# BORN-SERIES-INSPIRED RESIDUAL METRIC FOR LEARNED PRECONDITIONERS*

**Juntao Wang**[a,c], **Xinliang Liu**[b], **Jiwei Jia**[a,c†]
[a] School of Mathematics, Jilin University, Changchun 130012, China
[b] Ocean University of China, Qingdao 266100, China
[c] Shenzhen Loop Area Institute, Shenzhen 518038, China
`jiajiwei@jlu.edu.cn`

## ABSTRACT

Learned PDE preconditioners typically train by minimizing an unpreconditioned residual norm, yet iterate in preconditioned coordinates at inference. For indefinite operators such as the high-frequency Helmholtz equation, this mismatch between training and inference leaves near-resonant error modes poorly controlled. We close this gap by recasting the training objective as a *preconditioned residual metric*: the standard Born-series / shifted-Laplacian relation $\mathbf{I} - G_\eta V_\eta = G_\eta A = L_\eta^{-1} A$ shows that measuring the residual in the $G_\eta^* G_\eta$-weighted inner product (a Riesz-map metric) inherits the conditioning benefits of classical shifted-Laplacian preconditioning, so that training and inference operate in the same geometry. Building on this, we propose (i) a *neural preconditioned Born-series* solver that replaces the CBS scalar correction with a learned operator, and (ii) a *Born-series-inspired loss* that trains in the matched preconditioned metric. The framework is architecture-agnostic: any learned preconditioner that is linear in the residual can be plugged in. We validate on three PDE classes—high-frequency Helmholtz, convection–diffusion–reaction, and linearized Newton systems from nonlinear PDEs—with controlled experiments that separate the solver-level benefit of replacing CBS with a learned operator (up to $68.6\times$ fewer iterations on OpenFWI) from the pure loss-design benefit of switching to the preconditioned metric ($\sim 28\%$ fewer iterations on high-contrast media at fixed architecture).

## 1 INTRODUCTION

The heterogeneous Helmholtz equation is notoriously difficult for both classical and learning-based solvers in high-wavenumber and high-contrast regimes. Standard discretizations exhibit pollution errors Babuška & Sauter (1997); multilevel methods suffer from coarse-grid dispersion mismatch Stolk (2016); and indefiniteness degrades relaxation and coarse correction Ernst & Gander (2011). Robust classical remedies—shifted-Laplacian preconditioning Erlangga et al. (2004), wave-ray multigrid Brandt & Livshits (1997); Livshits & Brandt (2006), CBS Osnabrugge et al. (2016), and domain decomposition Chen & Xiang (2013); Leng & Ju (2022)—have made progress, but efficiency at scale in strongly heterogeneous media remains an open challenge.

Learning-based PDE solvers have progressed from PINNs Raissi et al. (2019) to neural operators Azizzadenesheli et al. (2024); Lu et al. (2021); Li et al. (2020; 2022); Rahman et al. (2022); He et al. (2024); Cao (2021); He et al. (2025); Cao et al. (2025). For Helmholtz, the most promising recent direction embeds learned modules as iterative preconditioners rather than end-to-end predictors Lerer et al. (2024); Rudikov et al. (2024); Zhang et al. (2024); Cui et al. (2025); Xie et al. (2025); Stanziola et al. (2021). In this paradigm, a learned operator $\mathcal{M}_\theta$ produces a correction from the residual at each iteration, $u^{m+1} = u^m + \mathcal{M}_\theta(f - Au^m)$, progressively reducing the physical residual. The standard training objective minimizes the unpreconditioned Euclidean residual $\|A\mathcal{M}_\theta(r) - r\|_2$.

---

[*] This paper was accepted to the AI&PDE Workshop at ICLR 2026.

[†] Corresponding author.

This loss is simple and widely used, but it exposes optimization directly to the hardest spectrum of $A$ and does not match the preconditioned coordinates used at inference.

**The training–inference mismatch.** Despite the appeal of this paradigm, a subtle but consequential gap persists between how learned preconditioners are *trained* and how they are *used*. At inference, preconditioned iterations naturally operate in transformed residual coordinates—for example, letting $L_\eta := -\Delta - (k_0^2 + i\eta)$ be a damped reference operator and $G_\eta := L_\eta^{-1}$ its inverse, CBS corrects in the Born/integral-residual coordinates $G_\eta(f - Au^m)$, and shifted-Laplacian methods precondition with $L_\eta^{-1}$. Yet training almost universally minimizes the unpreconditioned Euclidean residual $\|A\mathcal{M}_\theta(r) - r\|_2$. For well-conditioned (e.g., SPD) problems this mismatch is benign, but for indefinite operators like high-frequency Helmholtz, the Euclidean residual landscape is stiff with respect to near-resonant modes Liu et al. (2024); Rahaman et al. (2019), making optimization unnecessarily difficult.

**Key observation.** We resolve this mismatch through a single algebraic identity. Writing $V_\eta := L_\eta - A$ for the perturbation (so $A = L_\eta - V_\eta$; formal definitions in Section 2), the three frameworks are linked by

$$\mathbf{I} - G_\eta V_\eta \;=\; G_\eta A \;=\; L_\eta^{-1} A.$$

This identity reveals that the Born/integral residual $(\mathbf{I} - G_\eta V_\eta)u - G_\eta f$ is exactly the shifted-Laplacian left-preconditioned residual $L_\eta^{-1}(Au - f)$. Though the algebraic relation is implicit in CBS theory Osnabrugge et al. (2016), its consequences for *training objective design* have not been exploited. We show that it directly yields a preconditioned residual metric—equivalently, a Riesz-map loss in the $G_\eta^* G_\eta$-inner product—that aligns the training geometry with the inference geometry by the same conditioning mechanism that makes classical shifted-Laplacian preconditioning effective.

We instantiate this framework with two architectures, MGNO He et al. (2025; 2024) and PA-FNO, and validate across three PDE classes: Helmholtz, convection–diffusion–reaction, and linearized Newton systems from nonlinear PDEs. Our experimental design cleanly decomposes two orthogonal effects: the *solver-level* benefit of replacing CBS's scalar correction with a learned operator, and the *loss-level* benefit of switching from $\mathcal{L}_{\mathrm{dir}}$ to $\mathcal{L}_{\mathrm{bs}}$ at fixed architecture.

**Contributions.**

1. **Preconditioned-metric training objective:** We derive a Born-series-inspired loss that is equivalently a Riesz-map residual norm in the shifted-preconditioner inner product, ensuring that training optimizes in the same geometry used at inference. We provide consistent formulas for Helmholtz, convection–diffusion–reaction (CDR), and linearized Newton systems from nonlinear PDEs, all with $\mathcal{O}(N \log N)$ evaluation via FFT/DST/DCT.
2. **Neural preconditioned Born-series solver:** We formulate a learned correction iteration in Born/preconditioned residual coordinates, unifying CBS, shifted-Laplacian, and learned-operator viewpoints under the identity $\mathbf{I} - G_\eta V_\eta = G_\eta A$.
3. **Controlled empirical decomposition:** We separately quantify the architecture effect (MgNO-BS vs CBS: up to $68.6\times$ iteration reduction on OpenFWI) and the loss-design effect ($\mathcal{L}_{\mathrm{bs}}$ vs $\mathcal{L}_{\mathrm{dir}}$: $\sim 28\%$ improvement on high-contrast media), with extensions to CDR and nonlinear Newton solvers.

The paper is organized as follows: Section 2 derives the Born/integral equation and establishes the key identity linking it to shifted-Laplacian left-preconditioning; Section 3 introduces the neural preconditioned Born-series solver and the Riesz-map loss; Section 4 presents the experiments.

## 2 FROM BORN SERIES TO PRECONDITIONED COORDINATES

We begin from the classical Born/Lippmann–Schwinger integral formulation and analyze its algebraic relation to shifted-Laplacian left-preconditioning. The key result (Proposition 1) shows that the integral residual and the shifted-Laplacian left-preconditioned residual are the *same* object. This identity motivates both the neural preconditioned Born-series (NPBS) solver (Section 3) and the preconditioned training metric (Section 3.1).

### 2.1 OPERATOR SPLITTING AND INTEGRAL EQUATION

Let

$$Au = f \tag{1}$$

be the discrete heterogeneous Helmholtz system (with absorbing treatment included in $A$). In our Helmholtz experiments, $A$ denotes the physical Helmholtz operator discretized either by finite differences or by FFT-based pseudospectral methods with absorbing layers on a bounded box; in both cases, A is *not* assumed to be translation invariant.

