EQUATION-COORDINATE-INVARIANT RESIDUAL METRICS FOR PHYSICS-INFORMED NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Physics-informed neural networks (PINNs) can learn different solutions from algebraically equivalent formulations of the same partial differential equation (PDE). For coupled systems, invertible scaling and mixing preserve the physical constraints but change standard residual losses and their optimization geometry. We formalize this dependence as equation-coordinate bias and introduce ECI-PINN, which replaces the algebraic residual norm with the minimum correction required to satisfy the locally linearized PDE constraints in a dimensionless jet space. The resulting full residual metric accounts for coupling between equations rather than normalizing each equation independently. When the residual Jacobian with respect to the jet has full row rank, we prove that the unregularized metric preserves loss values and detached parameter gradients under pointwise invertible residual-basis transformations independent of network parameters and jet state. Matched deterministic training therefore follows the same optimization trajectory. We evaluate ECI-PINN using PDE-Rep, a protocol spanning Wave, Reaction-Diffusion, and Kovasznay benchmarks, five equivalent residual representations, and three random seeds. ECI-PINN achieves the lowest identity, average, and worst-case relative L_2 errors among the compared methods on all three tasks, reducing the Wave worst-case error by more than 50-fold relative to Vanilla PINN. Comparisons with diagonal metrics and residual normalization support the importance of modeling the full coupled residual geometry. These results establish equation-coordinate invariance as a design principle for reliable physics-informed learning.

1 INTRODUCTION

Physics-informed neural networks (PINNs) approximate PDE solutions by minimizing governing-equation residuals together with boundary and initial-condition errors (Lagaris et al., 1998; Raissi et al., 2019; Karniadakis et al., 2021). For a coupled PDE, the same constraint can be written in several algebraically equivalent forms. These forms share a feasible solution set, but their squared-residual losses need not agree. We study how this choice of residual coordinates affects PINN training.

Let $r(z;\theta)$ denote a coupled residual vector and let $A(z)$ be pointwise invertible. The constraints $r = 0$ and $Ar = 0$ have the same solutions, whereas their standard losses are $r^\top r$ and $r^\top A^\top A r$. Anisotropic scaling and nondiagonal mixing can therefore change the loss geometry without changing the PDE constraint. This distinction is consistent with the role of algebraic representations in preconditioning and weighted least squares (Benzi, 2002; van der Sluis, 1969; Björck, 1996).

PINN optimization has been studied through gradient imbalance, neural-tangent-kernel dynamics, and ill-conditioned loss landscapes (Wang et al., 2021; 2022; Rathore et al., 2024). Adaptive weighting and specialized optimizers improve training under a specified residual formulation. Our question concerns the formulation itself: how can the PDE term remain unchanged when its residual components are rescaled or invertibly mixed?

ECI-PINN measures the minimum correction to the locally linearized PDE constraints in dimensionless jet space. The construction extends geometric-error and Sampson-type approximations for implicit constraints to coupled PDE residuals (Sampson, 1982; Taubin, 1991; Rydell et al., 2024).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

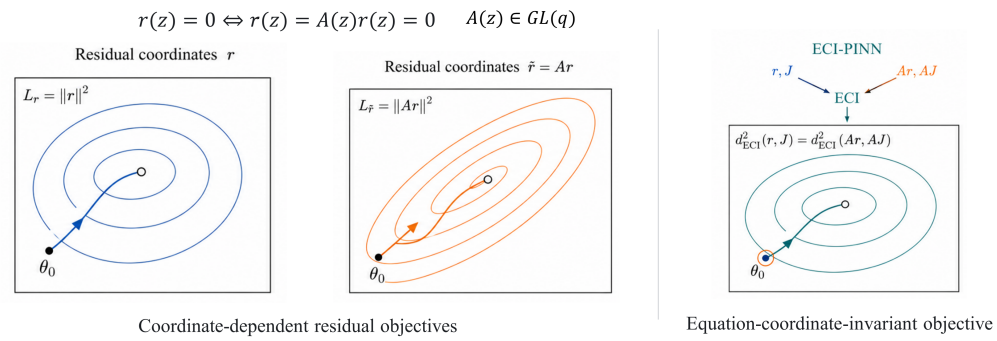

Figure 1: Conceptual illustration of residual-coordinate dependence. Equivalent coupled residual systems share a zero set but can induce different squared-residual loss geometries. ECI uses a full-Gram minimum-correction metric in a fixed jet space.

