

Direction-Adaptive Plane-Wave Discontinuous Galerkin Methods for the Helmholtz Equation

Shelvean Kapita^{1*}

^{1*}Department of Mathematics, Texas A&M University, College Station, Texas, USA.

Abstract

We consider plane-wave discontinuous Galerkin (PWDG) approximations of the Helmholtz equation in which the local propagation directions are allowed to change. The directions are determined by minimizing a weighted residual on the mesh skeleton. We consider two formulations. In Part A the PWDG equations are retained and, for each set of directions, the plane-wave coefficients are obtained from the Galerkin system. In Part B the same skeleton residual is minimized over both coefficients and directions, with the coefficient variables eliminated by variable projection. Complex angles are used so that propagating and evanescent waves are included in the same local Trefftz family. We also use the Trefftz-DG norm to normalize the discrete system and a local Cauchy-trace Gramian to remove numerically dependent directions. On straight edges the plane-wave products in the residual can be integrated exactly. For an all-Dirichlet problem the residual is the squared DG error, and its local contributions can also be used for adaptivity. We prove a local quadratic-growth result for the reduced direction problem in a fixed-rank neighborhood of an identifiable zero-residual solution. The numerical experiments first show a limitation: on an exact circular DtN boundary the dominant phase direction of a Hankel wave is recovered accurately, but a small plane-wave fan does not give a comparably small field error. We then show that the high- p DtN floor is affected separately by the coefficient basis, the trace cutoff and the arithmetic used to compute the trace spectrum. Finally, for a finite sum of plane waves, residual-based ENRICH-MOVE continuation recovers all directions to roundoff for $M = 1, \dots, 19$ in the family considered. At $M = 20$ the automatic birth step enters a false basin, while a nearby birth at the same dimension again gives roundoff error. These experiments show that direction adaptation is most effective when the solution has low directional complexity.

Keywords: Helmholtz equation, plane-wave discontinuous Galerkin method, Trefftz method, direction adaptivity, variable projection

MSC Classification: 65N30 , 65N50 , 65N15

1 Introduction

The numerical solution of the Helmholtz equation becomes increasingly expensive as the wavenumber is increased. One approach is to use oscillatory basis functions that already contain part of the behavior of the solution. If the local basis functions satisfy the Helmholtz equation exactly, the resulting method is a Trefftz method. Plane waves give a particularly simple Trefftz space and lead to numerical formulations involving only traces on the mesh skeleton. The ultra-weak variational formulation of Cessenat and Després [1] is an early example. Plane-wave discontinuous Galerkin methods were subsequently analyzed in [4–6]; see also the survey [7]. Residual based mesh adaptivity for PWDG was considered in [9].

In the usual PWDG method, the number of plane waves and their directions are chosen when the discrete space is constructed. Uniformly spaced directions are convenient and work well in many examples. They can, however, be wasteful when the solution is dominated by only a few propagation directions. Adding more uniformly spaced waves is also not without cost. As the local dimension is increased, the traces of the plane waves become nearly dependent and the resulting algebraic problem can be severely ill-conditioned [11]. This suggests treating the directions as part of the approximation problem rather than fixing them beforehand.

This idea has been considered in several forms. Amara et al. [14] rotate local plane-wave bases in a least-squares method and determine the rotations from a nonlinear minimization problem. Agrawal and Hoppe [13] optimize real PWDG directions while keeping the PWDG variational equation as a constraint. Congreve, Houston and Perugia [10] use directional information together with local enrichment and mesh refinement in an hp -Trefftz DG method. Fang et al. [15] first compute a lower-frequency solution, extract dominant ray directions, and use those directions in a higher-frequency finite element space. Other methods using approximate propagation information are described in [16, 17].

Our aim is to study a PWDG calculation in which the directions are determined from the same residual that is later used to judge the quality of the approximation. The residual contains the jumps of the field and normal flux and the mismatch in the boundary conditions. We allow the angle of a plane wave to be complex. Hence an ordinary propagating wave and an evanescent wave are obtained from the same formula. No separate evanescent basis or critical-angle test is required.

We use two related formulations. In the first, which we call Part A, the PWDG equations are kept. For a given set of directions we solve the PWDG problem, substitute the computed field into the skeleton residual, and then vary only the directions. Thus the coefficients satisfy the Galerkin equations at every nonlinear iterate. This is closest to the formulation of Agrawal and Hoppe [13]. In Part B we remove the Galerkin constraint and minimize the skeleton residual over the broken Trefftz field itself. With the directions fixed, this is a linear least-squares problem of the type studied by Monk and Wang [2]. We then eliminate the linear coefficient variables by variable projection [3].

There are two algebraic issues that arise immediately in such a calculation. The first is conditioning. The Euclidean condition number of the PWDG matrix depends strongly on the scaling of the plane-wave coefficients and does not by itself describe the discrete operator in its natural norm. We therefore use the standard coercive Trefftz-DG norm as a Riesz metric. This gives a graph-Riesz condition number for the normalized operator. The second issue is numerical rank. If several local directions are close, some combinations of their Cauchy traces contain essentially no new information. We detect these combinations with a local trace Gram matrix and remove them before the global solve. Related conditioning and redundancy issues for Trefftz spaces are discussed in [12, 18–20].

For straight-sided elements, the edge integrals needed by the residual can be evaluated directly. Products of two plane waves reduce to exponential moments, and the same is true after differentiation with respect to a direction parameter. Therefore the least-squares matrix can be assembled without sampling the homogeneous edge residuals. There is a further simplification in the all-Dirichlet case. If u is the exact solution and v_h is any broken Trefftz field, then

$$\mathcal{J}(v_h) = \mathcal{G}_h(u - v_h, u - v_h) = \|u - v_h\|_{\mathcal{G},h}^2.$$

Thus the quantity minimized by the direction search is also the exact squared DG/graph error. Its edge contributions give local error quantities without introducing a second estimator. For mixed Dirichlet–impedance conditions this exact identity is lost. In that case we obtain an L^2 bound from a Helmholtz stability estimate stated later in the paper.

The numerical experiments are arranged to show first where direction movement helps and then where it does not. We first use three hidden plane waves to test exact recovery, and a transmission problem to check that the same complex-angle parameterization recovers propagating and evanescent waves. We next return to the exact circular DtN geometry of [8]. For a Hankel field the learned directions follow the local phase very well, but the field error remains limited by curvature and amplitude variation. We then increase p on the same geometry and investigate separately the effects of the raw coefficient basis, the trace cutoff and arithmetic precision. Our main recovery experiment is considered after these tests. The exact field is a finite sum of M plane waves, and M is increased one direction at a time. For the family used here the residual-based ENRICH–MOVE iteration gives roundoff error for $M = 1, \dots, 19$. At $M = 20$ the automatic birth step enters a false basin, while a nearby birth at the same dimension again converges to roundoff. Thus the important distinction in the experiments is between a field with a small number of identifiable directions and a field with general angular content. For the more nonconvex examples we use continuation, initializing a more difficult problem with directions obtained from the preceding one [30].

The paper is organized as follows. In Section 2 we give the PWDG formulation and introduce the complex-angle Trefftz space. Section 3 describes the graph-Riesz normalization and local trace-rank reduction. In Section 4 we define the skeleton residual and the two nonlinear formulations, and we derive the local fixed-rank result and the direction derivatives. The numerical algorithms are summarized in Section 5. Section 6 contains the numerical experiments. We end with conclusions in Section 7.

2 PWDG with Variable Complex Directions

2.1 Model problem

We begin with a standard truncated acoustic scattering problem. Let $\Omega \subset \mathbb{R}^2$ be a bounded Lipschitz domain with boundary

$$\partial\Omega = \Gamma_D \cup \Gamma_R, \quad \Gamma_D \cap \Gamma_R = \emptyset.$$

Here Γ_D denotes the boundary on which Dirichlet data are prescribed and Γ_R is an artificial exterior boundary, as in the usual truncated scattering problem [1, 7]. If n is the outward unit normal to Ω , we consider

$$\begin{aligned} \Delta u + \kappa^2 u &= 0 && \text{in } \Omega, \\ u &= g && \text{on } \Gamma_D, \\ \partial_n u + i\kappa u &= 0 && \text{on } \Gamma_R, \end{aligned} \tag{1}$$

where $\kappa > 0$ is the wavenumber. We use the time dependence $e^{+i\omega t}$. With this convention an outgoing radial wave behaves like $e^{-i\kappa r}$, so the first-order absorbing condition is $\partial_n u + i\kappa u = 0$. Under the alternative convention $e^{-i\omega t}$ the outgoing factor is $e^{+i\kappa r}$ and the corresponding impedance sign is reversed. The PWDG fluxes below use the $e^{+i\omega t}$ convention consistently.

Two special cases will be used in the numerical experiments. In the first, $\Gamma_R = \emptyset$ and Dirichlet data are prescribed on the whole boundary of a square. In the second, the wavenumber is piecewise constant and the field and normal flux are continuous across a material interface [21]. The PWDG formulation below is written first for (1); the changes needed for these two examples are stated when they are used.

For an interior edge $e = \partial K^+ \cap \partial K^-$ with outward normals $n^\pm$, we use the standard two-sided DG averages and jumps [5, 7, 9]. For a scalar trace v ,

$$\{\{v\}\} = \frac{1}{2}(v^+ + v^-), \quad \llbracket v \rrbracket = v^+ n^+ + v^- n^-, \tag{2}$$

so that $\llbracket v \rrbracket$ is vector-valued. For a vector trace $\boldsymbol{\tau}$,

$$\{\{\boldsymbol{\tau}\}\} = \frac{1}{2}(\boldsymbol{\tau}^+ + \boldsymbol{\tau}^-), \quad \llbracket \boldsymbol{\tau} \rrbracket_n = \boldsymbol{\tau}^+ \cdot n^+ + \boldsymbol{\tau}^- \cdot n^-, \tag{3}$$

so that $\llbracket \boldsymbol{\tau} \rrbracket_n$ is scalar-valued. We reserve the double-line symbols $\llbracket \cdot \rrbracket$ and $\llbracket \cdot \rrbracket_n$ for these DG jumps throughout the paper. Ordinary single brackets $[\cdot]$ are used later for an oriented scalar difference on a fixed edge, where $[v] = v^+ - v^-$ after choosing one normal n_e . They are not DG jumps. The Monk–Wang comparison uses this oriented scalar difference for the full-gradient jump. The normals are real, hence conjugation commutes with the trace operators.

Edge-scale convention.

For the constant-wavenumber problem we take $\xi_e = \kappa$. The solution jump is then weighted by $\alpha\kappa$ and the normal-flux jump by β/κ , as in the usual Trefftz-DG scaling [5, 7]. In the transmission experiment the two neighboring elements have different wavenumbers, and we use the arithmetic interface scale

$$\xi_e := \frac{\kappa_{K^+} + \kappa_{K^-}}{2} \quad (e = \partial K^+ \cap \partial K^-), \tag{4}$$

and retain the reciprocal pair $\alpha\xi_e$ and β/ξ_e . This choice of average is made only for the numerical transmission problem. We do not use it as a general variable-coefficient flux prescription. The remaining flux parameters satisfy $\alpha, \beta > 0$ and $0 < \delta \leq 1/2$; admissible choices and their role in the Trefftz-DG norm are discussed in [4, 5]. We take $\alpha = \beta = 1/2$ in the all-Dirichlet computations. The parameter δ is needed only on Γ_R and determines the split of the Robin boundary terms.

2.2 Complex-angle Trefftz basis functions

We next describe the local Trefftz space. Let x_K be a reference point in the element K ; in the computations we take x_K to be the centroid. For a complex angle $z = \theta + i\eta \in \mathbb{C}$, define

$$q_K(z) = \kappa_K(\cos z, \sin z) \tag{5}$$

and

$$\phi_{K,z}(x) = \exp(iq_K(z) \cdot (x - x_K)). \quad (6)$$

The dot product in (6) is bilinear, without complex conjugation.

Proposition 2.1 (complete complex-angle Trefftz family) *For every $z \in \mathbb{C}$,*

$$\Delta\phi_{K,z} + \kappa_K^2 \phi_{K,z} = 0 \quad \text{in } K.$$

In particular, $\eta = 0$ gives an ordinary propagating plane wave, while $\eta \neq 0$ gives an inhomogeneous or evanescent Trefftz wave [18]. Conversely, every $q \in \mathbb{C}^2$ satisfying the complex Helmholtz dispersion relation $q \cdot q = \kappa_K^2$ can be written as $q = q_K(z)$ for some $z \in \mathbb{C}$.

Proof The complex trigonometric identity $\cos^2 z + \sin^2 z = 1$ gives $q_K(z) \cdot q_K(z) = \kappa_K^2$. Since $\nabla\phi_{K,z} = iq_K(z)\phi_{K,z}$,

$$\Delta\phi_{K,z} = -(q_K(z) \cdot q_K(z))\phi_{K,z} = -\kappa_K^2 \phi_{K,z}.$$

For the converse, let $q = (q_1, q_2) \in \mathbb{C}^2$ satisfy $q_1^2 + q_2^2 = \kappa_K^2$ and set $w = (q_1 + iq_2)/\kappa_K$. Then $w \neq 0$ and $w^{-1} = (q_1 - iq_2)/\kappa_K$. Choose any logarithm branch at the nonzero number w and take $z = -i \log w$. Since $e^{iz} = w$,

$$\cos z = \frac{w + w^{-1}}{2} = \frac{q_1}{\kappa_K}, \quad \sin z = \frac{w - w^{-1}}{2i} = \frac{q_2}{\kappa_K}.$$

Thus $q = q_K(z)$. □

Given local direction parameters

$$Z_K = (z_{K,1}, \dots, z_{K,r_K}),$$

define

$$V_K(Z_K) = \text{span}\{\phi_{K,z_{K,1}}, \dots, \phi_{K,z_{K,r_K}}\}, \quad V_h(Z) = \prod_{K \in \mathcal{T}_h} V_K(Z_K). \quad (7)$$

In a standard p -PWDG calculation the local directions are prescribed when the approximation space is defined [5, 6]. In the present method the collection Z will be varied by the nonlinear algorithm.

For numerical purposes the complex angles are restricted to a bounded set. Write $z = \theta + i\eta$ and fix $\eta_{\max} > 0$. We define

$$\mathcal{Z}_K(\eta_{\max}) := (\mathbb{R}/(2\pi\mathbb{Z})) \times [-\eta_{\max}, \eta_{\max}], \quad (8)$$

and, for the current local multiplicities $\mathbf{r} = (r_K)_{K \in \mathcal{T}_h}$,

$$\mathcal{Z}_h(\mathbf{r}, \eta_{\max}) := \prod_{K \in \mathcal{T}_h} \mathcal{Z}_K(\eta_{\max})^{r_K}. \quad (9)$$

All reported complex-angle experiments use $\eta_{\max} = 1.8$. This is the actual bound imposed in the reference code.