Introduce a damped reference operator

$$L_\eta := -\Delta - (k_0^2 + i\eta), \qquad \eta > 0, \tag{2}$$

used only for preconditioning/training coordinates. The reference is chosen to be transform-diagonalizable (periodic FFT or Dirichlet/Neumann DST/DCT) for fast application; any mismatch between this reference and the physical operator, including absorbing-layer terms, is absorbed into the perturbation $V_\eta := L_\eta - A$.

Let $g_\eta$ be the Green kernel for the *continuous* reference operator $L_\eta$ under the chosen boundary condition on $\Omega$. Define the associated continuous Green operator

$$(G_\eta q)(x) := \int_\Omega g_\eta(x, y)\, q(y)\, \mathrm{d}y, \qquad q \in L^2(\Omega), \tag{3}$$

so $G_\eta$ is the integral operator induced by $g_\eta$ and satisfies $G_\eta = L_\eta^{-1}$ on the admissible function class. In the discrete setting, we use the corresponding matrix (or FFT/DST/DCT operator) $G_\eta$ that approximates $L_\eta^{-1}$. For translation-invariant references (free space or periodic box), $g_\eta(x, y) = g_\eta(x - y)$ and equation 3 reduces to a convolution.

Recall the perturbation operator $V_\eta := L_\eta - A$, so the heterogeneous Helmholtz operator splits as

$$A = L_\eta - V_\eta. \tag{4}$$

For the coefficient-only Helmholtz split, $V_\eta$ reduces to multiplication by $k(x)^2 - k_0^2 - i\eta$. Starting from $Au = f$, we obtain

$$L_\eta u = V_\eta u + f. \tag{5}$$

Applying $G_\eta$ to both sides and using $G_\eta L_\eta = \mathbf{I}$ yields the preconditioned integral equation

$$u = G_\eta(V_\eta u + f), \tag{6}$$

or equivalently

$$(\mathbf{I} - G_\eta V_\eta)u = G_\eta f. \tag{7}$$

Equation 7 is the preconditioned operator equation used by CBS Osnabrugge et al. (2016). In translation-invariant settings it is exactly the standard Lippmann–Schwinger form. Solving equation 7 by fixed-point iteration amounts to a Neumann series in $G_\eta V_\eta$: convergence is guaranteed whenever $\|G_\eta V_\eta\| < 1$, and the rate and conditioning of the preconditioned system are controlled by this norm. This motivates the spectral bounds formalized below.

### 2.2 KEY IDENTITY: INTEGRAL RESIDUAL = SHIFTED-LAPLACIAN LEFT-PRECONDITIONED RESIDUAL

**Proposition 1** (Shifted-preconditioned equivalence and spectral motivation)**.** *For the decomposition in equation 4, define $G_\eta = L_\eta^{-1}$. Then:*

$$\mathbf{I} - G_\eta V_\eta = G_\eta A = L_\eta^{-1} A. \tag{8}$$

*Hence equation 7 is algebraically identical to the left-preconditioned system*

$$L_\eta^{-1} A\, u = L_\eta^{-1} f. \tag{9}$$

*Moreover, if $\|G_\eta V_\eta\|_2 \le \rho < 1$, then*

$$\sigma(\mathbf{I} - G_\eta V_\eta) \subset \{z \in \mathbb{C} : |z - 1| \le \rho\}, \tag{10}$$

$$(\mathbf{I} - G_\eta V_\eta)^{-1} = \sum_{j=0}^{\infty} (G_\eta V_\eta)^j, \tag{11}$$

*and*

$$\kappa_2(\mathbf{I} - G_\eta V_\eta) \le \frac{1+\rho}{1-\rho}. \tag{12}$$

*Proof.* Using $A = L_\eta - V_\eta$ and $G_\eta L_\eta = \mathbf{I}$:

$$\mathbf{I} - G_\eta V_\eta = G_\eta L_\eta - G_\eta V_\eta = G_\eta (L_\eta - V_\eta) = G_\eta A.$$

For the spectral part, $\rho(G_\eta V_\eta) \le \|G_\eta V_\eta\|_2 \le \rho$ implies each eigenvalue $\mu$ of $G_\eta V_\eta$ satisfies $|\mu| \le \rho$, so eigenvalues of $\mathbf{I} - G_\eta V_\eta$ are $1 - \mu$, giving equation 10. Since $\|G_\eta V_\eta\|_2 < 1$, the Neumann series equation 11 converges. Also

$$\|\mathbf{I} - G_\eta V_\eta\|_2 \le 1 + \rho, \qquad \|(\mathbf{I} - G_\eta V_\eta)^{-1}\|_2 \le \sum_{j \ge 0} \rho^j = \frac{1}{1-\rho},$$

which yields equation 12. □

### 2.3 CBS AS RICHARDSON ITERATION IN PRECONDITIONED COORDINATES

The CBS fixed-point iteration Osnabrugge et al. (2016) is

$$u^{m+1} = u^m + \gamma (G_\eta (V_\eta u^m + f) - u^m)\,, \tag{13}$$

where $\gamma$ is a scalar/diagonal relaxation. By equation 8:

$$u^{m+1} = u^m + \gamma\, G_\eta (f - A u^m), \tag{14}$$

which is Richardson iteration preconditioned by the shifted Laplacian $L_\eta$. Thus CBS, usually derived from integral-equation theory, is equivalently the simplest left-preconditioned stationary method in the differential-operator framework. The scalar $\gamma$ limits the correction to a single-parameter family; replacing it with a richer correction map is the subject of the next section.

**Role of $\eta$ in metric and iteration coordinates.** The shift $\eta$ defines both the left preconditioner $G_\eta = L_\eta^{-1}$ and the induced residual metric $R_\eta := G_\eta^* G_\eta$. Hence $\eta$ controls the geometry in which residuals are measured and corrected, not only numerical damping. In practice, we tune $\eta$ by a small validation sweep and then keep it fixed between training and inference for each dataset to preserve objective–inference consistency. Larger $\eta$ typically reduces $\|G_\eta V_\eta\|$, while excessively large $\eta$ can over-damp useful corrections.

## 3 NEURAL PRECONDITIONED BORN-SERIES METHOD

Proposition 1 reveals that CBS iterates in well-conditioned shifted-Laplacian coordinates but is limited by its scalar correction $\gamma$. We now replace $\gamma$ with a learned operator $\mathcal{M}_\theta$ that acts in the *same* preconditioned coordinates, yielding the neural preconditioned Born-series (NPBS) iteration.

At iterate $u^m$, the NPBS update is

$$u^{m+1} = u^m + \mathcal{M}_\theta (G_\eta (V_\eta u^m + f) - u^m)\,. \tag{15}$$

Here $V_\eta u^m$ is a pointwise multiplication and $G_\eta$ is applied once via FFT/IFFT, so each iteration requires only *one* FFT pair for the preconditioned residual. If $\mathcal{M}_\theta = \gamma \mathbf{I}$, equation 15 reduces to CBS equation 13. At the other extreme, if $\mathcal{M}_\theta$ approximates $(\mathbf{I} - G_\eta V_\eta)^{-1}$, one NPBS step recovers the exact Helmholtz solution.

**Equivalent differential-preconditioned form.** By the identity equation 8, equation 15 is equivalent to

$$u^{m+1} = u^m + \mathcal{M}_\theta (G_\eta (f - A u^m))\,. \tag{16}$$

This form makes the left-preconditioned interpretation transparent—the input to $\mathcal{M}_\theta$ is $r_{\mathrm{bs}}^m := G_\eta (f - A u^m) = L_\eta^{-1}(f - A u^m)$—but computing $A u^m$ requires an additional FFT for the $-\Delta$ term, so in practice we implement the integral form equation 15. Proposition 1 implies that the preconditioned coordinates enjoy spectral clustering when $\|G_\eta V_\eta\| < 1$, justifying their use for both iteration *and* training.

**Direct-residual baseline.** Given the same learned preconditioner $\mathcal{M}_\theta$, the standard direct-residual iteration is

$$u^{m+1} = u^m + \mathcal{M}_\theta(f - Au^m)\,, \tag{17}$$

which operates in unpreconditioned Euclidean coordinates. The corresponding direct training objective is

$$\mathcal{L}_{\mathrm{dir}}(\theta) = \mathbb{E}_{r\sim\mathcal{N}(0,\mathbf{I})} \frac{\|A\,\mathcal{M}_\theta(r) - r\|_2}{\|r\|_2}. \tag{18}$$

Training with equation 18 is simple and widely used, but it optimizes in unpreconditioned Euclidean residual coordinates. For indefinite Helmholtz, this leads to stiff gradients for near-resonant modes. The preconditioned-metric objective derived next resolves this mismatch.