Its full Gram matrix transforms together with the residual vector, giving an invariant value and detached update under the assumptions stated in Section 4.4. The geometric quantity and the update used in training are defined separately in Section 4.3.

We evaluate residual-representation sensitivity on Wave, Reaction-Diffusion, and Kovasznay problems. The study combines a pointwise loss-value diagnostic, comparisons across five residual representations, field visualizations, and component ablations. The contributions are a formulation of equation-coordinate bias, a full-Gram minimum-correction metric with an invariant detached update, and a representation-stress evaluation of the resulting training configuration.

2 Related Work and Positioning

2.1 PINN Optimization and Ill-Conditioning

PINN training difficulties have been analyzed in terms of optimization dynamics, coupled loss terms, and PDE-induced ill-conditioning. High-frequency, multiscale, stiff, and complex-boundary problems can produce slow convergence, gradient imbalance, and localized error accumulation (Wang et al., 2021; 2022; Krishnapriyan et al., 2021; Rathore et al., 2024; Cao & Zhang, 2025; Shin et al., 2023). Loss-landscape and Hessian analyses further show that differential operators can severely worsen conditioning and impede first-order optimization (Rathore et al., 2024; Cao & Zhang, 2025; Nocedal & Wright, 2006). Adaptive weights, inverse-Dirichlet weighting, self-adaptive losses, MultiAdam, residual attention, and augmented-Lagrangian relaxation address competing objectives or constraint satisfaction (McClenny & Braga-Neto, 2023; Maddu et al., 2022; Yao et al., 2023; Anagnostopoulos et al., 2024; Son et al., 2023). Geometry-aware PINN optimization also includes energy natural gradient descent, which uses a Hessian-induced metric (Müller & Zeinhofer, 2023), and PDE-aware Kronecker-factored curvature approximations (Dangel et al., 2024). SpecMuon regulates updates along spectral modes of matrix-valued gradients (Lu et al., 2026). These approaches study optimization directions and their scaling. ECI-PINN instead constructs a residual metric whose value and detached update are unchanged under the specified changes of equation basis.

2.2 Sampling, Architecture, and Constraint Encoding

Another line of work improves PINN generalization and stability through collocation sampling, network expressivity, and constraint encoding. Importance sampling, failure-informed sampling, and residual-based adaptive sampling reduce the coverage gap between finitely many collocation points and a continuous PDE domain (Nabian et al., 2021; Gao et al., 2023; Wu et al., 2023). Fourier features and transformer architectures improve the representation of high-frequency modes and complex dynamics (Tancik et al., 2020; Zhao et al., 2024). Variational and domain-decomposition methods, including hp-VPINNs, FBPINNs, cPINNs, and XPINNs, improve scalability and local modeling (Kharazmi et al., 2021; Moseley et al., 2023; Jagtap et al., 2020; Jagtap & Karniadakis, 2020), while distance-function constructions impose boundary conditions exactly (Sukumar & Srivastava,

2022). ECI-PINN acts on the coupled PDE residual block, whereas these methods modify sampling, function representation, domain decomposition, or boundary-condition encoding.

2.3 Invariance and Geometric Residual Metrics

Invariance has a long history in differential-equation learning and numerical analysis. Invariant PINNs constrain predictions under transformations of physical variables or Lie-group actions (Arora et al., 2024), whereas geometric integrators preserve dynamical structure under discretization (McLachlan & Quispel, 2006). In implicit geometric estimation, algebraic residuals generally do not equal distance to a constraint manifold, motivating Sampson corrections and related geometric-error approximations (Sampson, 1982; Taubin, 1991; Kanatani, 1996; Rydell et al., 2024; Hartley & Zisserman, 2004). Here, the transformed object is the equation residual basis. The full-Gram construction preserves the local correction value and the detached update for a fixed jet metric and the invertible transformation class specified below.

3 Problem Formulation: Equivalent Equation Coordinates

3.1 Coupled PINN Residual Learning

Consider a coupled PDE on a space–time domain $\Omega \subset \mathbb{R}^d$. The unknown physical field is represented by $u : \Omega \to \mathbb{R}^m$ and governed by

$$F(z, u, \partial u, \ldots, \partial^k u) = 0, \tag{1}$$

where z denotes the collection of spatial and temporal coordinates. A PINN $u_\theta(z)$ approximates the continuous solution and evaluates the governing residual through automatic differentiation (Raissi et al., 2019; Karniadakis et al., 2021; Cuomo et al., 2022; Shin et al., 2023). For q coupled equations, define

$$r(z;\theta) = [r_1(z;\theta), \ldots, r_q(z;\theta)]^\top \in \mathbb{R}^q. \tag{2}$$