The imaginary component has the useful physical interpretation

$$\text{Im } q_K(\theta + i\eta) = \kappa_K \sinh \eta (-\sin \theta, \cos \theta), \quad \|\text{Im } q_K(\theta + i\eta)\|_2 = \kappa_K |\sinh \eta|. \quad (10)$$

It follows that $h_K \kappa_K |\sinh \eta|$ is a useful measure of the exponential variation of the basis function over K . One could impose a bound on this quantity instead of a fixed bound on $|\eta|$, but we do not do so in the computations below. We also note that a permutation of the r_K angles does not change $V_K(Z_K)$. The nonlinear objective therefore has the corresponding permutation symmetry, although the implementation simply keeps the directions as labeled variables.

2.3 Skeleton formulation

We now write the PWDG formulation for fixed directions Z . We use the convention that $a_h(\cdot, \cdot)$ is linear in the first argument and conjugate-linear in the second. The mixed Dirichlet–impedance form is

$$\begin{aligned} a_h(u, v) = & \sum_{e \in \mathcal{E}_h^i} \int_e \left(\{u\} \llbracket \nabla_h \bar{v} \rrbracket_n - \llbracket \bar{v} \rrbracket \cdot \{\nabla_h u\} + i \frac{\beta}{\xi_e} \llbracket \nabla_h u \rrbracket_n \llbracket \nabla_h \bar{v} \rrbracket_n \right) ds \\ & + \sum_{e \in \mathcal{E}_h^i} \int_e i \alpha \xi_e \llbracket u \rrbracket \cdot \llbracket \bar{v} \rrbracket ds \end{aligned} \quad (11)$$

$$\begin{aligned}
& + \int_{\Gamma_D} \left(-\partial_n u \bar{v} + i\alpha\kappa u \bar{v} \right) ds \\
& + \int_{\Gamma_R} \left((1-\delta)i\kappa u \bar{v} + (1-\delta)u \partial_n \bar{v} - \delta \partial_n u \bar{v} - \delta(i\kappa)^{-1} \partial_n u \partial_n \bar{v} \right) ds.
\end{aligned}$$

The corresponding right-hand side for the homogeneous impedance condition in (1) is

$$F_h(v) = \int_{\Gamma_D} g \left(i\alpha\kappa \bar{v} - \partial_n \bar{v} \right) ds. \quad (12)$$

Equations (11)–(12) are the standard primal Trefftz-DG/PWDG flux formulation for a Dirichlet–Robin problem [5, 7]. When $\Gamma_R = \emptyset$ the Robin terms disappear, and this is the form used for the all-Dirichlet numerical examples.

For a complex wave vector q , if $v(x) = e^{iq \cdot (x - x_\kappa)}$, then the conjugated test traces are

$$\nabla_h \bar{v} = -i \bar{q} \bar{v}, \quad \partial_n \bar{v} = -i(\bar{q} \cdot n) \bar{v}.$$

The fixed-direction problem is

$$\text{find } u_h(Z) \in V_h(Z) : \quad a_h(u_h(Z), v_h) = F_h(v_h) \quad \forall v_h \in V_h(Z). \quad (13)$$

Since every function in $V_h(Z)$ satisfies the Helmholtz equation on each element, the volume terms cancel after elementwise integration by parts and the resulting discrete problem involves only the mesh skeleton [1, 7].

Fix an ordering of all active local Trefftz basis functions and write the resulting global broken Trefftz basis as

$$\Phi(Z) = \{\varphi_1(Z), \dots, \varphi_{N_h}(Z)\}.$$

With

$$u_h(Z) = \sum_{j=1}^{N_h} c_j(Z) \varphi_j(Z),$$

the assembled matrix and load vector are

$$A_{ij}(Z) = a_h(\varphi_j(Z), \varphi_i(Z)), \quad f_i(Z) = F_h(\varphi_i(Z)). \quad (14)$$

It follows that the fixed-direction PWDG problem can be written as

$$A(Z)c(Z) = f(Z). \quad (15)$$

For later comparison, we define the ordinary Euclidean condition number by

$$\kappa_2(A(Z)) = \|A(Z)\|_2 \|A(Z)^{-1}\|_2 \quad (16)$$

in the coefficient basis used to assemble A . This number changes with the choice and scaling of the basis. In the next section we introduce a second condition number after normalization by the DG/graph norm.

3 Graph Norm, Normalization and Numerical Rank

3.1 The DG/graph norm and graph–Riesz normalization

We first introduce the norm that will be used to normalize the discrete operator. With the same linear-first convention as for a_h , define

$$\begin{aligned}
\mathcal{G}_h(u, v) := & \sum_{e \in \mathcal{E}_h^I} \left[\alpha \xi_e \int_e \llbracket u \rrbracket \cdot \overline{\llbracket v \rrbracket} ds + \frac{\beta}{\xi_e} \int_e \llbracket \nabla_h u \rrbracket_n \cdot \overline{\llbracket \nabla_h v \rrbracket_n} ds \right] \\
& + \alpha\kappa \int_{\Gamma_D} u \bar{v} ds + \int_{\Gamma_R} \left[(1-\delta)\kappa u \bar{v} + \frac{\delta}{\kappa} \partial_n u \overline{\partial_n v} \right] ds.
\end{aligned} \quad (17)$$

For Trefftz functions, the standard PWDG Green identity gives [4, 5, 7]

$$\operatorname{Im} a_h(v, v) = \mathcal{G}_h(v, v). \quad (18)$$

If the retained Trefftz traces are independent and the fixed-direction PWDG problem is uniquely solvable, $\mathcal{G}_h(v, v) = 0$ implies $v = 0$. Hence $\|v\|_{\mathcal{G},h} = \mathcal{G}_h(v, v)^{1/2}$ is a norm on the retained space and its Gram matrix is Hermitian positive definite. When two or more local traces become nearly dependent, the smallest eigenvalues of this Gram matrix become very small. We therefore remove numerically dependent local trace combinations before applying the Cholesky normalization described below.

Remark 3.1 (Relation with the standard DG norm) For constant κ , this is exactly the usual Trefftz-DG/PWDG coercive norm [5, 7]:

$$\begin{aligned} \|v\|_{\text{DG}}^2 &= \frac{\beta}{\kappa} \|\llbracket \nabla_h v \rrbracket_n\|_{L^2(\mathcal{E}_h^1)}^2 + \alpha \kappa \|\llbracket v \rrbracket\|_{L^2(\mathcal{E}_h^1)}^2 + \alpha \kappa \|v\|_{L^2(\Gamma_D)}^2 \\ &\quad + \frac{\delta}{\kappa} \|\partial_n v\|_{L^2(\Gamma_R)}^2 + (1 - \delta) \kappa \|v\|_{L^2(\Gamma_R)}^2 = \|v\|_{\mathcal{G},h}^2. \end{aligned} \quad (19)$$

We use the name *graph norm* because the Gram matrix of this norm will be used as the coefficient-space Riesz map. This should not be confused with the stronger *DG+* norm that appears in continuity estimates. In the transmission tests, $\kappa^{\pm 1}$ on interior edges is replaced by $\xi_e^{\pm 1}$ from (4). Thus the heterogeneous implementation changes only the positive edge scale. It preserves the reciprocal solution-jump/flux-jump balance used in the coercive identity.

Weight conventions.

The PWDG operator and graph metric use $\alpha \xi_e$ and β/ξ_e on the interior solution and flux jumps. The residual $\mathcal{J}$ uses the same interior and Dirichlet weights but the strong Robin residual with factor $(2\kappa)^{-1}$. The local rank metric m_K instead uses only the dimensional pair κ_K, κ_K^{-1} . It is a rank diagnostic and does not enter the Galerkin form or the residual objective.

Let $A(Z)$ be the PWDG matrix in the active basis. The Gram matrix of (17) is

$$G_{ij}(Z) = \mathcal{G}_h(\varphi_j, \varphi_i) = \left[\frac{A - A^*}{2i} \right]_{ij}, \quad G = \frac{A - A^*}{2i}. \quad (20)$$

If $G > 0$, choose B so that

$$B^* G B = I. \quad (21)$$

In the computations $G = LL^*$ and B is obtained from the triangular solve $L^* B = I$. Writing $A = H + iG$ with $H = H^*$ gives the exact normal form

$$\hat{A} := B^* A B = S + iI, \quad S = B^* H B = S^*. \quad (22)$$

It follows that $\hat{A}$ is normal and $\sigma_{\min}(\hat{A}) \geq 1$. With $c = B\hat{c}$,

$$(B^* A B)\hat{c} = B^* f, \quad c = B\hat{c}, \quad (23)$$

so the Galerkin field is unchanged. The dense graph–Riesz realization is used here as a reference normalization. Large-scale variants can approximate the Riesz map locally or by Schwarz/multilevel techniques [31].

We measure intrinsic conditioning by

$$\kappa_{\text{GR}} = \kappa_2(B^* A B). \quad (24)$$

If λ_j are the eigenvalues of S , normality gives

$$\kappa_{\text{GR}} = \left(\frac{1 + \max_j \lambda_j^2}{1 + \min_j \lambda_j^2} \right)^{1/2} \leq \sqrt{1 + \gamma_h^2}, \quad \gamma_h = \|S\|_2. \quad (25)$$

The quantity κ_{GR} therefore measures the size of the Hermitian part of the discrete operator relative to the coercive skeleton metric. In particular, it is not simply a measure of how well the original plane-wave coefficients have been scaled.

Table 1 shows the difference between the two condition numbers for the uniform 32-triangle transmission mesh at $\omega = 12$. As p is increased, the Euclidean condition number grows very rapidly. The graph–Riesz

condition number grows much more slowly until the local traces themselves lose numerical rank. Once this happens, normalization alone is not sufficient and a rank reduction is needed. At $p = 29$ the smallest

Table 1 Ordinary and graph-Riesz conditioning for uniformly directed PWDG on the uniform 32-triangle mesh at $\omega = 12$.

p	coefficients	$\kappa_2(A)$	κ_{GR}
3	96	2.87×10^1	2.70
7	224	5.90×10^1	4.01
11	352	3.21×10^2	1.53×10^1
15	480	4.47×10^4	1.99×10^1
19	608	3.37×10^7	4.46×10^1
23	736	7.76×10^{10}	8.12×10^2
27	864	3.43×10^{14}	1.69×10^3
29	928	2.48×10^{17}	—

computed graph eigenvalue is -2.0×10^{-14} , a roundoff signature of numerical rank loss. We therefore report κ_{GR} as undefined before trace compression.

3.2 Local trace-rank compression

We now consider what happens when several plane waves are present on the same element. Near linear dependence of high-order plane-wave traces is well known [11, 12, 18]. Rather than testing rank with the assembled global graph matrix, we use a local Cauchy-trace Gram matrix. For $K \in \mathcal{T}_h$, define

$$m_K(u, v) = \sum_{e \subset \partial K} \left(\kappa_K \int_e u \bar{v} ds + \kappa_K^{-1} \int_e \partial_n u \overline{\partial_n v} ds \right). \quad (26)$$

The factors κ_K and κ_K^{-1} balance the two traces in the same dimensions as the PWDG graph weights. For the homogeneous calculations with $\alpha = \beta = 1/2$ the two scalings are proportional before the element traces are assembled into global edge terms. We stress, however, that m_K is used only to determine numerical rank. It is independent of α, β, δ and ξ_e , and it is not the restriction of $\mathcal{G}_h$ to one element. For the local basis $\Phi_K = (\phi_{K,z_1}, \dots, \phi_{K,z_{r_K}})$ let $M_K = (m_K(\phi_{K,z_j}, \phi_{K,z_i}))_{ij} = Q_K \Lambda_K Q_K^*$, with $\lambda_{K,1} \geq \dots \geq 0$. The matrix M_K is the Gram matrix of the local Cauchy traces in the metric m_K . Hence M_K is Hermitian positive semidefinite and, for a coefficient vector $a \in \mathbb{C}^{r_K}$,

$$m_K(\Phi_K a, \Phi_K a) = a^* M_K a.$$

Thus an eigenvector $q_{K,j}$ gives a particular linear combination $\Phi_K q_{K,j}$ of the local plane waves, and $\lambda_{K,j}$ measures the squared size of its Cauchy trace in the metric m_K . If $\lambda_{K,j}$ is small, this combination has very small Dirichlet and normal traces on ∂K and contributes little independent information to the global skeleton problem. Since M_K is Hermitian, the eigenvectors are chosen Euclidean-orthonormal, so $Q_K^* Q_K = I$. In particular, $Q_{K,r}^* Q_{K,r} = I$, whereas $Q_{K,r}^* M_K Q_{K,r} = \Lambda_{K,r}$. Thus the retained eigenvectors are orthonormal in coefficient space but are not yet orthonormal in the trace metric. Multiplication by $\Lambda_{K,r}^{-1/2}$ converts them to an m_K -orthonormal trace basis. We retain the modes satisfying

$$\lambda_{K,j} > \tau_{\text{rank}} \lambda_{K,1}, \quad \tau_{\text{rank}} = 10^{-12} \quad (27)$$

in the reported calculations, and set $C_K = Q_{K,r} \Lambda_{K,r}^{-1/2}$. Then the retained local trace basis $\Phi_K C_K$ is m_K -orthonormal. With $C = \text{diag}(C_K)$ the compressed Galerkin matrices are

$$A_c = C^* A C, \quad f_c = C^* f, \quad G_c = C^* G C, \quad (28)$$

and the graph-Riesz solve is applied to (A_c, f_c, G_c) .

It is important that this compression changes the retained Galerkin space; it is not only a rescaling of coordinates. If $c_\perp$ lies in the discarded eigenspace, then

$$m_K(\Phi_K c_\perp, \Phi_K c_\perp) \leq \tau_{\text{rank}} \lambda_{K,1} \|c_\perp\|_2^2. \quad (29)$$

Thus r_K denotes the nominal local dimension and r_K^{eff} the retained trace rank. In the double-precision calculations below we initially take $\tau_{\text{rank}} = 10^{-12}$. This is only a numerical cutoff and is not part of the definition of the method. In Section 6.5 we decrease this value and also recompute the trace spectrum at higher precision. Those experiments show that useful high- p trace modes may lie below 10^{-12} , so the cutoff must be chosen together with the working precision and the accuracy required from the calculation.

4 Direction Optimization by the Skeleton Residual

4.1 The skeleton residual

We now introduce the residual that will be used in both direction optimization problems. For a coefficient vector c and a direction state Z , let

$$v_h(c, Z) = \sum_{j=1}^{N_h} c_j \varphi_j(Z)$$

be the associated broken Trefftz field. For the canonical constant-wavenumber model (1), define the direction objective

$$\begin{aligned} \mathcal{J}(c, Z) = & \sum_{e \in \mathcal{E}_h^I} \int_e \left[\alpha \kappa \left| \llbracket v_h(c, Z) \rrbracket \right|^2 + \frac{\beta}{\kappa} \left| \llbracket \nabla_h v_h(c, Z) \rrbracket_n \right|^2 \right] ds \\ & + \alpha \kappa \int_{\Gamma_D} |v_h(c, Z) - g|^2 ds \\ & + \frac{1}{2\kappa} \int_{\Gamma_R} \left| \partial_n v_h(c, Z) + i\kappa v_h(c, Z) \right|^2 ds. \end{aligned} \quad (30)$$

No volume residual is needed because $v_h(c, Z)$ satisfies the Helmholtz equation in every element. The interior terms measure the jumps of the field and normal flux, the Dirichlet term measures the boundary mismatch, and the final term measures the impedance condition on Γ_R . For piecewise-constant transmission experiments with $\Gamma_R = \emptyset$, the first two interior weights are replaced by $\alpha \xi_e$ and β/ξ_e , with ξ_e from (4), exactly as in the implemented PWDG form.