### 3.1 BORN-SERIES-INSPIRED LOSS AS A PRECONDITIONED RESIDUAL METRIC

The identity established in Section 2 directly motivates a training objective that operates in the same preconditioned geometry as the inference iteration. Rather than minimizing the Euclidean residual $\|A\mathcal{M}_\theta(r) - r\|_2$ as in equation 18, we feed preconditioned probes $G_\eta r$ and minimize the integral-form residual. Let $\mathcal{D}$ denote the training distribution of probe residuals:

$$\mathcal{L}_{\mathrm{bs}}(\theta) = \mathbb{E}_{r\sim\mathcal{D}} \frac{\|(\mathbf{I} - G_\eta V_\eta)\,\mathcal{M}_\theta(G_\eta r) - G_\eta r\|_2}{\|G_\eta r\|_2}. \tag{19}$$

Using equation 8, this is equivalently

$$\mathcal{L}_{\mathrm{bs}}(\theta) = \mathbb{E}_{r\sim\mathcal{D}} \frac{\|G_\eta A\,\mathcal{M}_\theta(G_\eta r) - G_\eta r\|_2}{\|G_\eta r\|_2}. \tag{20}$$

**Preconditioned residual norm / Riesz-map interpretation.** The following result characterizes the loss as a residual norm in a preconditioned inner product, which is the central insight of this work.

**Proposition 2** (Riesz-map equivalence)**.** *Define the preconditioned inner product and norm*

$$R_\eta := G_\eta^* G_\eta, \qquad \|x\|_{R_\eta} := \|G_\eta x\|_2, \qquad \langle x, y\rangle_{R_\eta} := \langle x, R_\eta y\rangle_2.$$

*Then the Born-series-inspired loss equation 20 admits the Riesz-map form*

$$\mathcal{L}_{\mathrm{bs}}(\theta) = \mathbb{E}_{r\sim\mathcal{D}} \frac{\|A\mathcal{M}_\theta(G_\eta r) - r\|_{R_\eta}}{\|r\|_{R_\eta}}. \tag{21}$$

which measures the residual of the original operator $A$ in the $R_\eta$-weighted geometry induced by the left preconditioner $G_\eta = L_\eta^{-1}$. Proposition 2 reveals a dual interpretation: the Born-series-inspired loss is simultaneously an integral residual norm and a Riesz-map loss. In contrast to the direct objective equation 18, which uses unpreconditioned probes $r$ and the Euclidean residual norm, $\mathcal{L}_{\mathrm{bs}}$ both (i) feeds preconditioned probes $G_\eta r$ into $\mathcal{M}_\theta$ and (ii) evaluates residuals in the preconditioned metric $\|\cdot\|_{R_\eta}$. This alignment between training geometry and inference-time preconditioning is what alleviates stiff gradients for near-resonant Helmholtz modes and extends naturally to CDR and linearized Newton systems (Section B).

**Why this mitigates training difficulty.** Compared with equation 18, equation 20 (equivalently equation 21) optimizes against $G_\eta A$ in the $R_\eta$-metric, instead of raw $A$ in the Euclidean metric. This is precisely the same conditioning mechanism used in shifted-Laplacian methods Erlangga et al. (2004). In Fourier coordinates for homogeneous reference medium,

$$\widehat{G_\eta}(\xi) = \frac{1}{|\xi|^2 - k_0^2 - i\eta},$$

which damps high-frequency components and reshapes the optimization landscape. The resulting objective is therefore conditioned in the same coordinates used by the inference iteration.

**Remark 1** (Training/inference consistency)**.** *At inference, the Born residual is*

$$r_{\mathrm{bs}}^m = G_\eta(f - Au^m).$$

*The training input in equation 19 has the same* form $G_\eta r$*. Exact distribution matching holds only if $\mathcal{D}$ is chosen as the empirical distribution of iterative residuals (e.g., rollout replay). In this manuscript we use form-level alignment and treat residual-replay matching as an optional refinement.*

### 3.2 NPBS ITERATION

Sections 2–3.1 developed three ingredients—preconditioned residual coordinates (Section 2), a learned correction map (Section 3), and a matched training metric (Section 3.1)—that together define the NPBS framework. We now collect them into a concise solver statement and discuss computational cost.

For implementation clarity, we restate the NPBS solver in its efficient integral form:

$$u^{m+1} = u^m + \mathcal{M}_\theta(G_\eta(V_\eta u^m + f) - u^m)\,. \tag{22}$$

Since $V_\eta u^m$ is pointwise and $G_\eta$ is applied once via FFT/IFFT (Proposition 3), each iteration uses a single $\mathcal{O}(N \log N)$ transform pair for the preconditioned residual. If $\mathcal{M}_\theta = \gamma \mathbf{I}$, equation 22 reduces to CBS equation 13. If $\mathcal{M}_\theta$ approximates $(\mathbf{I} - G_\eta V_\eta)^{-1}$, then equation 22 approaches the exact Helmholtz correction map.

The method combines:

1. a CBS-compatible preconditioned iteration structure,
2. coefficient-adaptive corrections through $\mathcal{M}_\theta$,
3. a training objective in matched preconditioned coordinates.

This separates solver design (choice of $\mathcal{M}_\theta$) from objective design (choice of residual metric), which is the organizing principle used in the experiments.

**Practical training/inference loop.** For clarity, the implementation is:

1. **Setup:** given PDE coefficients, build/cache setup features for $\mathcal{M}_\theta$ and fix the reference operator $L_\eta$.
2. **Training:** sample residual-like inputs $r$, form $G_\eta r$, and minimize equation 19 (or equation 20).
3. **Inference:** iterate equation 22 until $\|f - Au^m\|_2/\|f\|_2 \leq \mathrm{rTol}$ (definition in Appendix A).

Full derivations, loss formulas, and implementation details are moved to Appendix B. Fast-application verification of the Green operators (including Proposition 3, referenced in Section 3) is moved to Appendix C. The detailed Fourier pseudospectral discretization recipe remains in Appendix D.

## 4 NUMERICAL RESULTS

The preceding sections developed a unified framework with two separable design axes: the *solver architecture* (which correction map $\mathcal{M}_\theta$) and the *training metric* (Euclidean $\mathcal{L}_{\mathrm{dir}}$ vs. preconditioned $\mathcal{L}_{\mathrm{bs}}$). We report solver-level comparison (MgNO-BS vs CBS), loss-level ablation at fixed architecture, and transfer beyond Helmholtz. For Helmholtz, we use OpenFWI CurveFault-B (2D high-contrast). OpenFWI is a large-scale FWI benchmark Deng et al. (2022); the CurveFault-B subset used here contains sharp velocity contrasts and discontinuous interfaces. In our setup, velocities range from approximately 1500 to 4500 m/s, models are interpolated to a $256 \times 256$ grid, an absorbing layer of width 32 grid points is used, and $\texttt{ppw} = 12$. Detailed runtime and reproducibility settings are listed in Appendix A.

### 4.1 HELMHOLTZ SOLVER/ARCHITECTURE BENEFIT (MGNO-BS VS CBS)

Table 1 reports the primary solver-level comparison on OpenFWI. Across the two precision settings, MgNO-BS reduces iterations by $55.5\times$–$68.6\times$ and wall time by $2.4\times$–$4.5\times$. Here, `Iter Ratio` = (CBS iters)/(MgNO-BS iters) and `Time Ratio` = (CBS time)/(MgNO-BS time).

Figure 1 provides representative OpenFWI residual-decay and inference examples consistent with Table 1.