Given interior collocation points $\{z_i\}_{i=1}^{N_r}$, the standard objective is

$$\mathcal{L}_{\mathrm{PINN}}(\theta) = \frac{1}{N_r} \sum_{i=1}^{N_r} \|r(z_i;\theta)\|_2^2 + \lambda_b \mathcal{L}_b(\theta) + \lambda_0 \mathcal{L}_0(\theta). \tag{3}$$

The squared residual selects an inner product in equation space. Boundary and initial-condition terms remain in the physical variables; the coordinate transformation acts on the governing-equation residual block.

3.2 Equivalent Coordinates and Equation-Coordinate Bias

Definition 1 (Equivalent equation coordinates). For each $z \in \Omega$, let $A(z) \in \mathrm{GL}(q)$ be independent of both the network parameters and the current jet state. Residual systems r and $\widetilde{r}$ are equivalent equation coordinates when $\widetilde{r}(z;\theta) = A(z)r(z;\theta)$. Invertibility gives $r = 0 \Leftrightarrow \widetilde{r} = 0$. Thus, permutations, nonzero row scaling, nondiagonal invertible mixing, and position-dependent invertible basis changes preserve the physical constraint while changing its residual coordinates.

Proposition 1 (Coordinate dependence of the standard objective). In general,

$$\|\widetilde{r}\|_2^2 = r^\top A^\top A r, \tag{4}$$

which equals $\|r\|_2^2$ only for isometric transformations or special residual subspaces. If $J_r = \partial r/\partial\theta$, then

$$\nabla_\theta \|\widetilde{r}\|_2^2 = 2J_r^\top A^\top A r, \tag{5}$$

instead of $2J_r^\top r$. Physically equivalent residual systems can therefore produce different losses, parameter gradients, and deterministic optimization trajectories. We call this mismatch *equation-coordinate bias*. Componentwise normalization cannot remove it in general because diagonal scaling does not account for nondiagonal equation mixing. Moreover, dynamically estimated weights are themselves representation-dependent (Wang et al., 2021; 2022; McClenny & Braga-Neto, 2023; Maddu et al., 2022; Yao et al., 2023; Anagnostopoulos et al., 2024).

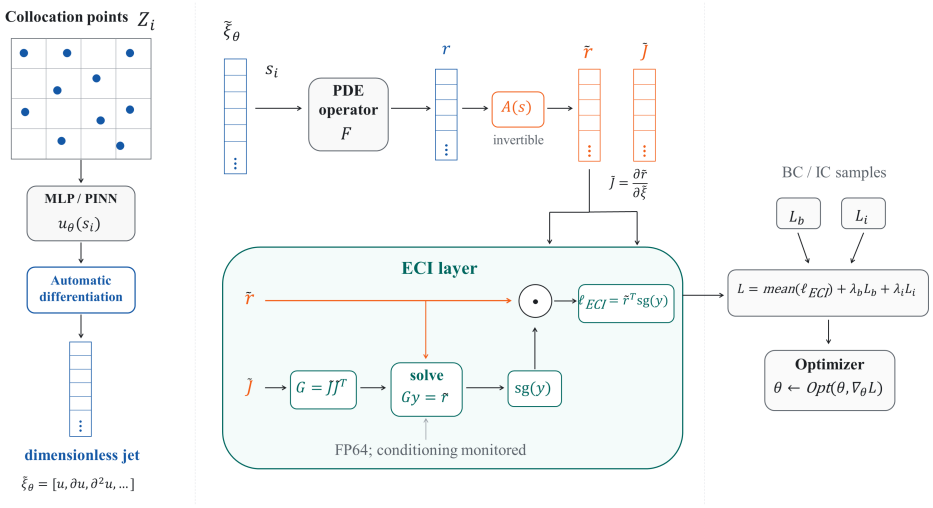

Figure 2: Core ECI computation at a collocation point. The network outputs and required derivatives form the dimensionless jet. The residual Jacobian defines a full Gram system, whose detached solution is used to evaluate the local correction term. Boundary and initial-condition losses are added in the physical variables.

3.3 Scope and Evaluation Principle

We consider pointwise invertible changes of the governing-equation residual basis. The transformation may depend on position, while the physical variables, boundary and initial conditions, and dimensionless jet metric remain fixed. The invariance results require full-row-rank residual Jacobians and transformations independent of the network parameters and jet state. The evaluation applies diagonal scaling, constant dense mixing, position-dependent mixing, and anisotropic mixing to the same canonical residual systems.