The factor $(2\kappa)^{-1}$ on Γ_R is a least-squares scaling, not a PWDG flux parameter. Indeed,

$$\frac{1}{2\kappa} |\partial_n v + i\kappa v|^2 = \frac{1}{2\kappa} |\partial_n v|^2 + \frac{\kappa}{2} |v|^2 + \text{Im}(\partial_n v \bar{v}). \quad (31)$$

For the symmetric Robin split $\delta = 1/2$, the first two diagonal terms have the same coefficients as the two separate Robin traces in $\mathcal{G}_h$, but the cross term in (31) remains. Thus $\mathcal{J}$ and $\mathcal{G}_h$ agree term by term only when $\Gamma_R = \emptyset$. On Γ_R , $\mathcal{J}$ measures the strong impedance mismatch while $\mathcal{G}_h$ is the coercive PWDG trace metric.

When $\Gamma_R = \emptyset$, as in the manufactured square-domain examples, the relation between the residual and the DG/graph norm is simpler and in fact exact.

Proposition 4.1 (exact graph-error identity for the all-Dirichlet specialization) *Assume $\Gamma_R = \emptyset$ and let u be the exact solution, with the usual continuity of u and normal flux across any material interfaces [21]. Then for every broken Trefftz field $v_h(c, Z)$,*

$$\mathcal{J}(c, Z) = \mathcal{G}_h(u - v_h(c, Z), u - v_h(c, Z)) = \|u - v_h(c, Z)\|_{\mathcal{G}_h}^2. \quad (32)$$

The identity is independent of how c is obtained. It requires, of course, that the same positive weights α and β be used in $\mathcal{J}$ and $\mathcal{G}_h$. This is how both quantities are defined here. All all-Dirichlet computations use the common choice $\alpha = \beta = 1/2$.

Proof Across an interior edge,

$$\llbracket u - v_h \rrbracket = -\llbracket v_h \rrbracket, \quad \llbracket \nabla(u - v_h) \rrbracket_n = -\llbracket \nabla_h v_h \rrbracket_n,$$

because the exact field and its normal flux are continuous. On the Dirichlet boundary $u = g$, so $u - v_h = g - v_h$. Substitution into (17) with $\Gamma_R = \emptyset$ gives (32) term by term. $\square$

All terms in Proposition 4.1 are computed directly from edge integrals for the plane-wave spaces used here. No pointwise sampling of the interior edges is required.

4.2 Residual control for mixed Dirichlet–impedance boundaries

For the mixed problem, the impedance residual contains a cross term and $\mathcal{J}$ is no longer identical to the PWDG graph norm. We can still obtain an L^2 error bound provided a stability estimate is available for broken Trefftz fields. We state this requirement explicitly.

Assumption 4.2 (broken Trefftz stability for the mixed problem) Let w be elementwise Helmholtz on $\mathcal{T}_h$. There is a constant $C_{\text{stab}}(\kappa, \Omega)$, independent of the particular broken representation of w , such that

$$\begin{aligned} \|w\|_{L^2(\Omega)} \leq C_{\text{stab}}(\kappa, \Omega) & \left(\sum_{e \in \mathcal{E}_h^I} [\kappa \| [w] \|_{L^2(e)}^2 + \kappa^{-1} \| [\nabla_h w]_n \|_{L^2(e)}^2] \right. \\ & \left. + \kappa \|w\|_{L^2(\Gamma_D)}^2 + \kappa^{-1} \|\partial_n w + i\kappa w\|_{L^2(\Gamma_R)}^2 \right)^{1/2}. \end{aligned} \quad (33)$$

Assumption 4.2 plays the same role as a Helmholtz stability estimate in a standard residual argument. Continuous impedance estimates give explicit information on the dependence on κ and the geometry [22]. We state the broken estimate as an assumption because the form above is not proved here as a direct consequence of the continuous result.

Proposition 4.3 (mixed-boundary residual control) *Under Assumption 4.2, let u solve (1) and let v_h be any broken Trefftz field. Then*

$$\|u - v_h\|_{L^2(\Omega)} \leq C_J(\alpha, \beta) C_{\text{stab}}(\kappa, \Omega) \mathcal{J}(v_h)^{1/2}, \quad (34)$$

where

$$C_J(\alpha, \beta) := \sqrt{\max\{\alpha^{-1}, \beta^{-1}, 2\}}. \quad (35)$$

In particular, $C_J(1/2, 1/2) = \sqrt{2}$ for the weights used in the numerical experiments.

Proof Set $w = u - v_h$. The exact field has zero interior jumps, satisfies $u = g$ on Γ_D , and satisfies $\partial_n u + i\kappa u = 0$ on Γ_R . Hence every trace residual of w is, up to sign, exactly the corresponding term appearing in (30). Comparing the four groups in (33) with (30) gives the factors α^{-1} , β^{-1} , α^{-1} , and 2, respectively. Therefore the unweighted quantity in (33) is bounded by $C_J(\alpha, \beta) \mathcal{J}(v_h)^{1/2}$ with (35), which proves the result. $\square$

For the all-Dirichlet problem we therefore have the identity (32), whereas for the mixed problem we obtain the upper bound (34). We do not prove a converse efficiency estimate. The estimate is nevertheless sufficient to justify using $\mathcal{J}$ as a computable error quantity for the mixed problem, with the qualification that the bound may become less sharp when the Helmholtz stability constant is large.

4.3 Evaluation on straight edges and local contributions

We next describe the computation of the residual matrices on a straight edge. Let the edge be parameterized by

$$x(t) = a + t\ell, \quad 0 \leq t \leq 1, \quad |e| = |\ell|.$$

For two local Trefftz functions

$$\phi_j(x) = e^{iq_j \cdot (x - x_{K_j})}, \quad \phi_i(x) = e^{iq_i \cdot (x - x_{K_i})},$$

including complex directions, one has the closed formula

$$\int_e \phi_j \overline{\phi_i} ds = |e| e^{iq_j \cdot (a - x_{K_j}) - i\overline{q_i} \cdot (a - x_{K_i})} \Psi(i(q_j - \overline{q_i}) \cdot \ell), \quad (36)$$

where

$$\Psi(\zeta) = \frac{e^\zeta - 1}{\zeta}, \quad \Psi(0) = 1.$$

In the implementation we evaluate $\Psi(\zeta)$ as $\text{expm1}(\zeta)/\zeta$ when $\zeta \neq 0$ and set $\Psi(0) = 1$. Normal derivatives introduce only the constant factors $iq_j \cdot n$ and $-i\overline{q_i} \cdot n$. Hence all basis–basis contributions from interior edges and homogeneous straight boundary edges are available in closed form. Differentiating with respect to a direction parameter only adds the first moment $\int_0^1 t e^{\zeta t} dt$, so the corresponding matrix derivatives are computed in the same way.

Numerical quadrature is used only for data-dependent terms for which no closed expression is available, for example

$$\int_{\Gamma_D} g \overline{\phi_i} ds, \quad \int_{\Gamma_D} |g|^2 ds,$$

when g has no convenient closed form. For manufactured plane-wave boundary data these terms may also be evaluated analytically. The interior residual matrix is therefore assembled from exact edge moments rather than from sampled residual values.

For fixed directions, expansion of (30) gives the Hermitian quadratic

$$\mathcal{J}(c, Z) = c^* M(Z) c - 2\Re\{c^* d(Z)\} + \gamma. \quad (37)$$

The weights in (30) are retained in each part of this expansion. With $v_h = \sum_j c_j \varphi_j$, the entries of the Gram matrix for the canonical constant-wavenumber problem are

$$\begin{aligned} M_{ij}(Z) = & \sum_{e \in \mathcal{E}_h^I} \left[\alpha \kappa \int_e \llbracket \varphi_j \rrbracket \cdot \overline{\llbracket \varphi_i \rrbracket} ds + \frac{\beta}{\kappa} \int_e \llbracket \nabla_h \varphi_j \rrbracket_n \overline{\llbracket \nabla_h \varphi_i \rrbracket_n} ds \right] \\ & + \alpha \kappa \int_{\Gamma_D} \varphi_j \overline{\varphi_i} ds \\ & + \frac{1}{2\kappa} \int_{\Gamma_R} (\partial_n \varphi_j + i\kappa \varphi_j) \overline{(\partial_n \varphi_i + i\kappa \varphi_i)} ds. \end{aligned} \quad (38)$$

The data terms are

$$d_i(Z) = \alpha \kappa \int_{\Gamma_D} g \overline{\varphi_i} ds, \quad \gamma = \alpha \kappa \int_{\Gamma_D} |g|^2 ds. \quad (39)$$

Thus α enters M , d , and γ , while β enters the normal-flux block of M . The parameter δ does not occur in (38): on Γ_R the optimization functional uses the impedance residual in (30), rather than the separately weighted Dirichlet and Neumann traces used in the PWDG graph form. For the piecewise-constant transmission problem, the two interior weights in (38) are replaced by $\alpha \xi_e$ and β/ξ_e . In the all-Dirichlet specialization, $M(Z)$ is precisely the matrix of the DG/graph inner product on the current Trefftz space. For the mixed Dirichlet–impedance problem it is instead the Gram matrix of the residual functional (30), and should not be confused with the PWDG graph matrix G .

For the all-Dirichlet problem, Proposition 4.1 also gives an elementwise error decomposition. Write

$$\eta_e^2 = \alpha \xi_e \|\llbracket v_h \rrbracket\|_{L^2(e)}^2 + \frac{\beta}{\xi_e} \|\llbracket \nabla_h v_h \rrbracket_n\|_{L^2(e)}^2, \quad e \in \mathcal{E}_h^I, \quad (40)$$

$$\eta_e^2 = \alpha \kappa \|v_h - g\|_{L^2(e)}^2, \quad e \subset \Gamma_D. \quad (41)$$

Assign half of each interior contribution to either adjacent element and the full boundary contribution to its element:

$$\eta_K^2 = \frac{1}{2} \sum_{e \subset \partial K \cap \mathcal{E}_h^I} \eta_e^2 + \sum_{e \subset \partial K \cap \Gamma_D} \eta_e^2. \quad (42)$$

Then

$$\sum_K \eta_K^2 = \mathcal{J}(c, Z) = \|u - v_h\|_{\mathcal{G}, h}^2. \quad (43)$$

Thus, for these tests, the numbers η_K localize the exact DG/graph error. We use them first to decide which elements require attention and, when only mesh refinement is allowed, in the usual Dörfler marking procedure [23]. Residual based PWDG refinement is studied in [9, 10].

4.4 Part A: direction-optimized PWDG as a constrained minimization problem

We first consider the formulation in which the PWDG equations are retained during the direction search. For each admissible direction state Z , let $c_{\text{PW}}(Z)$ be the coefficient vector obtained from the fixed-direction PWDG problem,

$$A(Z) c_{\text{PW}}(Z) = f(Z), \quad (44)$$

computed in practice by graph–Riesz normalization. Define the reduced continuous PWDG skeleton-residual objective

$$\hat{\mathcal{J}}_{\text{PW}}(Z) := \mathcal{J}(c_{\text{PW}}(Z), Z). \quad (45)$$

$$Z_{\text{PW}}^* \in \arg \min_{Z \in \mathcal{Z}_h(\mathbf{r}, \eta_{\max})} \widehat{\mathcal{J}}_{\text{PW}}(Z), \quad (46)$$

Changes of local graph rank are treated as outer active-set events. Equivalently, before compression events are introduced, it is the equality-constrained problem

$$\min_{c, Z} \mathcal{J}(c, Z) \quad \text{subject to} \quad A(Z)c = f(Z). \quad (47)$$

Equation (47) has the form of a finite-dimensional equality-constrained optimization problem. Following the real-angle PWDG control formulation of Agrawal and Hoppe [13], we introduce a complex Lagrange multiplier μ . The directions are now allowed to be complex and the objective is the skeleton residual. For real direction coordinates ρ , define

$$\mathcal{L}(c, \rho, \mu) = \mathcal{J}(c, \rho) + 2\Re[\mu^*(A(\rho)c - f(\rho))]. \quad (48)$$

On a fixed-rank stratum, the first-order constrained stationarity conditions are

$$A(\rho)c = f(\rho), \quad (49)$$

$$\partial_{\bar{c}}\mathcal{J} + A(\rho)^*\mu = 0, \quad (50)$$

$$\partial_{\rho_\ell}\mathcal{J}|_c + 2\Re[\mu^*(A_{\rho_\ell}c - f_{\rho_\ell})] = 0 \quad (\ell = 1, \dots, m). \quad (51)$$

Since c is constrained by the PWDG equations, $\partial_{\bar{c}}\mathcal{J}$ is not required to vanish. Instead, (50) states that this gradient is balanced by $A^*\mu$. We later use the equivalent adjoint variable $\lambda = -\mu$ when evaluating the reduced derivatives.

The multiplier can also be interpreted as the sensitivity to a perturbation of the discrete PWDG equation. Suppose that

$$A(\rho)c = f(\rho) + \varepsilon r.$$

Differentiating the constrained objective with respect to ε gives the dual pairing with μ , with the sign determined by (48). Thus μ measures the change in the optimized residual produced by a defect in the discrete state equation. It is an adjoint variable and is not another approximation of the Helmholtz solution. This point is also useful when comparing the two formulations. In Part B the coefficients are unconstrained least-squares variables and satisfy $\partial_{\bar{c}}\mathcal{J} = 0$, while in Part A this gradient is balanced by $A^*\mu$. If the residual is zero and A is nonsingular, then $\partial_{\bar{c}}\mathcal{J} = 0$ and consequently $\mu = \lambda = 0$.

The bounds on the angles do not by themselves guarantee that $Z \mapsto c_{\text{PW}}(Z)$ is well defined for every admissible point. For example, two local directions can coalesce and make the coefficient representation singular. For existence statements we therefore restrict to a rank-safe closed subset

$$\mathcal{Z}_h^\delta := \{Z \in \mathcal{Z}_h(\mathbf{r}, \eta_{\max}) : \lambda_{\min}(G(Z)) \geq \delta\}, \quad \delta > 0, \quad (52)$$

with a fixed active basis. On $\mathcal{Z}_h^\delta$, $A(Z)$ is nonsingular, $c_{\text{PW}}(Z)$ and $\widehat{\mathcal{J}}_{\text{PW}}(Z)$ are continuous, and a minimizer exists by compactness. In the actual algorithm, approaching a rank-deficient configuration triggers local trace compression rather than an attempt to differentiate through the rank change.