Table 1: Solver-level comparison (architecture effect): CBS vs learned-preconditioner MgNO-BS on OpenFWI CurveFault-B.

| Case | Precision / rTol | CBS Iters | MgNO-BS Iters | Iter Ratio | CBS Time | MgNO-BS Time | Time Ratio |
|---|---|---|---|---|---|---|---|
| OpenFWI CurveFault-B (2D) | Mixed / $10^{-6}$ | 1124.5 | 16.4 | 68.6 | 0.157 | 0.035 | 4.5 |
| OpenFWI CurveFault-B (2D) | FP32 / $10^{-4}$ | 533.0 | 9.6 | 55.5 | 0.050 | 0.021 | 2.4 |

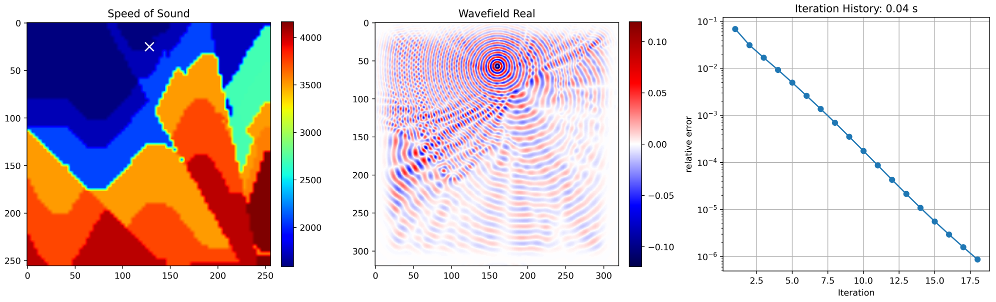

Figure 1: OpenFWI CurveFault-B: MgNO-BS prediction and residual decay.

### 4.2 LOSS-DESIGN BENEFIT AT FIXED SOLVER

Table 2 isolates objective design while keeping architecture and inference fixed. On OpenFWI CurveFault-B (high-contrast), $\mathcal{L}_{\text{bs}}$ reduces iteration count from $27.0$ to $19.5$ and wall time from $0.056$s to $0.041$s relative to $\mathcal{L}_{\text{dir}}$, indicating a clear loss-design effect in this regime.

Table 2: Loss ablation at fixed learned-preconditioner architecture: direct residual loss $\mathcal{L}_{\text{dir}}$ vs Born-inspired loss $\mathcal{L}_{\text{bs}}$.

| Dataset | Loss | Avg. Iters | Avg. Time (s) |
|---|---|---|---|
| OpenFWI CurveFault-B | $\mathcal{L}_{\text{dir}}$ | 27.0 | 0.056 |
| OpenFWI CurveFault-B | $\mathcal{L}_{\text{bs}}$ | 19.5 | 0.041 |

### 4.3 TRANSFER BEYOND HELMHOLTZ

#### 4.3.1 CDR WITH PHYSICS-AWARE FNO

For the linear CDR equation

$$-\nabla \cdot (\kappa \nabla u) + \mathbf{v} \cdot \nabla u + \sigma u = f \quad \text{in } \Omega \subset \mathbb{R}^d, \tag{23}$$

with periodic boundary conditions, we use PA-FNO (Eq. equation 45) as the learned preconditioner $\mathcal{M}_\theta$. We test two complementary settings:

**Loss-isolation on a synthetic CDR dataset.** To isolate the training-metric effect in the CDR setting, we generate a $128 \times 128$ periodic dataset with heterogeneous $\kappa, \mathbf{v}, \sigma$ (construction details in Appendix E). Table 3 compares three methods at the tighter tolerance rTol $= 10^{-6}$: direct-loss PA-FNO, Born-series-loss PA-FNO (PA-FNO-BS), and the non-learned BS fixed-point baseline. PA-FNO-BS reduces iteration count by $34\%$ relative to direct PA-FNO (12.2→8.0), and both learned variants substantially outperform the fixed-point baseline (18.6 iterations).

Figure 2 shows representative residual histories.

#### 4.3.2 NONLINEAR PDE WITH NEWTON-BORN PRECONDITIONING

We consider

$$\begin{cases} -\Delta u(x,y) - u(x,y)^2 = -s\sin(\pi x)\sin(\pi y), & (x,y) \in \Omega, \\ u(x,y) = 0, & (x,y) \in \partial\Omega, \end{cases} \tag{24}$$

| **Method** | **Iters** | **rTol** |
|---|---|---|
| PA-FNO (direct loss) | 12.2 | $1 \times 10^{-6}$ |
| PA-FNO-BS (Born-series loss) | 8.00 | $1 \times 10^{-6}$ |
| BS fixed-point | 18.60 | $1 \times 10^{-6}$ |

Table 3: CDR loss-isolation: iteration counts on the synthetic $128 \times 128$ periodic dataset under the same relative tolerance.

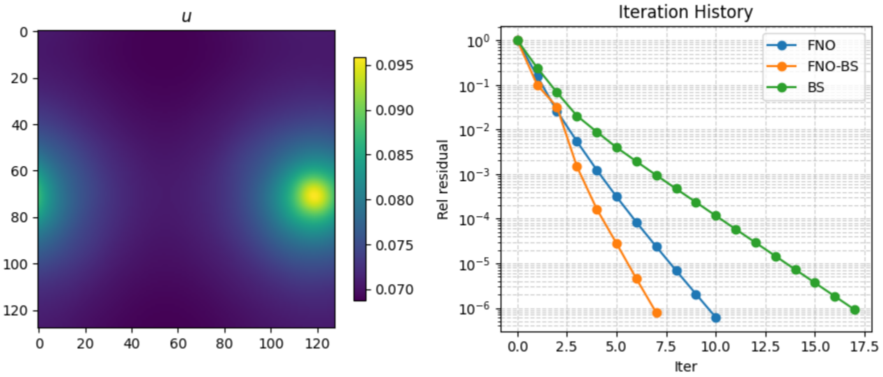

Figure 2: CDR residual histories on the synthetic dataset: PA-FNO-BS vs BS fixed-point vs direct PA-FNO.

with $f(x,y) := -s\sin(\pi x)\sin(\pi y)$ and nonlinear residual

$$\mathcal{F}_{\mathrm{nl}}(u) := -\Delta u - u^2 - f.$$

At Newton iterate $u^{(m)}$, we solve

$$\begin{cases} -\Delta(\delta u^{(m)}) - 2u^{(m)}\,\delta u^{(m)} = -\mathcal{F}_{\mathrm{nl}}(u^{(m)}), & (x,y) \in \Omega, \\ \delta u^{(m)} = 0, & (x,y) \in \partial\Omega. \end{cases} \tag{25}$$

and update $u^{(m+1)} = u^{(m)} + \delta u^{(m)}$.

Dataset construction and Newton-trajectory generation details are given in Appendix E, following the nonlinear PDE benchmark setting in Hao et al. (2024).

Figure 3 and Fig. 4 show faster inner-loop residual reduction for the preconditioned model. Table 4 reports the same trend over 50 test samples: Born-series preconditioning reduces average inner-loop iterations by $4.2\times$ (58.0→13.9).

**Empirical storyline summary.** The two design axes are now empirically confirmed. At the solver level, replacing the CBS scalar correction with a learned operator gives $55.5\times$–$68.6\times$ fewer itera-

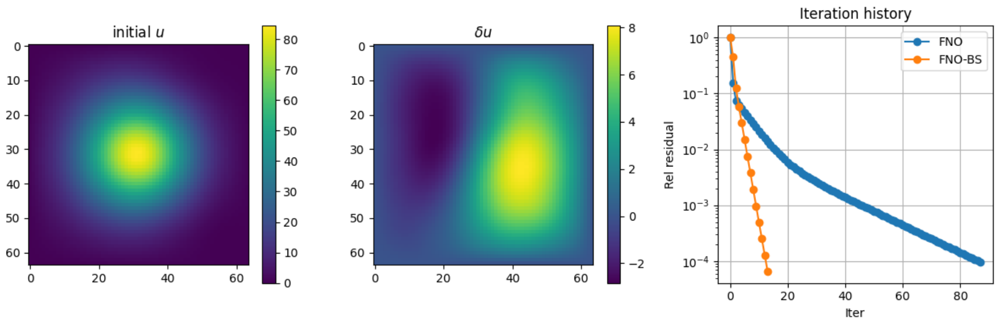

Figure 3: Nonlinear inference comparison: PA-FNO-BS vs direct PA-FNO.

