinct directions $\{\mathbf{d}_\ell, \mathbf{d}_\ell \neq \mathbf{d}_{\tilde{\ell}} \forall \ell \neq \tilde{\ell}\}_{\ell\in I}$ for some index set I, consider the polynomial space of Taylor truncations of degree p of the corresponding plane waves, namely $\{T_p\psi_\ell\}_{\ell\in I}$. It is known that its dimension has an upper bound, $\dim\{T_p\psi_\ell\}_{\ell\in I} \leq D_d$, depending on the dimension d: $D_{d=2} = 2p+1$ and $D_{d=3} = (p+1)^2$ [9,10] . In fact, this is related to trigonometric functions for $d = 2$, [3], and to spherical harmonics for $d = 3$, [6]. It is also known that spaces of plane waves with dimension at least equal to D_d, that is $\dim I \geq D_d$, have approximation properties of order p.

5.3 Best approximation properties

Consider a set of directions $\{\mathbf{d}_\ell, \mathbf{d}_\ell \neq \mathbf{d}_{\tilde{\ell}} \forall \ell \neq \tilde{\ell}\}_{1\leq \ell \leq D_d}$. For the corresponding candidate family from Section 5.1, best approximation properties have been studied via the polynomial space of its functions' Taylor truncations of degree p, namely $\{T_p\varphi_\ell\}_{1\leq\ell\leq D_d}$. The surjectivity of the quasi-Trefftz operator is fundamental to prove that:

- $T_p u \in \{T_p\varphi_\ell\}_{1\leq\ell\leq D_d}$,
- $\dim\{T_p\varphi_\ell\}_{1\leq\ell\leq D_d} = \dim\{T_p\psi_\ell\}_{1\leq\ell\leq D_d}$.

Theorem 3. *Assume $\gamma \in \mathbb{N}$ and $p \in \mathbb{N}$ with $p \geq \gamma$. If the coefficients of $\mathcal{L}$ are in $\mathcal{C}^{p-\gamma}$, consider a set of directions $\{\mathbf{d}_\ell, \mathbf{d}_\ell \neq \mathbf{d}_{\tilde{\ell}} \forall \ell \neq \tilde{\ell}\}_{1\leq \ell \leq D_d}$ and the corresponding GPW quasi-Trefftz space $\mathbb{QT}_p$. Then, for any function $u \in \mathcal{C}^{p+\gamma}$ such that $\mathcal{L}u = 0$ in a neighborhood of $\mathbf{x}_0$, there exists a quasi-Trefftz function $u_a \in \mathbb{QT}_p$ and $C \in \mathbb{R}$ independent of $\mathbf{x}$, but depending on u, $\mathbf{x}_0$ and p such that, in a neighborhood of $\mathbf{x}_0$,*

$$|u(\mathbf{x}) - u_a(\mathbf{x})| \leq C|\mathbf{x} - \mathbf{x}_0|^{p+1}. \tag{43}$$

6. FUTURE WORK

This work shines a new light on quasi-Trefftz spaces based on a GPW ansatz. Thanks to an abstract framework, it underlines the distinction between GPW-specific properties and more general quasi-Trefftz properties. The latter can be applied to other choices of ansatz for quasi-Trefftz functions in the future, among them the surjectivity of the quasi-Trefftz operator is absolutely fundamental. The former are mainly related to the ansatz, and therefore to plane waves. Hence for other ansatz we anticipate that they would relate to a relevant reference case.

7. ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Award Number DE-SC0024246.

8. REFERENCES

[1] L.-M. Imbert-Gérard and B. Després, "A generalized plane-wave numerical method for smooth nonconstant coefficients," *IMA Journal of Numerical Analysis*, vol. 34, pp. 1072–1103, 10 2013.

[2] L.-M. Imbert-Gérard, "Interpolation properties of generalized plane waves," *Numerische Mathematik*, vol. 131, no. 4, pp. 683–711, 2015.

[3] L.-M. Imbert-Gérard and G. Sylvand, "A roadmap for generalized plane waves and their interpolation properties," *Numerische Mathematik*, vol. 149, no. 1, pp. 87–137, 2021.

[4] L.-M. Imbert-Gérard, "Taylor-based polynomial quasi-trefftz spaces for scalar linear equations," *submitted*.

[5] R. Hiptmair, A. Moiola, and I. Perugia, *A Survey of Trefftz Methods for the Helmholtz Equation*, pp. 237–279. Springer International Publishing, 2016.

[6] L.-M. Imbert-Gérard and G. Sylvand, "Three types of quasi-Trefftz functions for the 3D convected Helmholtz equation: construction and approximation properties," *IMA Journal of Numerical Analysis*, p. drae060, 2024.

[7] L.-M. Imbert-Gérard, A. Moiola, C. Perinati, and P. Stocker, "Polynomial quasi-trefftz DG for PDEs with smooth coefficients: elliptic problems," *IMA Journal of Numerical Analysis*, p. drae094, 01 2025.

[8] S. Gómez and A. Moiola, "A space–time DG method for the Schrödinger equation with variable potential," *Advances in Computational Mathematics*, vol. 50, no. 2, 2024.

[9] O. Cessenat, "Application d'une nouvelle formulation variationnelle aux équations d'ondes harmoniques. problèmes de helmholtz 2d et de maxwell 3d.," *PhD thesis, Université Paris IX Dauphine*, 1996.

[10] A. Moiola, "Trefftz-discontinuous galerkin methods for time-harmonic wave problems.," *Diss. ETH Zurich*, 2011.