In the complex-angle transmission calculation we add $10^{-12} \sum_j \eta_j^2$ to the nonlinear objective. This very small term is used only to select between residual-equivalent representations. It is not included in the reported values of $\mathcal{J}$.

4.5 Local analysis for fixed rank

We now give a local result for the reduced direction problem. There are two reasons for restricting the statement to a fixed-rank neighborhood. A permutation of the rays on one element does not change the local Trefftz space, so the ordered direction vector is not uniquely determined. Also, if two rays coalesce or a trace becomes redundant, the dimension of the retained space changes. We therefore fix the active rank and measure the distance to the finite set of permutations of the target direction state.

We assume that the residual determines the directions locally. More precisely, if $R(\rho)$ denotes a smooth residual coordinate vector, we require $DR(\rho_*)$ to have full column rank after one representative of the permutation orbit has been fixed. Equivalently, $DR(\rho_*)^*DR(\rho_*)$ is positive definite. This assumption rules out a nonzero first-order change of the directions that leaves the residual unchanged. It is not satisfied when rays coalesce or become redundant, which is why those cases are treated separately by the rank reduction.

Let $\rho \in \mathbb{R}^m$ collect the free real direction coordinates and define

$$U(\rho) := v_h(c_{\text{PW}}(\rho), \rho).$$

Local ray permutations represent the same Trefftz space. We therefore measure distance to the finite permutation orbit $\mathcal{O}(\rho_\star)$ and work on a neighborhood in which the active trace rank is fixed.

Theorem 4.4 (local quadratic growth and direction-to-field stability) *Let $\rho_\star$ satisfy $\widehat{\mathcal{J}}_{\text{PW}}(\rho_\star) = 0$. Assume that, in a neighborhood of $\rho_\star$, the active graph rank is constant, $A(\rho)$ is uniformly invertible, and a smooth residual coordinate map R with*

$$\widehat{\mathcal{J}}_{\text{PW}}(\rho) = \|R(\rho)\|_2^2$$

has full-column-rank Jacobian $DR(\rho_\star)$. Then there are constants $c_0, C_0, C_U > 0$ such that

$$c_0 \text{dist}(\rho, \mathcal{O}(\rho_\star))^2 \leq \widehat{\mathcal{J}}_{\text{PW}}(\rho) \leq C_0 \text{dist}(\rho, \mathcal{O}(\rho_\star))^2, \quad (53)$$

and

$$\|U(\rho) - U(\rho_\star)\|_{\mathcal{G},h} \leq C_U \text{dist}(\rho, \mathcal{O}(\rho_\star)) \quad (54)$$

for ρ sufficiently close to $\rho_\star$. Consequently, standard Gauss–Newton or trust-region/Levenberg–Marquardt iteration is locally convergent, with the usual quadratic zero-residual Gauss–Newton behavior once the full-rank neighborhood is entered [25].

Proof With the active rank fixed, the plane-wave traces depend smoothly on ρ , and consequently so do $A(\rho)$ and $f(\rho)$. The assumed uniform invertibility of $A(\rho)$ then gives a smooth coefficient map $c_{\text{PW}}(\rho) = A(\rho)^{-1}f(\rho)$. Hence the field $U(\rho)$ and residual vector $R(\rho)$ are also smooth.

Choose the nearest representative $\tilde{\rho}_\star$ in the finite permutation orbit $\mathcal{O}(\rho_\star)$ and set $d = \rho - \tilde{\rho}_\star$. Because $R(\tilde{\rho}_\star) = 0$, Taylor expansion gives

$$R(\rho) = DR(\tilde{\rho}_\star)d + \mathcal{O}(|d|^2).$$

Since $DR(\tilde{\rho}_\star)$ has full column rank, its smallest and largest singular values give positive lower and upper bounds for the linear term. After reducing the neighborhood, the quadratic remainder can be absorbed into these bounds and we obtain $c_0|d|^2 \leq \|R(\rho)\|_2^2 \leq C_0|d|^2$, which proves (53). Applying a first-order Taylor estimate to U gives (54). The final statement follows from the usual local theory for full-rank zero-residual least squares [25]. $\square$

If rays coalesce or become redundant, the fixed-rank hypothesis is no longer satisfied. In the algorithm this is treated by recomputing the local trace rank as described in Section 3.2; we do not differentiate through the change of rank.

4.6 Relation to the Monk–Wang least-squares method

Monk and Wang [2] prescribe discontinuous local Helmholtz solutions and minimize a quadratic skeleton functional over their coefficients. Their interior term uses the full gradient jump, whereas (30) uses the normal-flux DG jump aligned with the PWDG graph form. Here a different, purely oriented notation is useful. Fix one unit normal n_e and tangent t_e on a straight edge and define the ordinary scalar difference

$$[v] := v^+ - v^-.$$

Then $[\cdot]$ depends on this chosen orientation, unlike the DG double-line jump above, and

$$[\nabla v] = [\partial_{n_e} v] n_e + \partial_{t_e} [v] t_e.$$

The residual normalization in (30) is chosen so that this comparison has no additional arbitrary constants. For constant κ , $\alpha = \beta = 1/2$, and matching impedance sign conventions, multiplication of (30) by 2κ gives unit weight on the normal-flux jump and impedance residual and weight κ^2 on the solution jump and Dirichlet residual. Decomposing the full gradient jump into its normal and tangential parts therefore yields

$$\mathcal{J}_{\text{MW}}(c, Z) = 2\kappa \mathcal{J}(c, Z) + \sum_{e \in \mathcal{E}_h^1} \|\partial_{t_e} [v_h(c, Z)]\|_{L^2(e)}^2. \quad (55)$$

With this scaling, the relation between the two functionals is given exactly by (55). The present functional retains the normal component of the gradient jump, whereas the Monk–Wang functional also contains the tangential derivative of the solution jump. It should therefore be viewed as a normal-flux version of the Monk–Wang residual, and not as the same functional.

For fixed directions, the coefficient least-squares solution is therefore characterized directly by the exact Hermitian system

$$M(Z)c_{\text{LS}}(Z) = d(Z), \quad c_{\text{LS}}(Z) = \arg \min_{c \in \mathbb{C}^{N_h}} \mathcal{J}(c, Z). \quad (56)$$

In Part B we keep the coefficient projection (56) but allow its local directions to vary. For every fixed Z ,

$$\mathcal{J}(c_{\text{LS}}(Z), Z) \leq \mathcal{J}(c_{\text{PW}}(Z), Z), \quad (57)$$

but this residual ordering need not agree with the L^2 -error ordering.

For Part A, the equality constraint $A(Z)c = f(Z)$ may be expressed with

$$\mathcal{L}(c, Z, \lambda) = \mathcal{J}(c, Z) - 2\Re\{\lambda^*(A(Z)c - f(Z))\}.$$

Stationarity in c gives the adjoint equation $A^*\lambda = \partial_c \mathcal{J}$. The resulting reduced direction derivative is given below. The adjoint term is present because the Part A coefficients must remain on the PWDG solution manifold; it is absent from the free coefficient minimization in Part B.

4.7 Differentiation with respect to the directions

To differentiate with respect to a complex angle we use the Wirtinger derivatives [24], in which z and $\bar{z}$ are treated as independent variables for the purpose of differentiation. For $z = \theta + i\eta$,

$$z = \theta + i\eta, \quad \partial_z = \frac{1}{2}(\partial_\theta - i\partial_\eta), \quad \partial_{\bar{z}} = \frac{1}{2}(\partial_\theta + i\partial_\eta).$$

If J is real valued, then $\partial_{\bar{z}} J = \overline{\partial_z J}$. The notation is particularly convenient here because a trial plane wave depends holomorphically on z , while the conjugated plane wave in the test position depends on $\bar{z}$. For

$$q(z) = \kappa(\cos z, \sin z), \quad q'(z) = \kappa(-\sin z, \cos z),$$

the local Trefftz basis function

$$\phi_z(x) = \exp(iq(z) \cdot (x - x_K))$$

is entire in z . Hence

$$\partial_z \phi_z = i q'(z) \cdot (x - x_K) \phi_z, \quad \partial_{\bar{z}} \phi_z = 0. \quad (58)$$

The conjugated test function satisfies

$$\partial_z \overline{\phi_z} = 0, \quad \partial_{\bar{z}} \overline{\phi_z} = -i \overline{q'(z)} \cdot (x - x_K) \overline{\phi_z}. \quad (59)$$

The gradient and normal-trace derivatives follow by the product rule:

$$\begin{aligned} \partial_z \nabla \phi_z &= i q'(z) \phi_z + i q(z) \partial_z \phi_z \\ &= \left(i q'(z) - q(z) [q'(z) \cdot (x - x_K)] \right) \phi_z, \end{aligned} \quad (60)$$

$$\partial_z \partial_n \phi_z = i (q'(z) \cdot n) \phi_z + i (q(z) \cdot n) \partial_z \phi_z. \quad (61)$$

These tangent functions remain Trefftz because ∂_z commutes with $-\Delta - \kappa^2$.

Although each plane wave is holomorphic in its own angle, the Galerkin matrix depends on both Z and $\bar{Z}$, since a_h is conjugate-linear in the test argument. For independent direction parameters and

$$A_{ij}(Z, \bar{Z}) = a_h(\phi_{z_j}, \phi_{z_i}), \quad f_i(Z, \bar{Z}) = F_h(\phi_{z_i}),$$

we obtain

$$\partial_{z_\ell} A_{ij} = \delta_{j\ell} a_h(\partial_z \phi_{z_j}, \phi_{z_i}), \quad (62)$$

$$\partial_{\bar{z}_\ell} A_{ij} = \delta_{i\ell} a_h(\phi_{z_j}, \partial_z \phi_{z_i}), \quad (63)$$

$$\partial_{z_\ell} f_i = 0, \quad \partial_{\bar{z}_\ell} f_i = \delta_{i\ell} F_h(\partial_z \phi_{z_i}). \quad (64)$$

For shared direction parameters these formulas are summed over the corresponding basis functions. Equation (62) differentiates the trial column associated with z_ℓ , whereas (63) differentiates the corresponding test row.

At fixed active rank,

$$A(Z, \bar{Z})c(Z, \bar{Z}) = f(Z, \bar{Z}).$$

Differentiation gives two coefficient sensitivities per complex angle,

$$Ac_{z_\ell} = -A_{z_\ell}c, \quad (65)$$

$$Ac_{\bar{z}_\ell} = f_{\bar{z}_\ell} - A_{\bar{z}_\ell}c. \quad (66)$$

The real sensitivities are

$$c_{\theta_\ell} = c_{z_\ell} + c_{\bar{z}_\ell}, \quad c_{\eta_\ell} = i(c_{z_\ell} - c_{\bar{z}_\ell}). \quad (67)$$

More generally, for either real parameter $\rho_\ell \in \{\theta_\ell, \eta_\ell\}$,

$$A c_{\rho_\ell} = f_{\rho_\ell} - A_{\rho_\ell}c. \quad (68)$$

The same factorization of A may be used for all the sensitivity right-hand sides. We differentiate the physical system $Ac = f$, and hence do not need to differentiate the graph-coordinate map B . During one smooth direction step the compression matrix C is fixed. If the local rank changes, the active space is recomputed before the next step.

For the exact skeleton quadratic (37), let M_ρ and d_ρ denote the derivatives induced by a real direction parameter $\rho \in \{\theta_\ell, \eta_\ell\}$. Combining the product rule with the coefficient sensitivity (68) gives the exact reduced derivative

$$\frac{d\hat{\mathcal{J}}_{\text{PW}}}{d\rho} = 2\Re[c_\rho^*(Mc - d)] + c^*M_\rho c - 2\Re(c^*d_\rho). \quad (69)$$

Interior terms in M_ρ follow by differentiating the closed edge moment (36); only nonanalytic boundary-data terms require quadrature.

If many direction variables are present, solving (68) separately for every parameter is unnecessary. An adjoint equation gives all derivatives after one additional global solve. If $g_c = \partial\mathcal{J}/\partial\bar{c}$, the Part A stationarity equation (50) gives $A^*\mu = -g_c$. With $\lambda = -\mu$,

$$A^*\lambda = g_c. \quad (70)$$

then for any real direction parameter ρ_ℓ the reduced derivative can be written

$$\frac{d\mathcal{J}}{d\rho_\ell} = \left. \frac{\partial\mathcal{J}}{\partial\rho_\ell} \right|_c + 2\Re[\lambda^*(f_{\rho_\ell} - A_{\rho_\ell}c)]. \quad (71)$$

After λ has been computed, the derivative with respect to every local direction parameter is obtained from the corresponding local matrix and load derivatives.

4.8 Part B: Trefftz least squares with variable directions

In the second formulation we remove the PWDG state equation and minimize the skeleton residual directly. This is related to the wave-tracking least-squares approach of Amara et al. [14], except that we eliminate the linear coefficients by variable projection. The problem is

$$(c^*, Z^*) \in \arg \min_{c, Z} \mathcal{J}(c, Z). \quad (72)$$

For fixed Z , (37) is quadratic in c and gives the usual linear Trefftz least-squares problem. The difference from Monk–Wang is that the local Trefftz functions are no longer fixed, since their directions are also unknown.

Full-space first-order conditions. Because $M = M^*$, Wirtinger differentiation with respect to the free coefficient vector gives

$$\partial_{\bar{c}}\mathcal{J} = Mc - d. \quad (73)$$

For a real direction parameter ρ , the product rule gives, with c held fixed,

$$\partial_\rho\mathcal{J} = c^*M_\rho c - 2\Re(c^*d_\rho). \quad (74)$$

Equivalently, in complex-angle coordinates the $\bar{z}_\ell$ equation is obtained by differentiating the test-side factors in the exact edge moments. The underlying basis derivative remains

$$\partial_z\phi_z = i q'(z) \cdot (x - x_K)\phi_z,$$

so (74) is exactly the product-rule condition expressing orthogonality of the residual functional to the directional tangent generated by $c_j \partial_z \phi_{z_j}$. A joint stationary point therefore satisfies the coefficient projection equation and all directional tangent equations simultaneously.

Variable projection. Since the dependence on c is linear, it is unnecessary to include the coefficients in the nonlinear iteration. We use the variable-projection reduction of Golub and Pereyra [3]. On a fixed-rank stratum,

$$M(Z)c_{\text{LS}}(Z) = d(Z), \quad (75)$$

with the minimum-norm solution used if the active matrix is singular. When M is positive definite,

$$\hat{\mathcal{J}}_{\text{LS}}(Z) = \gamma - d(Z)^* M(Z)^{-1} d(Z). \quad (76)$$

The reduced and full minimization problems have the same minimum, while the nonlinear problem contains only the direction variables. The envelope derivative contains no coefficient sensitivity:

$$\frac{d\hat{\mathcal{J}}_{\text{LS}}}{d\rho} = c_{\text{LS}}^* M_\rho c_{\text{LS}} - 2\Re(c_{\text{LS}}^* d_\rho). \quad (77)$$

Equation (77) is used in the exact-edge implementation. When the residual is very small, we evaluate the objective from the positive edge contributions in (40), since direct evaluation of (76) can lose accuracy through cancellation.