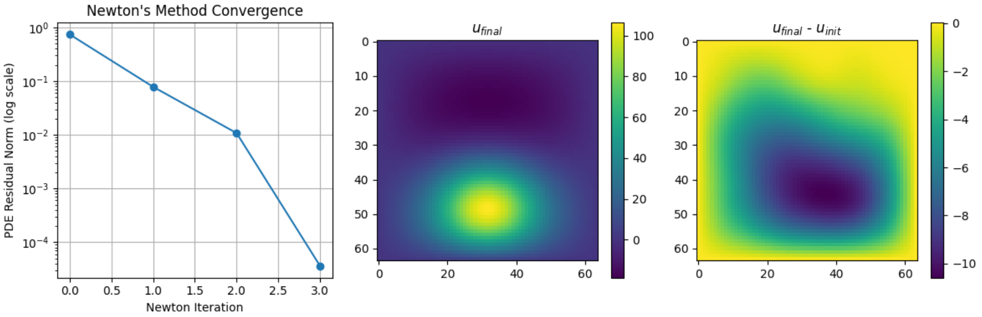

Figure 4: PA-FNO-BS preconditioner embedded in Newton iterations.

| **Method** | **Iters** | **rTol** |
|---|---|---|
| PA-FNO (direct loss) | 58.00 | $1 \times 10^{-4}$ |
| PA-FNO-BS (Born-series loss) | 13.85 | $1 \times 10^{-4}$ |

Table 4: Nonlinear Newton inner-loop iteration counts under the same relative tolerance.

tions on OpenFWI. At fixed architecture, training in the preconditioned $R_\eta$-metric (Proposition 2) yields $\sim 28\%$ fewer iterations on high-contrast media. The same preconditioned-coordinate construction transfers to CDR (PA-FNO-BS: $34\%$ fewer iterations than direct PA-FNO) and linearized Newton systems from nonlinear PDEs ($4.2\times$ iteration reduction).

## 5 CONCLUSION

We have shown that the training of learned PDE preconditioners can be cast as a *metric-design problem*: the identity $\mathbf{I} - G_\eta V_\eta = G_\eta A = L_\eta^{-1} A$ that unifies CBS, Born-series, and shifted-Laplacian viewpoints also defines a preconditioned Riesz-map metric for training (Proposition 2). By aligning the objective geometry with the inference geometry, the resulting Born-series-inspired loss inherits the conditioning benefits of classical shifted-Laplacian preconditioning.

Our controlled experiments decompose the improvement into two orthogonal axes—a solver-level benefit from replacing scalar CBS corrections with a learned operator (up to $68.6\times$ fewer iterations on OpenFWI), and a loss-level benefit from switching to the preconditioned metric ($\sim 28\%$ improvement on high-contrast media)—and demonstrate that both effects transfer to convection–diffusion–reaction and linearized Newton systems from nonlinear PDEs. The framework is agnostic to the choice of learned preconditioner architecture and requires only that a constant-coefficient reference inverse be available via fast transforms, making it a broadly applicable drop-in improvement for learned iterative solvers.

## REFERENCES

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, pp. 1–9, 2024.

Ivo Babuška and Stefan A Sauter. Is the pollution effect of the fem avoidable for the helmholtz equation considering high wave numbers? *SIAM Journal on Numerical Analysis*, 34(6):2392–2423, 1997.

John P Boyd. *Chebyshev and Fourier spectral methods*. Courier Corporation, 2001.

Achi Brandt and Irene Livshits. Wave-ray multigrid method for standing wave equations. *Electronic Transactions on Numerical Analysis*, 6:162–181, 1997.

Claudio Canuto, M. Yousuff Hussaini, Alfio Quarteroni, and Thomas A Zang. *Spectral Methods: Fundamentals in Single Domains*. Springer, 2006.

Shuhao Cao. Choose a transformer: Fourier or galerkin. *Advances in Neural Information Processing Systems*, 34, 2021.

Xiang Cao, Qiaoqiao Ding, Xinliang Liu, Lei Zhang, and Xiaoqun Zhang. Diff-ano: Towards fast high-resolution ultrasound computed tomography via conditional consistency models and adjoint neural operators. *arXiv preprint arXiv:2507.16344*, 2025.

Zhiming Chen and Xueshuang Xiang. A source transfer domain decomposition method for helmholtz equations in unbounded domain. *SIAM Journal on Numerical Analysis*, 51(4):2331–2356, 2013.

Chen Cui, Kai Jiang, and Shi Shu. A neural multigrid solver for helmholtz equations with high wavenumber and heterogeneous media. *SIAM Journal on Scientific Computing*, 47(3):C655–C679, 2025.

Chengyuan Deng, Shihang Feng, Hanchen Wang, Xitong Zhang, Peng Jin, Yinan Feng, Qili Zeng, Yinpeng Chen, and Youzuo Lin. OpenFWI: Large-scale multi-structural benchmark datasets for full waveform inversion. *Advances in Neural Information Processing Systems*, 35:6007–6020, 2022.

Yogi A Erlangga, Cornelis Vuik, and Cornelis Willebrordus Oosterlee. On a class of preconditioners for solving the helmholtz equation. *Applied Numerical Mathematics*, 50(3-4):409–425, 2004.

Oliver G Ernst and Martin J Gander. Why it is difficult to solve helmholtz problems with classical iterative methods. *Numerical analysis of multiscale problems*, pp. 325–363, 2011.

David Ha, Andrew Dai, and Quoc V Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016.

Wenrui Hao, Xinliang Liu, and Yahong Yang. Newton informed neural operator for solving nonlinear partial differential equations. In *Advances in Neural Information Processing Systems*, 2024. URL `https://openreview.net/forum?id=F9mNL6vR27`.

Juncai He, Xinliang Liu, and Jinchao Xu. Mgno: Efficient parameterization of linear operators via multigrid. In *International Conference on Learning Representations*, 2024.

Juncai He, Xinliang Liu, and Jinchao Xu. Self-composing neural operators with depth and accuracy scaling via adaptive train-and-unroll approach. *arXiv preprint arXiv:2508.20650*, 2025.

Xu Jia, Bert De Brabandere, Tinne Tuytelaars, and Luc Van Gool. Dynamic filter networks. In *Advances in neural information processing systems*, pp. 667–675, 2016.

Wei Leng and Lili Ju. Trace transfer-based diagonal sweeping domain decomposition method for the helmholtz equation: Algorithms and convergence analysis. *Journal of Computational Physics*, 455:110980, 2022.

Bar Lerer, Ido Ben-Yair, and Eran Treister. Multigrid-augmented deep learning preconditioners for the helmholtz equation using compact implicit layers. *SIAM Journal on Scientific Computing*, 46 (5):S123–S144, 2024.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *arXiv preprint arXiv:2207.05209*, 2022.

Xinliang Liu, Bo Xu, Shuhao Cao, and Lei Zhang. Mitigating spectral bias for the multiscale operator learning. *Journal of Computational Physics*, 506:112944, 2024.

Irene Livshits and Achi Brandt. Accuracy properties of the wave-ray multigrid algorithm for helmholtz equations. *SIAM Journal on Scientific Computing*, 28(4):1228–1251, 2006.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Gerwin Osnabrugge, Saroch Leedumrongwatthanakun, and Ivo M Vellekoop. A convergent born series for solving the inhomogeneous helmholtz equation in arbitrarily large media. *Journal of computational physics*, 322:113–124, 2016.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International conference on machine learning*, pp. 5301–5310. PMLR, 2019.

Md Ashiqur Rahman, Zachary E Ross, and Kamyar Azizzadenesheli. U-no: U-shaped neural operators. *arXiv e-prints*, pp. arXiv–2204, 2022.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Alexander Rudikov, Vladimir Fanaskov, Ekaterina Muravleva, Yuri M Laevsky, and Ivan Oseledets. Neural operators meet conjugate gradients: The fcg-no method for efficient pde solving. *arXiv preprint arXiv:2402.05598*, 2024.

Antonio Stanziola, Simon R Arridge, Ben T Cox, and Bradley E Treeby. A helmholtz equation solver using unsupervised learning: Application to transcranial ultrasound. *Journal of Computational Physics*, 441:110430, 2021.

Christiaan C Stolk. A dispersion minimizing scheme for the 3-d helmholtz equation based on ray theory. *Journal of Computational Physics*, 314:618–646, 2016.

Lloyd N Trefethen. *Spectral Methods in MATLAB*. SIAM, Philadelphia, 2000.

Yan Xie, Minrui Lv, and Chen-Song Zhang. Mgcfnn: A neural multigrid solver with novel fourier neural network for high wave number helmholtz equations. In *The Thirteenth International Conference on Learning Representations*, 2025.

Brandon Yang, Gabriel Bender, Quoc V Le, and Jiquan Ngiam. Condconv: Conditionally parameterized convolutions for efficient inference. In *Advances in neural information processing systems*, pp. 1305–1316, 2019.

Enrui Zhang, Adar Kahana, Alena Kopaničáková, Eli Turkel, Rishikesh Ranade, Jay Pathak, and George Em Karniadakis. Blending neural operators and relaxation methods in pde numerical solvers. *Nature Machine Intelligence*, pp. 1–11, 2024.

## A REPRODUCIBILITY PROTOCOL

All runs use PyTorch 2.6.0 with CUDA 12.6 on a workstation with Intel Xeon Gold 6444Y CPU and one NVIDIA RTX A6000 GPU; each table reports means over 50 test samples.

Runtime in Tables 1 and 2 is measured with batch size 1 and CUDA synchronization immediately before/after each solve. All iterative kernels (network forward, residual/operator evaluation, and FFT/DST/DCT applications of $G_\eta$) run on GPU; CPU is used for orchestration only. Reported wall time includes the full iterative solve loop and excludes dataset I/O and offline training/setup.

We define

$$\mathrm{rTol} = \frac{\|f - Au^m\|_2}{\|f\|_2},$$

i.e., the relative *unpreconditioned physical residual norm.* `FP32` denotes single precision throughout; `Mixed` denotes FP32 iterative updates with FP64 residual evaluation for stopping checks. For each dataset, the same validated shift $\eta$ is used in both training and inference.

# B FULL DERIVATIONS FOR CDR AND NONLINEAR EXTENSIONS

This appendix gives the full technical derivations deferred from the main text.

## B.1 CONVECTION-DIFFUSION-REACTION (LINEAR)

Consider

$$\mathcal{L}_{\mathrm{cdr}} u = -\nabla \cdot (\kappa(x) \nabla u) \; + \; \mathbf{v}(x) \cdot \nabla u \; + \; \sigma(x) u = f. \tag{26}$$

Choose background constants $\kappa_0, \mathbf{v}_0, \sigma_0$ and define

$$\kappa = \kappa_0 + \delta\kappa, \qquad \mathbf{v} = \mathbf{v}_0 + \delta\mathbf{v}, \qquad \sigma = \sigma_0 + \delta\sigma.$$

Set

$$\mathcal{L}_{0,\mathrm{cdr}} u := -\kappa_0 \Delta u \; + \; \mathbf{v}_0 \cdot \nabla u \; + \; \sigma_0 u, \tag{27}$$

and define $\mathcal{V}_{\mathrm{cdr}}$ by $\mathcal{L}_{\mathrm{cdr}} = \mathcal{L}_{0,\mathrm{cdr}} - \mathcal{V}_{\mathrm{cdr}}$, i.e.,

$$\mathcal{V}_{\mathrm{cdr}} u = \nabla \cdot (\delta\kappa \nabla u) \; - \; \delta\mathbf{v} \cdot \nabla u \; - \; \delta\sigma \, u. \tag{28}$$

Then

$$\mathcal{L}_{0,\mathrm{cdr}} u = \mathcal{V}_{\mathrm{cdr}} u + f. \tag{29}$$

With $G_{\mathrm{cdr}} := \mathcal{L}_{0,\mathrm{cdr}}^{-1}$:

$$(\mathbf{I} - G_{\mathrm{cdr}} \mathcal{V}_{\mathrm{cdr}}) u = G_{\mathrm{cdr}} f = G_{\mathrm{cdr}} \mathcal{L}_{\mathrm{cdr}} u. \tag{30}$$

In discrete form $(A_{\mathrm{cdr}}, V_{\mathrm{cdr}}, G_{\mathrm{cdr}})$:

$$\mathbf{I} - G_{\mathrm{cdr}} V_{\mathrm{cdr}} = G_{\mathrm{cdr}} A_{\mathrm{cdr}}. \tag{31}$$

Therefore, the Born-inspired training objective is

$$\mathcal{L}_{\mathrm{bs}}^{\mathrm{cdr}}(\theta) = \mathbb{E}_{r \sim \mathcal{D}} \frac{\|(\mathbf{I} - G_{\mathrm{cdr}} V_{\mathrm{cdr}}) \mathcal{M}_\theta(G_{\mathrm{cdr}} r) - G_{\mathrm{cdr}} r\|_2}{\|G_{\mathrm{cdr}} r\|_2} \tag{32}$$

or equivalently

$$\mathcal{L}_{\mathrm{bs}}^{\mathrm{cdr}}(\theta) = \mathbb{E}_{r \sim \mathcal{D}} \frac{\|G_{\mathrm{cdr}} A_{\mathrm{cdr}} \mathcal{M}_\theta(G_{\mathrm{cdr}} r) - G_{\mathrm{cdr}} r\|_2}{\|G_{\mathrm{cdr}} r\|_2}. \tag{33}$$

**Discussion (CDR).** For periodic domains, $G_{\mathrm{cdr}}$ is a convolution operator with Fourier symbol

$$\widehat{G}_{\mathrm{cdr}}(\xi) = \frac{1}{\kappa_0 |\xi|^2 + i \mathbf{v}_0 \cdot \xi + \sigma_0}, \tag{34}$$

so high-frequency diffusion-dominated modes are attenuated in training. As in Helmholtz, optimizing against $G_{\mathrm{cdr}} A_{\mathrm{cdr}}$ instead of raw $A_{\mathrm{cdr}}$ improves conditioning and better matches the residual distribution in preconditioned iterations.

## B.2 NONLINEAR PDE VIA NEWTON-BORN LINEARIZATION

For the nonlinear example

$$\mathcal{F}_{\mathrm{nl}}(u) := -\Delta u - u^2 - f = 0, \tag{35}$$

Newton at iterate $u^{(m)}$ solves

$$J_m \, \delta u^{(m)} = r^{(m)}, \qquad J_m := -\Delta - 2u^{(m)}, \qquad r^{(m)} := -\mathcal{F}_{\mathrm{nl}}(u^{(m)}), \tag{36}$$

followed by $u^{(m+1)} = u^{(m)} + \delta u^{(m)}$.

To build a Born-style preconditioned form for each Newton step, pick a reference Jacobian

$$J_{0,m} := -\Delta - 2\bar{u}^{(m)} + \alpha_m \mathbf{I}, \qquad \bar{u}^{(m)} := \text{spatial average of } u^{(m)}, \qquad \alpha_m \geq 0, \tag{37}$$

and define

$$V_m := J_{0,m} - J_m. \tag{38}$$

Then $J_m = J_{0,m} - V_m$, $G_m := J_{0,m}^{-1}$, and

$$(\mathbf{I} - G_m V_m)\, \delta u^{(m)} = G_m r^{(m)} \quad \Leftrightarrow \quad \mathbf{I} - G_m V_m = G_m J_m. \tag{39}$$

Hence the Born-inspired loss for nonlinear problems (trained on Newton states) is

$$\mathcal{L}_{\mathrm{bs}}^{\mathrm{nl}}(\theta) = \mathbb{E}_{(u^{(m)}, r) \sim \mathcal{D}_{\mathrm{N}}} \frac{\|(\mathbf{I} - G_m V_m)\mathcal{M}_\theta(G_m r) - G_m r\|_2}{\|G_m r\|_2}, \tag{40}$$

equivalently

$$\mathcal{L}_{\mathrm{bs}}^{\mathrm{nl}}(\theta) = \mathbb{E}_{(u^{(m)}, r) \sim \mathcal{D}_{\mathrm{N}}} \frac{\|G_m J_m \mathcal{M}_\theta(G_m r) - G_m r\|_2}{\|G_m r\|_2}. \tag{41}$$

**Discussion (nonlinear).** The loss is computed on *linearized* systems along Newton trajectories, not directly on the nonlinear map. This has two effects:

1. It preserves the same preconditioned-coordinate advantage as the linear case (better-conditioned operator equation).
2. It aligns training with actual inference, where the solver repeatedly tackles Jacobian systems with changing $u^{(m)}$.

In practice, one may use an inner iteration

$$\delta u_{k+1}^{(m)} = \delta u_k^{(m)} \; + \; \mathcal{M}_\theta\Big(G_m \left(r^{(m)} - J_m \delta u_k^{(m)}\right)\Big), \tag{42}$$

then set $\delta u^{(m)} \approx \delta u_K^{(m)}$ and update $u^{(m+1)} = u^{(m)} + \delta u^{(m)}$.

## C VERIFICATION: FAST FFT IMPLEMENTATION OF GREEN OPERATORS (ALL CASES)

We verify, case by case, when Green-operator application is fast. A concrete Fourier pseudospectral discretization recipe for $L_\eta$ and $G_\eta$ is provided in Appendix D.

**Proposition 3** (Transform-diagonalizable Green operators). *On a uniform grid, if the chosen reference operator has constant coefficients and compatible boundary conditions (periodic, or transform-compatible Dirichlet/Neumann), then its discrete matrix is diagonalizable by a fast trigonometric transform. Consequently, applying the corresponding discrete Green operator costs* $\mathcal{O}(N \log N)$.