Krylov realization. When the active residual seminorm is a norm on the retained Trefftz space, M is Hermitian positive definite and (75) can be solved directly by preconditioned conjugate gradients [2, 26], without normal equations. Trace-rank compression from Section 3.2 is applied if moving directions produce numerical rank loss. A symmetric eigendecomposition is used as a small-problem fallback.

When several directions are used on each element, optimizing every angle independently can give a large nonlinear problem. We therefore also use the structured fan

$$z_{K,j} = \mu_K + s_K \xi_j, \quad j = 1, \dots, r_K, \quad (78)$$

with fixed equally spaced centered nodes. In the reported implementation, $\xi_j = -1 + 2(j-1)/(r_K-1)$ for $r_K > 1$, while $\xi_1 = 0$ for $r_K = 1$. Thus s_K is the half-width of the angular fan. The chain rule is applied directly to M_ρ and d_ρ . The nonlinear dimension is then two fan parameters per element, independent of the number of linear coefficients.

4.9 Frequency continuation

The numerical experiments below show that direct minimization at the target wavenumber can converge to a false local minimum. To reduce this difficulty we use frequency continuation [30]. For

$$0 < \kappa^{(0)} < \dots < \kappa^{(M)} = \kappa_{\text{target}},$$

the converged directions at $\kappa^{(m)}$ initialize the solve at $\kappa^{(m+1)}$. Blind starts at the lowest frequency use Sobol points [29]. The $\omega = 12$ transmission tests use 1, 2, 4, 6, 8, 10, 12. Only the direction variables are passed from one frequency to the next. No auxiliary low-frequency field is postprocessed to estimate rays, as is done in the learning procedure of Fang et al. [15].

5 Numerical Algorithms

Before giving the numerical results, we summarize the two procedures used in the computations. Both use the same direction parameterization and the same skeleton residual. The difference is how the coefficients are obtained. Part A solves the PWDG equations for each direction state, whereas Part B solves the coefficient least-squares problem and removes it from the nonlinear iteration by variable projection.

Algorithm 1 Direction-optimized PWDG (Part A)

- 1: Choose an admissible initial direction state and, when needed, a frequency-continuation schedule.
 - 2: **for** each continuation frequency **do**
 - 3: Assemble the PWDG matrices and compress graph-small local trace modes by (27).
 - 4: Solve the retained PWDG system in graph–Riesz coordinates.
 - 5: Minimize $\hat{\mathcal{J}}_{\text{PW}}(Z)$ over the active directions using (69).
 - 6: Use the converged directions to initialize the next frequency.
 - 7: **end for**
 - 8: **return** the PWDG field, directions, effective trace ranks, $\kappa_2(A)$, and κ_{GR} .
-

Algorithm 2 Exact-edge variable-projection Trefftz least squares (Part B)

- 1: Choose an initial direction state and, when needed, a continuation schedule.
 - 2: **repeat**
 - 3: Assemble the exact skeleton matrix $M(Z)$ by (36) and assemble $d(Z)$ and γ using boundary-data quadrature only when required.
 - 4: Solve $M(Z)c = d(Z)$ by preconditioned CG on a full-rank active space.
 - 5: Assemble M_ρ and d_ρ from differentiated exponential moments and evaluate (77).
 - 6: Apply a bounded nonlinear step to the direction variables only.
 - 7: **until** the objective or direction step satisfies the stopping criterion
 - 8: **return** Z , $c_{\text{LS}}(Z)$, and the positive edge sum $\mathcal{J} = \sum_e \eta_e^2$.
-

The nonlinear objective in both algorithms is nonconvex. In examples where a direct solve at the final wavenumber is unreliable, we use frequency continuation. We also keep changes of numerical rank outside the smooth direction step: after an accepted state the local trace ranks are recomputed and, if necessary, the active space is changed.

6 Numerical experiments

6.1 Implementation details and derivative checks

Unless stated otherwise, all computations in this section are carried out in IEEE double precision on triangular meshes. For the manufactured square-domain examples we take $\Gamma_R = \emptyset$, so Proposition 4.1 gives $\mathcal{J}$ as the squared DG/graph error. The plane-wave products on interior edges, including those used in M and in the direction derivatives, are evaluated from (36). Numerical quadrature is used only for boundary terms for which a closed form is not available. In the transmission problem the wavenumber is piecewise constant and continuity is imposed across $y = 0$; the interior penalties use the arithmetic face scale (4). We take $\alpha = \beta = 1/2$ unless another value is stated. These parameters do not enter the local rank metric (26). We begin the double-precision calculations with $\tau_{\text{rank}} = 10^{-12}$ and revisit this choice in the high- p DtN experiment.

We report

$$E_{L^2} = \frac{\|u - u_h\|_{L^2(\Omega)}}{\|u\|_{L^2(\Omega)}}.$$

As a check on the implementation, the analytic direction derivatives were compared with centered finite differences at points that were not stationary. Differentiation of the exact exponential moments agrees with finite differences at approximately 10^{-9} – 10^{-10} relative accuracy in the transmission tests. The original PWDG matrix and coefficient-sensitivity checks remain below 6×10^{-9} .

Nonlinear solver and constraints.

Part A uses trust-region nonlinear least squares with analytic residual derivatives [25]. Angular variables are represented modulo 2π , with $|\eta_j| \leq 1.8$. After each accepted state the trace Gramians are recomputed and compression is applied before the next fixed-rank solve. Part B uses the same framework after variable projection. Stopping tests use the scaled residual and step norms. The reported value of N_{fev} is the number of nonlinear residual evaluations. Since the derivatives are analytic, no additional finite-difference residual evaluations are included.

6.2 Three hidden plane waves

We first consider an exact solution consisting of three plane waves,

$$u(x) = \sum_{j=1}^3 A_j e^{i\kappa d_j \cdot x}, \quad (79)$$

where the three directions are 17° , 123° , and 251° . These angles are used to generate the boundary data but are not supplied to the nonlinear solver. The direction search starts from the same Sobol low-discrepancy angles [29], 43.929° , 130.738° , and 269.353° , for every wavenumber. Table 2 uses a uniform mesh with 72 triangles and 216 PWDG coefficients. For all three wavenumbers the nonlinear iteration recovers the three directions to the digits shown. After this has occurred, the exact solution belongs to the discrete Trefftz space and the remaining L^2 error is at the level of roundoff.

Table 2 Three-wave recovery on a uniform mesh with 72 triangles.

κ	uniform $p = 3$ error	optimized error	N_{fev}	$\kappa_2(A)$	κ_{GR}
4	1.1875×10^{-1}	1.25×10^{-15}	6	106.69	7.49
8	3.7541×10^{-1}	1.19×10^{-15}	8	37.61	5.98
12	5.5874×10^{-1}	2.01×10^{-15}	12	40.35	5.47

The preceding calculation is an exact recovery problem, but the corresponding nonlinear minimization is not globally easy. For a second hidden triple ($41^\circ, 177^\circ, 303^\circ$), a direct $\kappa = 8$ solve from a generic start converges to ($279.080^\circ, 143.976^\circ, 9.238^\circ$) with $E_{L^2} = 1.1513$ and $\mathcal{J} = 50.43$. Continuation from $\kappa = 1$ to 8 recovers the exact triple and gives $E_{L^2} = 2.17 \times 10^{-15}$ in 21 cumulative residual evaluations.

The corresponding direct solve required 67 residual evaluations. In this case continuation reaches a different local minimum and also uses fewer residual evaluations. The comparison is only one initialization and does not give a success probability for the nonlinear solver. It does, however, show why the analysis in Section 4.5 is local and why continuation is used in the more difficult experiments below.

6.3 Transmission at $\omega = 12$

We next consider the transmission problem on $(-1, 1)^2$. The purpose of this example is to check the complex-angle parameterization in the two cases that occur at an interface: a propagating transmitted wave and an evanescent one. We take $n_1 = 2$ below $y = 0$ and $n_2 = 1$ above it, so that $\kappa_j = \omega n_j$ with $\omega = 12$. There are two plane-wave components in the lower material and one in the upper material. The exact transmission solution is used only to prescribe the boundary data and to measure the error. A direct solve at $\omega = 12$ can converge to a false minimum, so we use the continuation sequence

$$1, 2, 4, 6, 8, 10, 12.$$

For 69° incidence the recovered directions are

$$69^\circ, \quad 291^\circ, \quad 44.2143563895^\circ, \quad (80)$$

with imaginary components at roundoff. Tangential phase matching predicts

$$\theta_t = \arccos\left(\frac{\kappa_1 \cos 69^\circ}{\kappa_2}\right) = 44.2143563895^\circ, \quad (81)$$

which agrees with the recovered direction in the upper medium to the digits shown.

At 29° incidence, total internal reflection gives

$$z_1 = 29^\circ + O(10^{-16})i, \quad z_2 = 331^\circ + O(10^{-16})i, \quad (82)$$

and

$$z_3 = 0 + 1.15828058554i. \quad (83)$$

The physical decay parameter is

$$\eta_{\text{tr}} = \operatorname{arccosh}\left(\frac{\kappa_1 \cos 29^\circ}{\kappa_2}\right) = 1.1582805855, \quad (84)$$

The computed imaginary part agrees with (84). Thus the same direction search finds the decaying branch without a critical-angle test or a separate evanescent ansatz.

The real parts of the two solutions are shown in Fig. 1. The mesh plays essentially no approximation role once the three directions have been recovered. This is worth displaying explicitly. Figure 2 shows the 8- and 32-triangle meshes used in Table 3. Already on the 8-triangle mesh the relative L^2 errors are 1.41×10^{-15} in the propagating case and 2.00×10^{-15} in the evanescent case. Refining to 32 triangles does not improve these numbers in any meaningful way; both calculations are already at working precision. This is exactly what should happen: after the correct directions are found, the manufactured transmission field belongs to the piecewise Trefftz space on either mesh.

The conditioning tells a different story. For the 8-triangle evanescent calculation, $\kappa_2(A) = 1.22 \times 10^9$, whereas $\kappa_{\text{GR}} = 2.25$. Thus the coarse mesh is already sufficient for approximation, but the choice of coefficient coordinates still has a large effect on the algebraic problem.

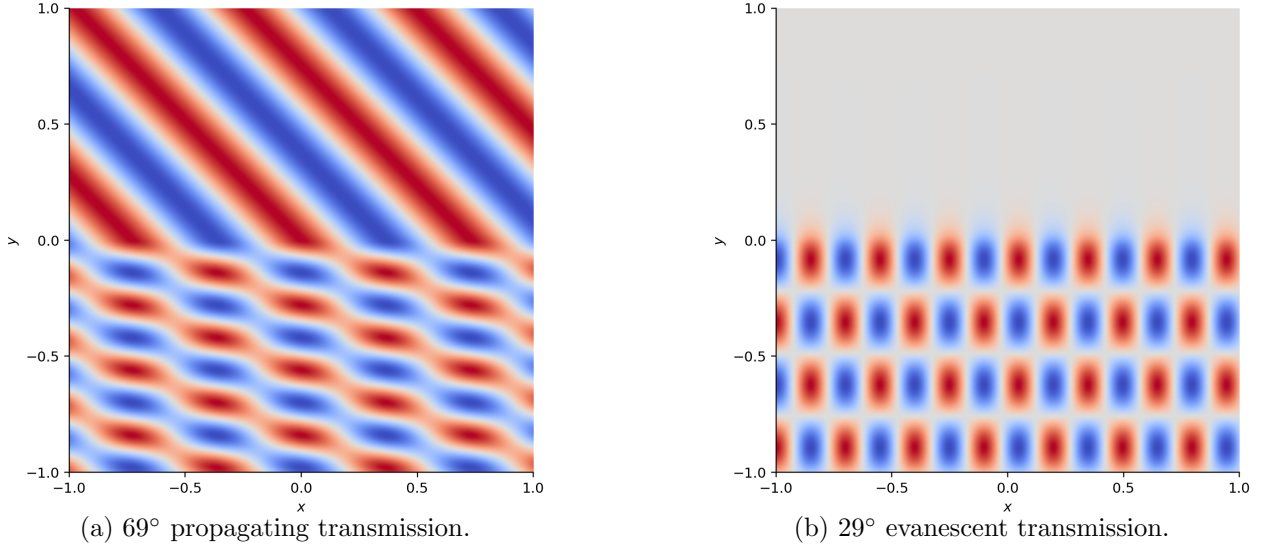


Fig. 1 Real parts of the $\omega = 12$ transmission fields.

Table 3 Part A transmission recovery at $\omega = 12$ on uniform square meshes. N_{fev} is cumulative over continuation.

triangles	incidence	coefficients	N_{fev}	E_{L^2}	$\kappa_2(A)$	κ_{GR}
8	69°	12	15	1.41×10^{-15}	10.06	1.84
32	69°	48	15	1.13×10^{-15}	22.95	2.92
8	29°	12	21	2.00×10^{-15}	1.22×10^9	2.25
32	29°	48	20	4.26×10^{-15}	4.28×10^4	3.52

From Table 3 we also observe that the raw condition number of the 29° calculation decreases on the finer mesh. This is consistent with the smaller elementwise exponential variation of the evanescent basis. The graph–Riesz condition number remains small on both meshes.

6.4 Exact circular DtN boundary: Hankel waves and the limitation of a sparse local space

We next return to the circular artificial boundary used in the DtN–PWDG experiments of [8]. We take a sound-soft circle of radius $a = 0.5$ and a concentric artificial boundary $r = R$ with $R = 1$. As in the earlier work, both circular boundaries are represented by exact curved edges so that there is no polygonal geometry error. With the time dependence $e^{+i\omega t}$, outgoing cylindrical modes are represented by Hankel functions of the second kind.

Transmission meshes; labels are relative L^2 errors

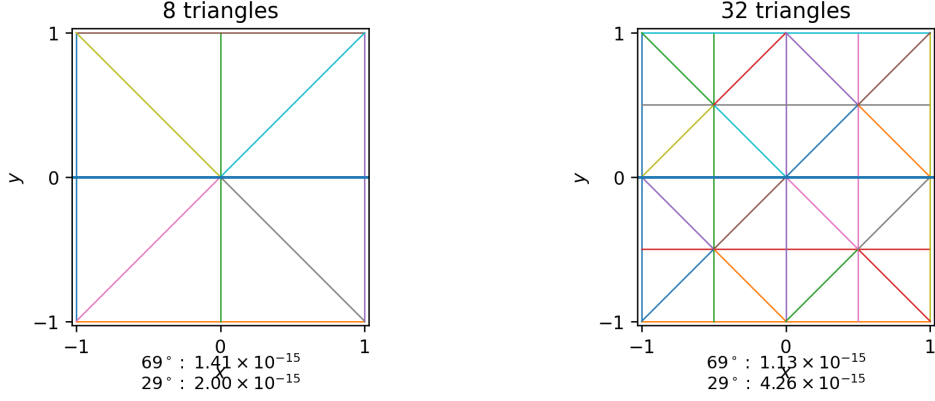


Fig. 2 The two uniform transmission meshes used in Table 3. The interface is $y = 0$. The numbers printed below each mesh are the relative L^2 errors for the 69° and 29° calculations. The 8-triangle mesh is already at machine precision in both cases; refinement to 32 triangles has no observable approximation effect.