*Constructive verification.* Let $L_0$ be any constant-coefficient reference operator and $G_0 = L_0^{-1}$. Under periodic boundary conditions, the discrete $L_0$ is block-circulant with circulant blocks, hence

$$L_0 = F^{-1} \Lambda F, \qquad G_0 = F^{-1} \Lambda^{-1} F,$$

where $F$ is the multidimensional DFT matrix and $\Lambda$ contains the symbol values. Therefore

$$G_0 q = \mathcal{F}^{-1}\left(\frac{\widehat{q}}{\lambda(\xi)}\right),$$

implemented by FFT/IFFT in $\mathcal{O}(N \log N)$. For Dirichlet or Neumann boundaries, the same statement holds with sine/cosine transforms (DST/DCT), which are FFT-based and have the same complexity order. ☐

**Helmholtz case.** With $L_\eta = -\Delta - (k_0^2 + i\eta)$, the symbol is

$$\lambda_{\mathrm{H}}(\xi) = |\xi|^2 - (k_0^2 + i\eta),$$

so

$$G_\eta q = \mathcal{F}^{-1}\left(\frac{\widehat{q}}{\lambda_{\mathrm{H}}(\xi)}\right).$$

Hence the Born residual $G_\eta r$ is FFT-fast under the stated boundary setting.

**CDR case.** With $\mathcal{L}_{0,\mathrm{cdr}} = -\kappa_0\Delta + \mathbf{v}_0 \cdot \nabla + \sigma_0$, the symbol is exactly equation 34 denominator:

$$\lambda_{\mathrm{cdr}}(\xi) = \kappa_0|\xi|^2 + i\mathbf{v}_0 \cdot \xi + \sigma_0.$$

Thus

$$G_{\mathrm{cdr}}q = \mathcal{F}^{-1}\left(\frac{\widehat{q}}{\lambda_{\mathrm{cdr}}(\xi)}\right),$$

again $\mathcal{O}(N \log N)$.

**Newton-linearized system case.** At Newton step $m$, with reference Jacobian $J_{0,m} = -\Delta - 2\bar{u}^{(m)} + \alpha_m\mathbf{I}$,

$$\lambda_{\mathrm{N},m}(\xi) = |\xi|^2 + \alpha_m - 2\bar{u}^{(m)},$$

so

$$G_m q = \mathcal{F}^{-1}\left(\frac{\widehat{q}}{\lambda_{\mathrm{N},m}(\xi)}\right),$$

which is also FFT-fast.

**Boundary-condition note for all cases.** In scattering settings, we use an absorbing layer (e.g., sponge/PML-style damping) on a finite box and choose a *constant-coefficient* reference operator on that computational domain; this is the operator inverted by FFT in the Born-inspired loss/iteration. If strict Dirichlet/Neumann boundaries are required, use DST/DCT (or FFT-based embedding), preserving near-identical complexity.

## D FOURIER PSEUDOSPECTRAL DISCRETIZATION OF $L_\eta$ AND $G_\eta$

This appendix gives the concrete discretization used when $L_\eta$ and $G_\eta = L_\eta^{-1}$ are implemented by FFT under periodic boundary conditions. In numerical-analysis terminology, this is a *Fourier pseudospectral* (or Fourier collocation) discretization: derivatives are represented by exact Fourier multipliers, while coefficient multiplications are applied pointwise in physical space Boyd (2001); Trefethen (2000); Canuto et al. (2006).

**Periodic grid and Fourier modes.** Let $\Omega = \prod_{\alpha=1}^{d}[0, L_\alpha)$ and use a uniform grid with $N_\alpha$ points in direction $\alpha$:

$$x_{j_\alpha} = j_\alpha h_\alpha, \qquad h_\alpha = L_\alpha/N_\alpha, \qquad j_\alpha = 0, \dots, N_\alpha - 1.$$

For grid field $u_j = u(x_j)$, define the DFT coefficients $\widehat{u}_n$ with multi-index

$$n = (n_1, \dots, n_d), \qquad n_\alpha \in \Big\{-\frac{N_\alpha}{2}, \dots, \frac{N_\alpha}{2} - 1\Big\},$$

and physical wavenumbers

$$\xi_{\alpha,n} = \frac{2\pi n_\alpha}{L_\alpha}, \qquad |\xi_n|^2 = \sum_{\alpha=1}^{d} \xi_{\alpha,n}^2.$$

**Discrete symbol of $L_\eta$.** For

$$L_\eta = -\Delta - (k_0^2 + i\eta), \qquad \eta > 0,$$

the Fourier symbol is

$$\lambda_n = |\xi_n|^2 - (k_0^2 + i\eta),$$

so

$$\widehat{L_\eta u}_n = \lambda_n\, \widehat{u}_n.$$

Hence $L_\eta$ is diagonal in Fourier space, and its inverse acts modewise:

$$\widehat{(G_\eta q)}_n = \frac{\widehat{q}_n}{\lambda_n}.$$

Because $\eta > 0$, $\lambda_n$ has nonzero imaginary part and the modal inversion is well-defined.

**FFT implementation of $G_\eta q$.** Given $q$, compute:

1. $\widehat{q} = \mathcal{F}(q)$ (FFT),
2. $\widehat{z}_n = \widehat{q}_n / \lambda_n$ (pointwise complex division),
3. $z = \mathcal{F}^{-1}(\widehat{z})$ (IFFT).

Then $z = G_\eta q$. The complexity is $\mathcal{O}(N \log N)$ with $N = \prod_\alpha N_\alpha$.

**Relation to $A$ and $V_\eta$.** For heterogeneous Helmholtz,

$$Au = -\Delta u - k(x)^2 u, \qquad V_\eta u = (k(x)^2 - k_0^2 - i\eta)u.$$

In Fourier pseudospectral form:

$$-\Delta u = \mathcal{F}^{-1}\big(|\xi|^2 \widehat{u}\big),$$

while $k(x)^2 u$ and $V_\eta u$ are computed pointwise in physical space. This derivative-in-Fourier / product-in-physical split is exactly the pseudospectral pattern Boyd (2001); Trefethen (2000). For strongly non-smooth coefficients, standard de-aliasing (e.g., $2/3$ truncation) can be used Canuto et al. (2006).

**Boundary-condition scope.** The FFT diagonalization above is exact for periodic boundaries. For homogeneous Dirichlet/Neumann boundaries, the same constant-coefficient reference operators are diagonalized by sine/cosine transforms (DST/DCT), with the same $\mathcal{O}(N \log N)$ complexity class. In this manuscript, $G_\eta$ in CBS/MgNO-BS is implemented with the periodic/transform-compatible constant-coefficient reference operator to preserve fast application.

# E TRANSFER EXPERIMENTAL SETUP DETAILS

This appendix collects dataset-generation and protocol details for the transfer experiments.

## E.1 CDR SYNTHETIC DATASET

For the CDR loss-isolation test, we generate data on a $128 \times 128$ periodic grid. Diffusion $\kappa$ is sampled from a Gaussian random field (GRF). The incompressible velocity is built from a GRF stream function $\psi$:

$$v_x = \frac{\partial \psi}{\partial y}, \qquad v_y = -\frac{\partial \psi}{\partial x}.$$

The source term is a Gaussian bump with randomized center and scale. Representative samples are shown in Figure 5.

## E.2 NONLINEAR NEWTON DATASET

For the nonlinear Newton test, we use $\Omega = [0,1]^2$ with homogeneous Dirichlet boundary conditions and $s = 1600$, following the dataset/benchmark family in Hao et al. (2024). The Laplacian is discretized by a 5-point finite-difference stencil on $N = 63$ interior points, so the Jacobian is $J(u) = (-\Delta) - 2u$. Newton trajectories are generated with sparse direct solves and used as supervision. Initial guesses are perturbed from a reference solution via low-frequency sine-series noise:

$$u^0 = u^\star + \alpha \sum_{i,j=1}^{4} c_{ij} \sin(i\pi x)\sin(j\pi y), \quad c_{ij} \sim \mathcal{N}(0, 0.1^2),\ \alpha = 4.0.$$

Figure 6 shows representative solution branches from different initial guesses.

# F ARCHITECTURAL INSTANTIATIONS

The theory in the main text applies to any learned preconditioner $\mathcal{M}_\theta$ that is linear in the residual input. For completeness, we describe the two concrete instantiations used in our experiments: MgNO (for Helmholtz) and PA-FNO (for CDR).