For $v(R, \theta) = \sum_{m \in \mathbb{Z}} v_m e^{im\theta}$, the truncated DtN map is

$$S_N v = \sum_{|m| \leq N} k \frac{H_m^{(2)'}(kR)}{H_m^{(2)}(kR)} v_m e^{im\theta}, \quad v_m = \frac{1}{2\pi} \int_0^{2\pi} v(R, \phi) e^{-im\phi} d\phi. \quad (85)$$

On Γ_R we use the DtN fluxes from [8],

$$\begin{aligned} \hat{u}_h &= u_h - \frac{\delta}{ik} (\partial_n u_h - S_N u_h), \\ ik \hat{\sigma}_h &= S_N u_h n - \frac{\delta}{ik} S_N^* (\nabla_h u_h - S_N u_h n). \end{aligned} \quad (86)$$

For direction selection the boundary part of the skeleton objective is replaced by the nonlocal residual

$$\mathcal{J}_{R,N}(v_h) = \frac{\delta_D}{k} \|\partial_n v_h - S_N v_h\|_{L^2(\Gamma_R)}^2, \quad (87)$$

where δ_D denotes the optimization weight and is kept distinct from the flux parameter δ . The DtN term couples all outer boundary edges, so the corresponding part of $M(Z)$ is a global boundary block rather than an edge-local contribution.

We first use the DtN problem to expose a limitation of direction movement. We ask whether a small local fan can follow the phase of an outgoing cylindrical field and, if it can, whether this is enough to give a small field error. We consider the two exact solutions

$$u(x) = H_0^{(2)}(k|x - x_s|), \quad k = 8,$$

with $x_s = 0$ and with an off-center source x_s lying inside the removed disk. In both cases the exact field satisfies the homogeneous Helmholtz equation in the annulus and the exact DtN condition on $r = R$. The arrows used below represent the learned plane-wave wave vectors q , equivalently the local spatial phase-gradient directions. This distinction is important for the present time convention: since $e^{+i\omega t} H_0^{(2)}(kr) \sim e^{i\omega t - ikr} / \sqrt{r}$, an outgoing cylindrical wave has q directed toward the source, whereas its physical energy propagates outward. Thus the plotted arrows are not energy-flow arrows. We use one radial layer, eight exact curved sectors, and a three-wave local fan with offsets $(-18^\circ, 0, 18^\circ)$; only the fan center is moved. Frequency continuation uses $k = 1, 2, 4, 8$, and $N = 16$ at the target frequency.

The same exact annular mesh is used throughout the DtN experiments. For reference, Fig. 3 shows this mesh together with the exact scattered field for the standard sound-soft disk problem with incident wave

$u^{\text{inc}} = e^{ikx}$. With the present $e^{+i\omega t}$ convention, the scattered field is

$$u^{\text{scat}}(r, \theta) = -\frac{J_0(ka)}{H_0^{(2)}(ka)}H_0^{(2)}(kr) - 2\sum_{m=1}^{\infty} i^m \frac{J_m(ka)}{H_m^{(2)}(ka)}H_m^{(2)}(kr)\cos(m\theta). \quad (88)$$

The field shown in Fig. 3 uses $M = 20$ terms, which is more than sufficient for the visualization at $k = 8$. This truncation is used only for the plot; the reported numerical errors use the converged reference series in the computational code. The Hankel-source calculations below are retained as local phase-recovery diagnostics. The high- p experiment in Section 6.5 returns to this plane-wave-incidence scattering problem.

Plane-wave scattering by the sound-soft disk, $k = 8$

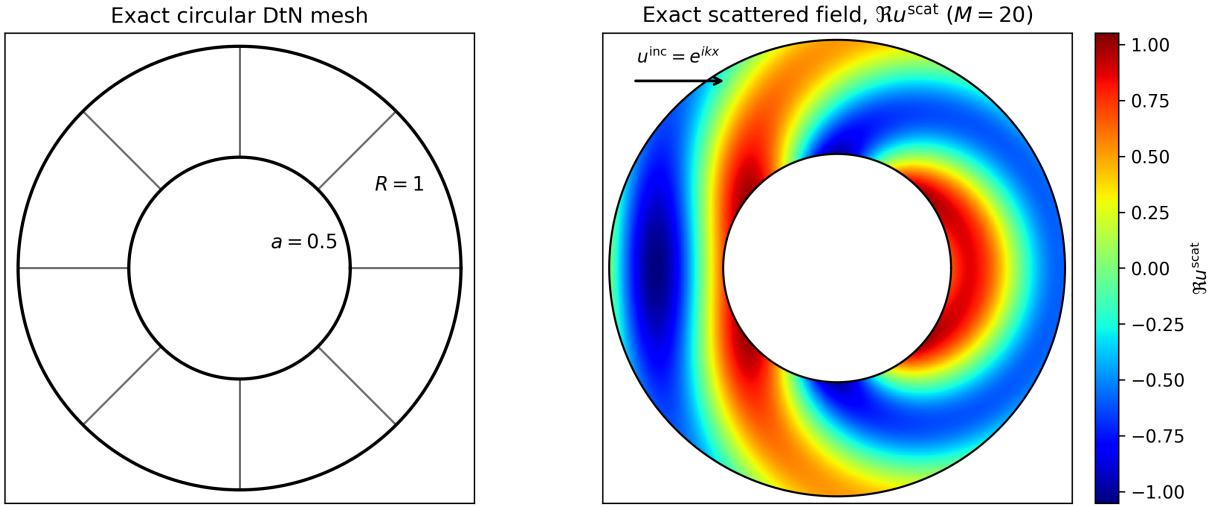


Fig. 3 DtN geometry and plane-wave scattering reference field at $k = 8$. Left: the exact annular mesh used throughout the DtN experiments, with one radial layer and eight exact curved sectors. Right: the real part of the exact scattered field for the incident plane wave $u^{\text{inc}} = e^{ikx}$, evaluated from (88) with $M = 20$ terms for display. The $M = 20$ truncation is used only for this plot and does not alter any of the numerical results reported below.

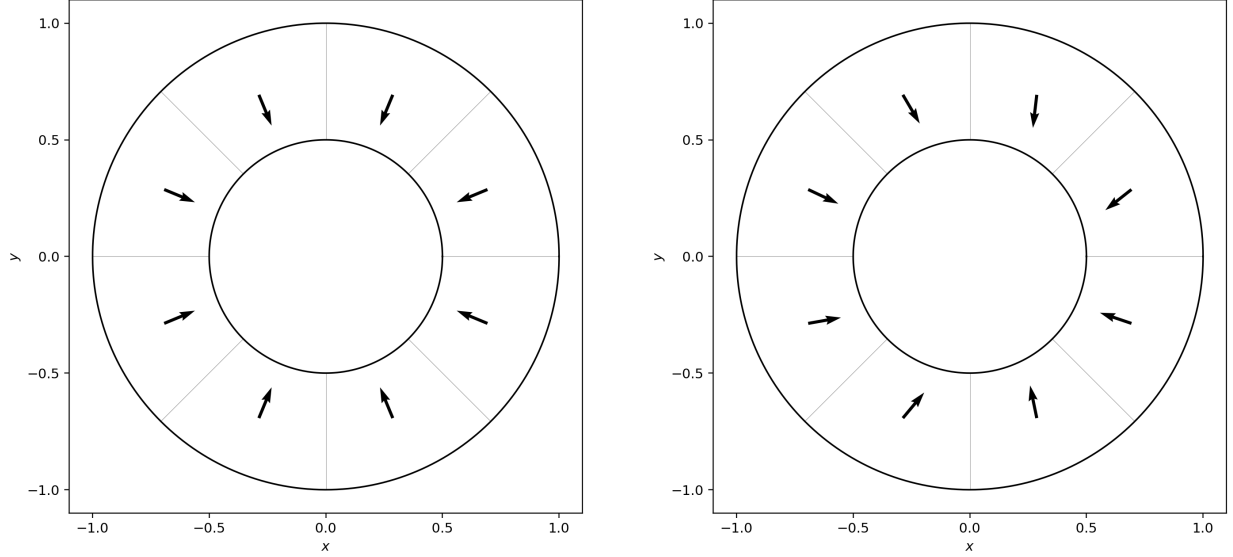
Table 4 DtN Hankel tests on the exact annulus. The angular errors compare the learned outer-ring fan center with the exact local wave-vector direction and are used only as a posteriori diagnostics.

case	uniform $p = 3 E_{L^2}$	adaptive E_{L^2}	mean angle error	max. angle error	κ_{GR}
centered source	7.653×10^{-1}	7.663×10^{-2}	$7 \times 10^{-6^\circ}$	$7 \times 10^{-6^\circ}$	1.307
off-center source	7.626×10^{-1}	9.407×10^{-2}	2.86°	4.87°	1.339

From Table 4 and Fig. 4 we observe that the direction recovery itself is good. For the centered source the fan centers are radially inward to the digits shown. For the off-center source they follow the source-centered phase rather than the normal to the artificial circle. Nevertheless, the relative L^2 error remains of order 10^{-1} . The conclusion is simple: recovering the dominant local phase direction is not enough for a cylindrical field. A three-wave fan does not represent the curvature of the wavefront or the radial amplitude variation. We next increase the local dimension and ask whether the remaining high- p floor is approximation error or numerical error.

6.5 Finite-precision, compression and trace-rank barriers for DtN–PWDG

Our next DtN experiment examines the high- p calculation. In [8] the relative error decreased rapidly with p and then stopped decreasing because of numerical instability. Here we investigate the source of this behavior. We use the same exact circular geometry, but we do not try to reproduce the earlier mesh point for point.



(a) Centered cylindrical wave.

(b) Off-center cylindrical wave.

Fig. 4 Learned plane-wave wave vectors on the exact circular DtN boundary. The arrows show q , equivalently the local spatial phase-gradient direction, not the physical energy-propagation direction. With the $e^{+i\omega t}$ convention and outgoing $H^{(2)}$ waves, q points toward the source while energy propagates outward. Thus, in the centered case the recovered arrows point radially inward to numerical accuracy; in the off-center case they follow the source-centered phase direction rather than the normal of the artificial circle.

Instead, we fix the present discretization and compare the raw coefficient system, the graph–Riesz system, different trace cutoffs, and different arithmetic precisions. In this way we can distinguish approximation error from numerical effects.

We now return to the sound-soft disk scattering problem shown in Fig. 3, with incident field $u^{\text{inc}} = e^{ikx}$ and $u^{\text{scat}} = -u^{\text{inc}}$ on $r = a$. We take the exact annulus $0.5 < r < 1$, $k = 8$, one radial layer and eight exact curved sectors. We fix the DtN truncation at $N = 40$ and use 44-point Gauss–Legendre quadrature on the circular and radial edges. The local directions are uniformly distributed and we use

$$p = 19, 23, 25, 27, 29, 31, 33, 35, 37, 39.$$

The same nominal spaces are solved first in the raw coefficient basis in IEEE double precision. We then apply the local Cauchy-trace normalization of Section 3.2 and solve in graph–Riesz coordinates, again in double precision. A companion calculation uses 80-bit extended arithmetic. For $p = 25, 27, 29, 31, 33$ we additionally assemble and factor the full system in IEEE binary128 arithmetic. No high-precision values are interpolated or extrapolated.

Results are shown in Fig. 5, with selected values in Table 5. Up to $p = 27$ the different calculations give essentially the same error. At $p = 29$ the raw double-precision error increases instead of decreasing. The graph–Riesz calculation does not show this large increase, which indicates that the first loss of accuracy is caused by the coefficient basis. Increasing the arithmetic precision gives a further improvement. For example, at $p = 33$ the binary128 error is 2.56×10^{-5} , compared with approximately 2.3×10^{-4} for graph–Riesz in double precision. Thus graph–Riesz normalization improves the double-precision calculation, but it does not by itself recover the smaller modes that are visible at higher precision.

Table 5 Selected points from the audited DtN precision continuation. The graph–Riesz column uses double precision and a near-machine trace cutoff.

p	raw double E_{L^2}	GR double E_{L^2}	80-bit E_{L^2}	binary128 E_{L^2}
25	4.506×10^{-4}	4.507×10^{-4}	4.507×10^{-4}	4.507×10^{-4}
27	2.417×10^{-4}	2.394×10^{-4}	2.394×10^{-4}	2.394×10^{-4}
29	1.191×10^{-3}	2.347×10^{-4}	1.065×10^{-4}	1.063×10^{-4}
31	3.340×10^{-4}	2.358×10^{-4}	8.302×10^{-5}	5.420×10^{-5}
33	3.087×10^{-4}	2.341×10^{-4}	9.274×10^{-5}	2.555×10^{-5}

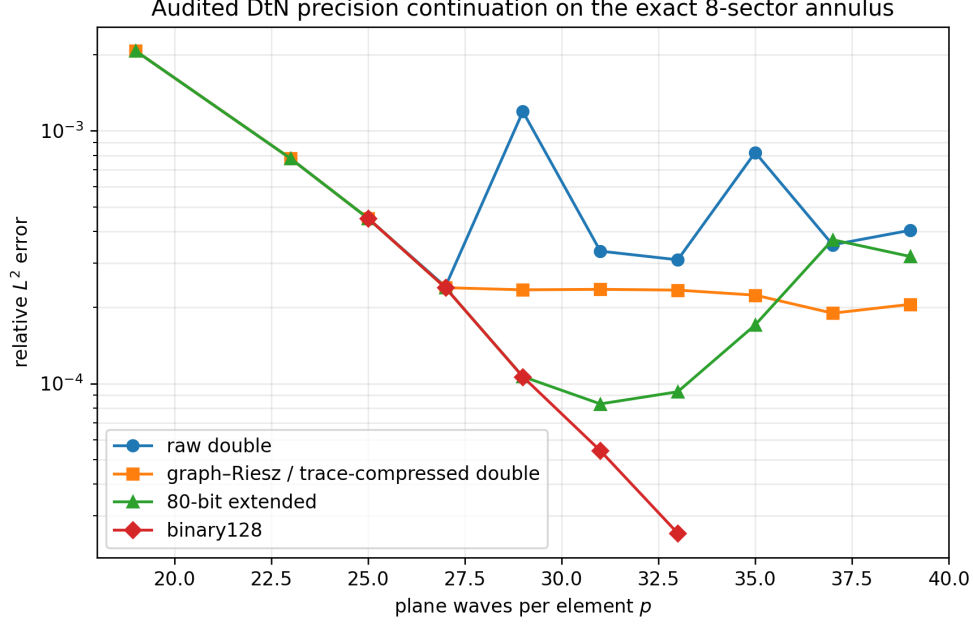


Fig. 5 Audited precision continuation on the exact eight-sector DtN annulus. Raw double, graph-Riesz double and 80-bit arithmetic contain ten computed points. Binary128 assembly and factorization were carried out only at $p = 25, 27, 29, 31, 33$. The raw double basis fails first; graph-Riesz normalization stabilizes the trace space still visible in double precision; higher arithmetic resolves additional modes and moves the p -floor.

We next investigate the effect of the trace cutoff. We keep the trace eigensolve in double precision and decrease τ_{rank} . Results for three values of p are given in Table 6. From the table we see that decreasing the cutoff continues to add useful modes. For $p = 35$, the relative error decreases from 4.27×10^{-4} when $\tau_{\text{rank}} = 10^{-14}$ to 6.66×10^{-5} when all trace modes having a computed positive eigenvalue are retained. Hence the choice 10^{-12} is too large for this high- p , high-accuracy calculation.

Table 6 Effect of decreasing the trace cutoff in the double-precision graph-Riesz calculation. “all +” means that every trace eigenvector with a computed positive eigenvalue is retained.

p	τ_{rank}	effective DOFs	E_{L^2}
29	10^{-12}	184	7.90×10^{-4}
29	10^{-15}	216	2.36×10^{-4}
29	10^{-17}	224	1.79×10^{-4}
29	all +	225	1.70×10^{-4}
31	10^{-15}	216	2.35×10^{-4}
31	10^{-16}	220	1.94×10^{-4}
31	10^{-17}	233	1.58×10^{-4}
31	all +	236	1.33×10^{-4}
35	10^{-14}	200	4.27×10^{-4}
35	10^{-16}	223	1.75×10^{-4}
35	10^{-17}	248	8.13×10^{-5}
35	all +	249	6.66×10^{-5}

Decreasing the cutoff in double precision eventually stops being sufficient because the smallest eigenvalues of the trace Gramian are no longer computed reliably. To check this directly, we recomputed the local trace Gramian and its eigendecomposition at 50 decimal digits for $p = 29, 31, 35$. In all three cases every eigenvalue is positive. The smallest relative eigenvalues are

$$5.01 \times 10^{-13}, \quad 1.00 \times 10^{-14}, \quad 2.53 \times 10^{-18}.$$

Consequently, at $p = 35$ the fixed value $\tau_{\text{rank}} = 10^{-12}$ retains only 27 of 35 local trace modes, 10^{-16} retains 33, and 10^{-20} retains all 35. Figure 6 shows this directly.

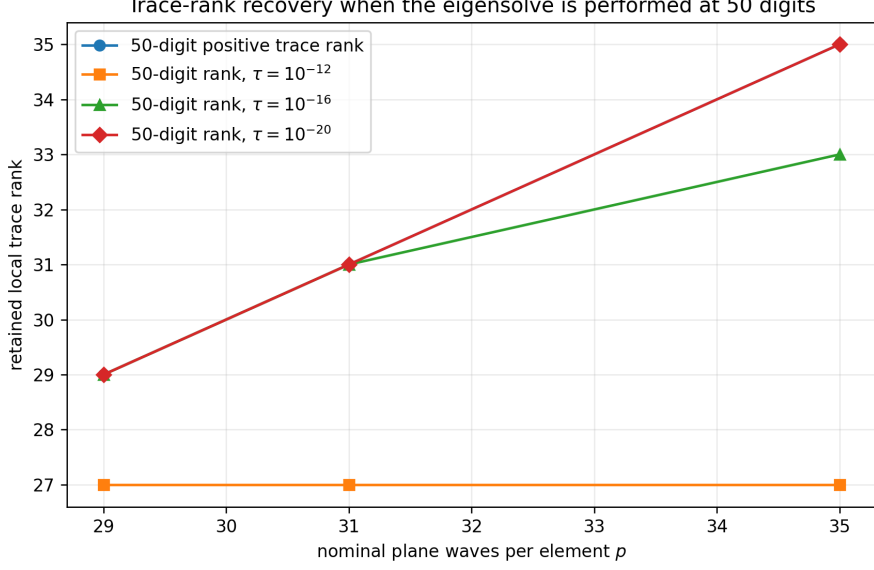


Fig. 6 Local trace rank when the Gramian and its eigendecomposition are computed at 50 decimal digits. The high-precision spectrum is positive in all three cases. A fixed cutoff 10^{-12} therefore creates an artificial high- p rank ceiling; decreasing the cutoff in proportion to the working precision restores the smaller but still resolvable trace modes.

From these experiments we draw two conclusions. First, τ_{rank} should not be regarded as a fixed universal value. It must be chosen with the working precision and the accuracy required from the calculation. Second, graph–Riesz normalization and trace truncation address different numerical problems. Graph–Riesz normalization improves the conditioning on the retained space, while the cutoff decides which trace modes are retained. Once the trace spectrum itself reaches the double-precision noise level, the eigensolve must also be carried out at higher precision. We do not pursue a full arbitrary-precision compressed global solve here.

6.6 Finite directional complexity and the limit of exact recovery

The Hankel calculation has broad directional content, so it does not isolate the ability of the nonlinear search to recover a finite set of rays. We now remove this ambiguity. We prescribe the exact solution to be a finite sum of plane waves and increase the number of directions one at a time. The exact field therefore remains in a finite Trefftz space for every value of M . Any loss of exact recovery can then be attributed to the direction search or the numerical linear algebra, not to curved-wavefront approximation. On $\Omega = (-1, 1)^2$ with $\kappa = 12$ we take

$$u_M(x) = \sum_{j=1}^M a_j \exp(ikd_j \cdot x), \quad d_j = (\cos \theta_j, \sin \theta_j), \quad (89)$$

with $|a_j| = 1$. The nested direction sequence is

$$(\theta_j)_{j=1}^{20} = (17, 123, 251, 68, 191, 315, 39, 158, 286, 340, \\ 92, 224, 5, 145, 273, 327, 54, 178, 235, 301)^\circ,$$

and the coefficient phases $a_j = e^{i\phi_j}$ are

$$(\phi_j)_{j=1}^{20} = (0, 0.37, -0.51, 0.81, -1.04, 1.33, -1.52, 0.22, 1.71, -0.93, \\ 0.58, -1.22, 1.48, -0.31, 0.96, -1.73, 0.11, 1.18, -0.77, 1.91).$$

The phases are fixed only to avoid an artificially symmetric sum. Exact Dirichlet data are imposed on the boundary of a uniform 8-triangle mesh.

For each M we use $p = M$ directions shared by all elements. Hence the adaptive and uniform calculations have the same $8M$ coefficients. The uniform directions are $11^\circ + 360^\circ j/M$. For the adaptive calculation we use continuation in M . After the $M - 1$ problem has converged, the new direction is chosen by testing a fixed angular dictionary with the skeleton residual, subject to a minimum separation from the directions already present. We then move all M directions by nonlinear least squares. When ten directions are active we refine the birth dictionary from 36 to 72 angles; this only refines the search grid and does not prescribe the number

of plane waves. The exact directions in (89) are not used by either ENRICH or MOVE. They are used only afterward to measure the angular error. All computations in this subsection are in IEEE double precision, including the $M = 20$ calculation and the control experiment below.

The result is shown in Table 7 and Fig. 7. For $M = 1, \dots, 19$ all directions are recovered to essentially machine precision and the relative L^2 error stays between 10^{-15} and 2×10^{-14} . The comparison with the uniform space is the important point: both calculations use exactly $8M$ coefficients. At $M = 10$ the uniform error is 0.830, at $M = 15$ it is 0.539, and at $M = 19$ it is still 2.09×10^{-2} . Thus the large gain is due to learning the correct directions, not to increasing the dimension of the approximation space.

Table 7 Finite-direction capacity experiment at $\kappa = 12$ on 8 triangles. Both methods use $p = M$ directions and $8M$ coefficients. The angle error is the maximum circular matching error between the recovered and exact direction sets.

M	adaptive E_{L^2}	uniform E_{L^2}	max. angle error	κ_{GR}
10	2.91×10^{-14}	8.30×10^{-1}	7.38×10^{-130}	—
15	4.28×10^{-15}	5.39×10^{-1}	1.97×10^{-130}	8.73
17	6.98×10^{-15}	1.16×10^{-1}	1.74×10^{-120}	16.08
18	2.16×10^{-14}	8.92×10^{-2}	2.56×10^{-120}	18.01
19	1.99×10^{-14}	2.09×10^{-2}	3.66×10^{-120}	32.40
20	1.62×10^{-3}	7.86×10^{-3}	28.99°	75.25

The final point, $M = 20$, behaves differently. This value is not an optimal dimension; it is simply the first value in this continuation for which the automatic birth step fails. The residual test chooses 117.5° for the new ray, whereas the twentieth exact direction is 301° . MOVE decreases the residual but remains in a false basin. The maximum angular error is then 28.99° and $E_{L^2} = 1.62 \times 10^{-3}$. The uniform calculation at the same dimension has error 7.86×10^{-3} . Thus the adaptive calculation is still better, but the large advantage seen for $M \leq 19$ is lost. We also checked whether this is a double-precision linear algebra effect. The final PWDG solve has relative algebraic residual 2.49×10^{-14} , with $\kappa_2(A) = 1.65 \times 10^7$ and $\kappa_{\text{GR}} = 75.25$. Since the field error is much larger than the algebraic residual, the loss of accuracy is caused by the nonlinear direction search rather than by the linear solve.

We next repeat only the twentieth step in order to check that the twenty-ray space can still represent the exact solution. The first nineteen recovered directions are kept and the new ray is initialized at 302.5° , which is 1.5° from the exact direction. This calculation is again carried out in double precision. After nine residual evaluations we obtain $E_{L^2} = 1.11 \times 10^{-14}$, $\mathcal{J} = 1.57 \times 10^{-24}$ and a maximum angular error of 3.20×10^{-11} degrees. Here $\kappa_2(A) = 1.60 \times 10^4$ and $\kappa_{\text{GR}} = 35.54$. This control shows that the twenty-ray Trefftz space has not reached an approximation or precision barrier. The failure of the automatic calculation is caused by the birth decision.

From Fig. 7 we see a clear change of behavior. When the solution contains a finite set of identifiable directions, the adapted space can be much more accurate than a uniform angular space of the same dimension. As more directions are added, the uniform space becomes a better generic approximation while the nonlinear birth problem becomes more difficult. We do not attach special significance to the number nineteen. It is the last exact-recovery point for this nested family and this continuation strategy. The main observation is that the largest gain is obtained when the directional content is sparse enough to be identified.

6.7 Trianglewise direction and h -adaptivity

Our final adaptive experiment combines direction changes with mesh refinement on the L-shaped corner-singular and crossing-wave example. We compare the number of Trefftz coefficients required by the combined method with residual based h -refinement alone. We measure the error by the relative graph norm

$$E_{\mathcal{G}} = \frac{\|u - u_h\|_{\mathcal{G},h}}{\|u\|_{\mathcal{G},h}}.$$

Both methods start from the same 24-triangle mesh with five equally spaced plane waves per element. The standard loop and its extension here are

$$\begin{aligned} & \text{SOLVE} \rightarrow \text{ESTIMATE} \rightarrow \text{MARK} \rightarrow \text{REFINE}, \\ & \text{SOLVE} \rightarrow \text{ESTIMATE} \rightarrow \text{MARK} \rightarrow \text{TEST} \rightarrow \text{ACT} \rightarrow \text{SOLVE}. \end{aligned}$$

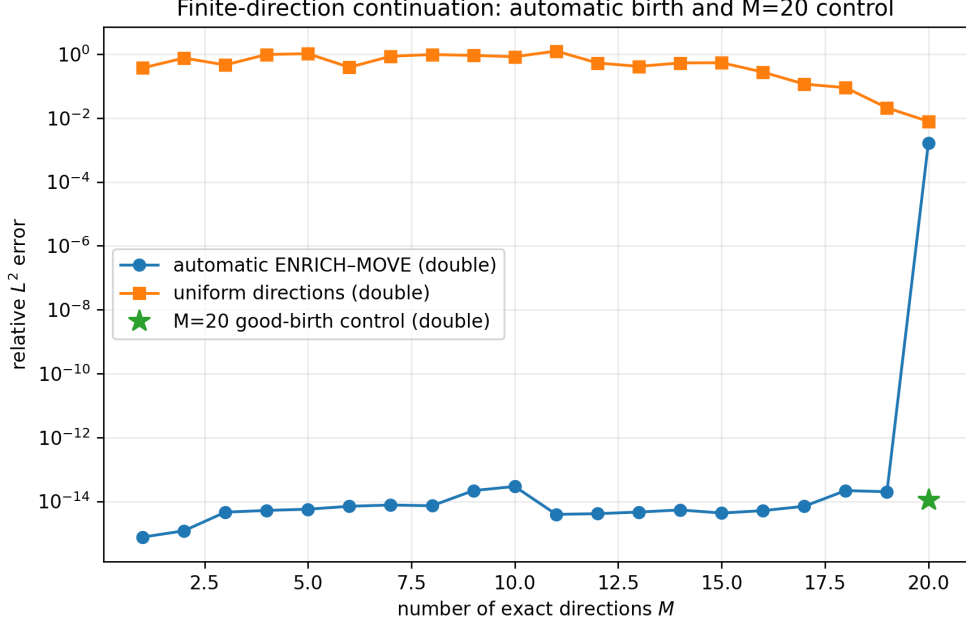


Fig. 7 Continuation in the number of exact directions. Every point in this figure is computed in IEEE double precision. The automatic ENRICH-MOVE path remains at roundoff through $M = 19$ and enters a false basin at $M = 20$. The isolated star at $M = 20$ is the double-precision control in which the twentieth ray is born at 302.5° ; the same twenty-ray space then returns to roundoff. This separates failure of the automatic birth decision from a finite-precision or representability barrier.

In the ESTIMATE step we compute the element quantities η_K^2 from (40). Each interior edge contribution is divided equally between the two adjacent triangles, while a boundary contribution is assigned to the adjacent triangle. This step uses only the current PWDG solution. No trial direction or trial mesh problem is solved during ESTIMATE.

For the present all-Dirichlet problem these indicators have an exact interpretation. Proposition 4.1 gives

$$\sum_K \eta_K^2 = \mathcal{J}(u_h) = \|u - u_h\|_{\mathcal{G},h}^2,$$

so their sum is the squared graph error of the current PWDG solution. For a mixed Dirichlet-impedance problem we would instead use the reliability estimate from Section 4.2. The role of ESTIMATE is therefore the same as in a standard adaptive method: it identifies the elements carrying the largest residual contribution. Only after MARK has selected these elements do we use TEST to decide whether to move directions, add directions, or refine the mesh.