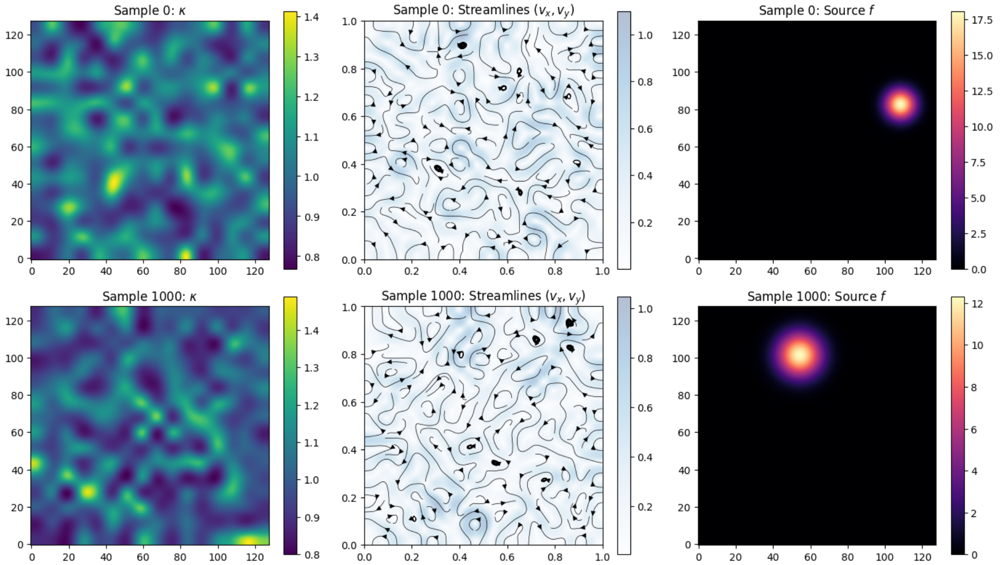

Figure 5: CDR dataset samples: diffusion $\kappa$, velocity field $\mathbf{v}$, and source term $f$.

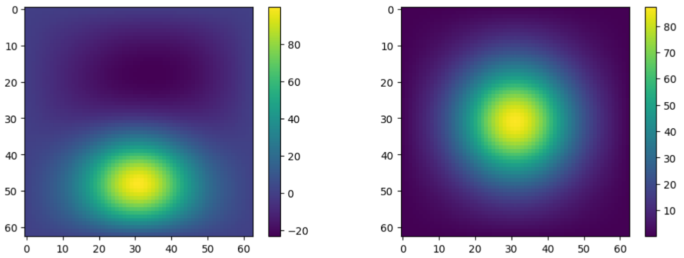

Figure 6: Multiple solutions of the nonlinear PDE under different initial guesses.

**Relation to prior architecture work.** The MGNO multigrid backbone follows the self-composing neural-operator framework of He et al. (2024) He et al. (2025), together with the multigrid-parameterized operator design in He et al. (2024). This design is related to recent multigrid-inspired operator/preconditioner methods Lerer et al. (2024); Xie et al. (2025); Cui et al. (2025), and is complementary to Diff-ANO Cao et al. (2025) under the same preconditioned-iteration viewpoint.

**MgNO preconditioner .** MgNO follows a V-cycle with an offline–online split. Offline, nonlinear modules encode coefficient fields into level-wise physics features. Online, linear-in-residual operators perform smoothing, restriction, coarse correction, and prolongation in multi-channel latent spaces. This preserves the preconditioner role while enabling coefficient-adaptive corrections.

**Physics-aware Convolution (PhysConv).** A key module is a coefficient-conditioned convolution that converts local physics into spatially varying stencil weights. This is related to dynamic-filter and hypernetwork ideas Jia et al. (2016); Ha et al. (2016); Yang et al. (2019); Perez et al. (2018), but used here inside multigrid operators and smoothers. Given a coefficient input $a$ and a residual-like field $x$, PhysConv has:

$$c(a) = \varphi(\mathcal{K}_a * a)\,, \tag{43}$$

$$\mathrm{PhysConv}(x; a) = c(a) \odot (\mathcal{K}_x * x)\,. \tag{44}$$

Eq. equation 43 is setup-only (cached per PDE instance), while Eq. equation 44 is linear in $x$ and reused every iteration. This is the mechanism by which MgNO keeps inference efficient while adapting stencils to the local medium.

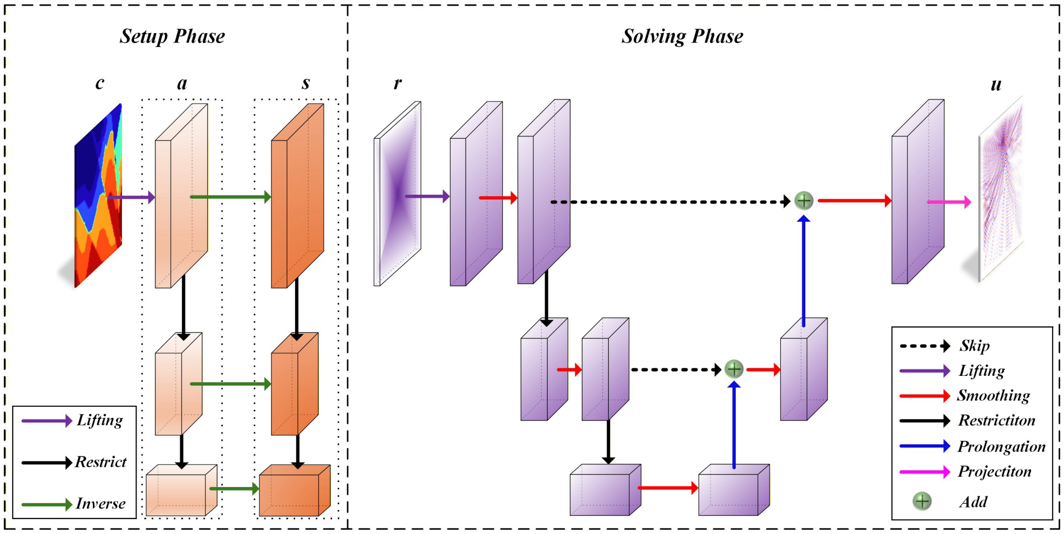

Figure 7: MgNO architecture adapted from prior MgNO work. Setup computes coefficient-dependent features once, and each solve call applies a linear residual-to-correction V-cycle.

**PA-FNO for CDR.** For the convection-diffusion-reaction case we use a physics-aware nonlinear FNO, denoted **PA-FNO**. Let $a = (\kappa, \mathbf{v}, \sigma)$ be coefficient fields and $r$ the residual. The module is nonlinear in $a$ and linear in $r$, matching the preconditioned-iteration requirement. A concrete realization consistent with our implementation is

$$\mathcal{M}_\theta(r;a) = \mathcal{P}\left[\mathrm{LFL}\Big(\mathrm{NLFL}(\Phi_\theta(a)) \odot \mathrm{LFL}(\Psi_\theta(r))\Big)\right], \tag{45}$$

with

$$\Phi_\theta(a) = W_a\,\mathrm{LN}(\mathrm{pad}(a)), \qquad \Psi_\theta(r) = W_r\,\mathrm{pad}(r), \qquad \mathcal{P} = W_p\,\mathrm{unpad}(\cdot), \tag{46}$$

where $W_a$ is a $3 \times 3$ convolutional lift for $a$, $W_r$ is a $1 \times 1$ lift for $r$, and $W_p$ is a $1 \times 1$ projection. The operator NLFL is a stack of nonlinear Fourier layers applied to $\Phi_\theta(a)$, while LFL denotes linear Fourier layers (no activation) applied to the residual branch. The elementwise product $\odot$ fuses coefficient-dependent features with the residual stream before the final linear Fourier map. This yields a residual-to-correction map that is linear in $r$ and coefficient-adaptive through $a$.

For a feature field $x$, we define the linear Fourier layer (LFL) and nonlinear Fourier layer (NLFL) as

$$\mathrm{LFL}(x) = \mathcal{F}^{-1}\big(\mathcal{W}_\theta^{\mathrm{F}} \odot \mathcal{F}(x)\big) \; + \mathcal{W}_\theta^{\mathrm{loc}} * x, \tag{47}$$

$$\mathrm{NLFL}(x) = \varphi\big(\mathrm{LFL}(x)\big), \tag{48}$$

where $\mathcal{W}_\theta^{\mathrm{F}}$ are the Fourier-mode weights (truncated to selected wavenumbers), $\mathcal{W}_\theta^{\mathrm{loc}}$ is the local $1 \times 1$ convolution kernel, $*$ is spatial convolution, and $\varphi$ is a pointwise nonlinearity (GELU in our implementation).