For consistency with the finite-direction study, we use the common nominal cap $p_K \leq 20$ during the direction-learning stage. After cycle 14 the local fans are frozen and the remaining iterations use residual based h -refinement only. The pure h calculation starts from the same mesh and uses the same Dörfler parameter $\theta = 0.4$ and the same refinement rule, but keeps five fixed equally spaced directions on every element.

In the first stage, MARK applies Dörfler marking [23] to the values η_K^2 . For each marked element, TEST considers two MOVE trials, obtained by rotating the current fan by $\pm 15^\circ$, and at most six ENRICH trials. Candidate directions are required to be separated by at least 7° . If J_K^{move} and J_K^{enrich} are the best trial residuals and η_K^2 the current local contribution, the action scores are

$$G_K^{\text{move}} = \frac{J - J_K^{\text{move}}}{\eta_K^2}, \quad G_K^{\text{enrich}} = \frac{J - J_K^{\text{enrich}}}{\eta_K^2}. \quad (90)$$

ACT chooses the stronger admissible directional action when

$$G_K^{\text{move}} \geq 0.20 \quad \text{or} \quad G_K^{\text{enrich}} \geq 0.40,$$

and otherwise chooses REFINE. Children inherit the parent direction set.

The main stages of the calculation are listed in Table 8. At the end of the direction-learning stage the relative graph error has decreased from 0.8471 to 0.1278 using 884 coefficients. The mean local dimension is 6.23 and the largest local dimension is 8, so the common cap $p_K \leq 20$ is inactive and does not affect this

trajectory. With the learned fans then held fixed, the hybrid calculation first satisfies $E_G < 0.05$ with 3329 coefficients. The pure h calculation requires 11290 coefficients to reach the same error level.

Table 8 Threshold-driven hybrid benchmark on the L-shaped domain.

state	cycle	triangles	coefficients	E_G
initial mesh	0	24	120	0.8471
hybrid after direction stage	14	142	884	0.1278
hybrid at threshold	19	538	3329	0.0404
pure h at threshold	16	2258	11290	0.0462

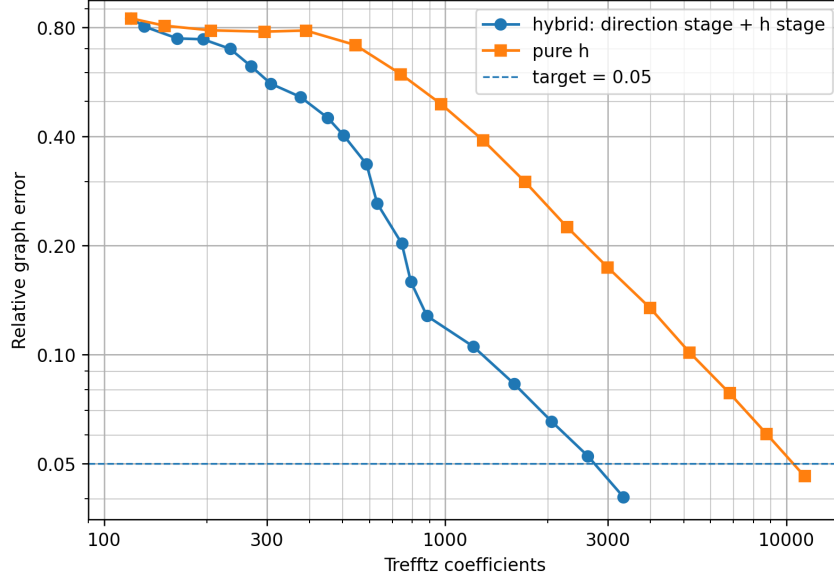


Fig. 8 Relative graph error versus Trefftz coefficients for the threshold-driven hybrid benchmark. The hybrid method uses a direction-learning stage followed by pure h -refinement with the learned local fans frozen.

The complete histories are shown in Fig. 8. At the first point below 5% relative graph error, the hybrid calculation uses about 70.5% fewer coefficients than pure h -refinement. During the first stage, MOVE and ENRICH are selected where changing the local directional space gives the larger reduction in the residual, while the mesh is still refined near the singular corner. The learned directions are then frozen. From the last part of the figure we see that the reduction in coefficients is retained as the mesh is further refined.

TEST-stage work.

The present TEST implementation is not cheaper than pure h -adaptivity in elapsed time. Every marked element may require two MOVE trials and as many as six ENRICH trials, and each accepted trial is evaluated through a global solve. Table 8 and Fig. 8 should therefore be interpreted as a comparison of approximation spaces, not as a timing comparison.

6.8 Variable-projection check

Part B is intended as an alternative nonlinear formulation rather than as a different approximation space. For the exact finite plane-wave recovery problems, Parts A and B therefore give the same field once the correct directions have been found. We use the Hankel field only as a nonzero-residual check of the variable-projection implementation. At $\kappa = 8$ on the eight-element mesh, one real direction per element initialized from the centroid-radial field gives $\mathcal{J} = 4.5604 \times 10^{-1}$ and $E_{L^2} = 3.5462 \times 10^{-1}$; a blind Sobol initialization converges to a poorer basin with $\mathcal{J} = 1.5792$ and $E_{L^2} = 9.8983 \times 10^{-1}$. With four local rays the same variable-projection calculation gives $\mathcal{J} = 2.5410 \times 10^{-2}$ and $E_{L^2} = 6.9700 \times 10^{-2}$. These checks confirm the implementation and the same nonconvexity already seen in Part A; they are not used as evidence for a separate approximation advantage.

7 Conclusion

In this paper we have considered PWDG approximations in which the propagation directions are determined during the computation. In Part A the PWDG equations are retained and the skeleton residual is minimized over the directions. In Part B the Galerkin constraint is removed and the linear coefficients are eliminated by variable projection. Complex angles allow the same local family to contain propagating and evanescent waves.

The numerical experiments show that the usefulness of direction adaptation depends strongly on the directional content of the solution. For the finite plane-wave sums, the result is particularly clear. In the nested family considered here, ENRICH–MOVE continuation recovers all directions to roundoff for $M = 1, \dots, 19$, while a uniform angular space with the same number of coefficients remains much less accurate. At $M = 20$ the automatic birth step enters a false basin. Repeating the same step with a nearby initial direction, still in double precision, again gives roundoff error. Thus this failure is caused by the nonlinear identification of the new direction and not by the twenty-ray approximation space. We do not interpret nineteen as a universal limit; the experiment is intended to show the regime in which direction adaptation gives its largest gain.

The Hankel calculation gives the complementary result. On the exact circular DtN boundary the learned fan follows the local phase direction accurately, including for an off-center source, but the field error remains much larger. A small number of plane waves cannot exactly represent the curvature and amplitude variation of a cylindrical field. Hence learning the correct local phase direction is not by itself sufficient when the directional content is broad.

We also revisited the high- p numerical instability observed in the earlier DtN–PWDG calculations. The experiments separate three effects. The raw coefficient basis loses accuracy first. Graph–Riesz normalization improves the calculation on the retained trace space. Decreasing τ_{rank} then restores additional trace modes. Finally, when the smallest trace eigenvalues are computed at 50 decimal digits, the modes that appeared nonpositive in double precision are positive. This shows that τ_{rank} should be regarded as a numerical parameter depending on the working precision and required accuracy, rather than as a fixed constant.

For straight edges, the products of the local plane waves and their direction derivatives reduce to exact exponential moments. In the all-Dirichlet case we also have

$$\mathcal{J}(v_h) = \|u - v_h\|_{\mathcal{G},h}^2,$$

so that the residual used to move the directions is the same quantity used to measure the graph error. For mixed Dirichlet–impedance problems we obtained the corresponding L^2 residual bound under the broken Trefftz stability assumption of Section 4.2.

The transmission experiment shows that the complex-angle formulation recovers both the propagating Snell direction and an evanescent transmitted wave. In the combined adaptive experiment, the usual SOLVE–ESTIMATE–MARK loop is supplemented by a TEST step which chooses between MOVE, ENRICH and REFINE on marked elements. The relative graph error falls below 5% with 3329 Trefftz coefficients, compared with 11290 coefficients for pure h -refinement. This is a comparison of approximation spaces; the present TEST step uses several global trial solves and is not yet a cheaper algorithm in elapsed time.

The nonlinear analysis given here is local. On a fixed-rank neighborhood of an identifiable zero-residual solution we obtain quadratic growth of the reduced objective and stability of the field with respect to the direction parameters. This agrees with the numerical behavior: once the iteration is in the correct basin the directions converge rapidly, while a poor birth or a direct solve at the final frequency may converge to another minimum. An important next step is therefore to replace the present global trial solves by cheaper local tests and to improve the continuation and birth strategies when several directional components are present.

References

- [1] O. Cessenat and B. Després. Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem. *SIAM J. Numer. Anal.*, 35(1):255–299, 1998.
- [2] P. Monk and D.-Q. Wang. A least-squares method for the Helmholtz equation. *Comput. Methods Appl. Mech. Engrg.*, 175:121–136, 1999.
- [3] G. H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal.*, 10(2):413–432, 1973.
- [4] C. J. Gittelsohn, R. Hiptmair, and I. Perugia. Plane wave discontinuous Galerkin methods: analysis of the h -version. *ESAIM Math. Model. Numer. Anal.*, 43(2):297–331, 2009.

- [5] R. Hiptmair, A. Moiola, and I. Perugia. Plane wave discontinuous Galerkin methods for the 2D Helmholtz equation: analysis of the p -version. *SIAM J. Numer. Anal.*, 49(1):264–284, 2011.
- [6] R. Hiptmair, A. Moiola, and I. Perugia. Plane wave discontinuous Galerkin methods: exponential convergence of the hp -version. *Found. Comput. Math.*, 16(3):637–675, 2016.
- [7] R. Hiptmair, A. Moiola, and I. Perugia. A survey of Trefftz methods for the Helmholtz equation. In *Building Bridges: Connections and Challenges in Modern Approaches to Numerical Partial Differential Equations*, Lecture Notes in Computational Science and Engineering 114, pp. 237–279. Springer, Cham, 2016.
- [8] S. Kapita and P. Monk. A plane wave discontinuous Galerkin method with a Dirichlet-to-Neumann boundary condition for the scattering problem in acoustics. *J. Comput. Appl. Math.*, 327:208–225, 2018.
- [9] S. Kapita, P. Monk, and T. Warburton. Residual-based adaptivity and PWDG methods for the Helmholtz equation. *SIAM J. Sci. Comput.*, 37(3):A1525–A1553, 2015.
- [10] S. Congreve, P. Houston, and I. Perugia. Adaptive refinement for hp -version Trefftz discontinuous Galerkin methods for the homogeneous Helmholtz problem. *Adv. Comput. Math.*, 45(1):361–393, 2019.
- [11] S. Congreve, J. Gedicke, and I. Perugia. Numerical investigation of the conditioning for plane wave discontinuous Galerkin methods. In F. A. Radu, K. Kumar, I. Berre, J. M. Nordbotten, and I. S. Pop (eds.), *Numerical Mathematics and Advanced Applications – ENUMATH 2017*, Lecture Notes in Computational Science and Engineering 126, pp. 493–500. Springer, Cham, 2019.
- [12] H. Barucq, A. Bendali, J. Diaz, and S. Tordeux. Local strategies for improving the conditioning of the plane-wave ultra-weak variational formulation. *J. Comput. Phys.*, 441:110449, 2021.
- [13] A. Agrawal and R. H. W. Hoppe. Optimization of plane wave directions in plane wave discontinuous Galerkin methods for the Helmholtz equation. *Port. Math.*, 74(1):69–89, 2017.
- [14] M. Amara, S. Chaudhry, J. Diaz, R. Djellouli, and S. L. Fiedler. A local wave tracking strategy for efficiently solving mid- and high-frequency Helmholtz problems. *Comput. Methods Appl. Mech. Engrg.*, 276:473–508, 2014.
- [15] J. Fang, J. Qian, L. Zepeda-Núñez, and H. Zhao. Learning dominant wave directions for plane wave methods for high-frequency Helmholtz equations. *Res. Math. Sci.*, 4:9, 2017.
- [16] C. Y. Lam and J. Qian. Numerical microlocal analysis by fast Gaussian wave packet transforms and application to high-frequency Helmholtz problems. *SIAM J. Sci. Comput.*, 41(5):A2717–A2746, 2019.
- [17] Q. Hu and Z. Wang. A plane wave method based on approximate wave directions for two dimensional Helmholtz equations with large wave numbers. arXiv:2107.09797, 2021.
- [18] E. Parolin, D. Huybrechs, and A. Moiola. Stable approximation of Helmholtz solutions in the disk by evanescent plane waves. *ESAIM Math. Model. Numer. Anal.*, 57(6):3499–3536, 2023.
- [19] T. Betcke and L. N. Trefethen. Reviving the method of particular solutions. *SIAM Rev.*, 47(3):469–491, 2005.
- [20] A. H. Barnett and T. Betcke. Stability and convergence of the method of fundamental solutions for Helmholtz problems on analytic domains. *J. Comput. Phys.*, 227(14):7003–7026, 2008.
- [21] A. Moiola and E. A. Spence. Acoustic transmission problems: wavenumber-explicit bounds and resonance-free regions. *Math. Models Methods Appl. Sci.*, 29(2):317–354, 2019.
- [22] D. Baskin, E. A. Spence, and J. Wunsch. Sharp high-frequency estimates for the Helmholtz equation and applications to boundary integral equations. *SIAM J. Math. Anal.*, 48(1):229–267, 2016.
- [23] W. Dörfler. A convergent adaptive algorithm for Poisson’s equation. *SIAM J. Numer. Anal.*, 33(3):1106–1124, 1996.

- [24] A. Hjørungnes and D. Gesbert. Complex-valued matrix differentiation: techniques and key results. *IEEE Trans. Signal Process.*, 55(6):2740–2746, 2007.
- [25] J. J. Moré. The Levenberg–Marquardt algorithm: implementation and theory. In G. A. Watson (ed.), *Numerical Analysis*, Lecture Notes in Mathematics 630, pp. 105–116. Springer, Berlin, 1978.
- [26] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.*, 49(6):409–436, 1952.
- [27] Y. Saad and M. H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7(3):856–869, 1986.
- [28] H. A. van der Vorst. Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 13(2):631–644, 1992.
- [29] I. M. Sobol’. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput. Math. Math. Phys.*, 7(4):86–112, 1967.
- [30] L. T. Watson. Globally convergent homotopy methods: a tutorial. *Appl. Math. Comput.*, 31:369–396, 1989.
- [31] K.-A. Mardal and R. Winther. Preconditioning discretizations of systems of partial differential equations. *Numer. Linear Algebra Appl.*, 18(1):1–40, 2011.

Declarations

Competing interests. The author declares no competing interests.

Data and code availability. The numerical experiments are generated by the accompanying research code and tabulated data supplied with the manuscript. The computational results reported in the paper are not extrapolated from fitted curves.

Author contributions. S. Kapita developed the methodology, analysis, implementation, numerical experiments, and manuscript.