4 ECI-PINN: An Equation-Coordinate-Invariant Residual Metric

4.1 Overview

ECI-PINN evaluates a local correction in dimensionless jet space. At each collocation point, the coupled PDE is linearized with respect to the jet, and the minimum squared correction defines the residual metric shown in Figure 2.

4.2 Dimensionless Jet and Minimum Correction

Placing states and derivatives of different orders in one Euclidean space introduces unit dependence. Let $\xi_\theta(z)$ collect the network outputs and all derivatives required by the PDE. A fixed diagonal scale S_ξ defines the dimensionless jet $\widehat{\xi}_\theta = S_\xi^{-1}\xi_\theta$. Scales are determined before training based on domain lengths, solution amplitudes, and derivative orders; they are shared by all residual representations and methods. For canonical residual $r = F(z, \widehat{\xi}_\theta)$, define

$$J(z;\theta) = \frac{\partial r(z;\theta)}{\partial \widehat{\xi}_\theta(z)}. \tag{6}$$

The local linearization is $r + J\delta\widehat{\xi} \approx 0$. ECI measures the smallest feasible jet correction:

$$d^2_{\mathrm{ECI}}(r, J) = \min_{\delta\widehat{\xi}} \|\delta\widehat{\xi}\|_2^2 \quad \text{s.t.} \quad r + J\delta\widehat{\xi} = 0. \tag{7}$$

When J has full row rank,

$$d^2_{\mathrm{ECI}}(r, J) = r^\top (JJ^\top)^{-1} r. \tag{8}$$

The Lagrangian is

$$\mathcal{Q}(\delta\widehat{\xi}, \lambda) = \delta\widehat{\xi}^\top \delta\widehat{\xi} + 2\lambda^\top (r + J\delta\widehat{\xi}). \tag{9}$$

The first-order conditions give $\delta\widehat{\xi} = -J^\top\lambda$ and $JJ^\top\lambda = r$. Hence,

$$\delta\widehat{\xi}^\star = -J^\top (JJ^\top)^{-1} r, \qquad \|\delta\widehat{\xi}^\star\|_2^2 = r^\top (JJ^\top)^{-1} r. \tag{10}$$

The optimization variable is a jet correction at one collocation point. The affine constraint and the selected jet metric determine its squared norm. The complete multiplier derivation is given in Appendix A. The scale matrix S_ξ fixes the jet metric and is held constant across residual representations. The invariance result holds for each fixed choice of this metric. For $q = 1$, the correction value reduces to $r^2/\|J\|_2^2$, connecting it to row-wise Jacobian normalization (Bento et al., 2026). ECI-Diag retains the diagonal of $JJ^\top$, whereas ECI-PINN uses the full Gram matrix, including the inner products between different residual-Jacobian rows.

4.3 Detached ECI update

At each residual evaluation, ECI-PINN solves $Gy = r$ and treats the complete solution vector as constant during backpropagation:

$$G = JJ^\top, \qquad y = \mathrm{stopgrad}(G^{-1}r), \qquad \ell_{\mathrm{ECI}} = r^\top y. \tag{11}$$

At an evaluation state θ_t, write $y_t = G(\theta_t)^{-1} r(\theta_t)$. Backpropagation differentiates the local surrogate $s_t(\theta) = r(\theta)^\top y_t$. Its value at θ_t equals $r(\theta_t)^\top G(\theta_t)^{-1} r(\theta_t)$, and its derivative is $g_t = (\partial_\theta r(\theta_t))^\top y_t$. The distinction can be seen by differentiating the geometric value $h(\theta) = r(\theta)^\top G(\theta)^{-1} r(\theta)$ in the usual way. Writing $y = G^{-1}r$, a parameter coordinate θ_a gives

$$\partial_a h = 2(\partial_a r)^\top y - y^\top (\partial_a G) y. \tag{12}$$

The local surrogate instead returns $(g_t)_a = (\partial_a r)^\top y_t$, with the complete solve vector fixed during that evaluation. The minimum-correction value provides the geometric quantity, and the refreshed surrogate specifies the update analyzed in Theorem 2. For a batch of interior collocation points, the PDE term is averaged and combined with the boundary and initial-condition losses:

$$\mathcal{L}_{\mathrm{ECI}} = \frac{1}{N_r} \sum_i \ell_{\mathrm{ECI}}(z_i; \theta) + \lambda_b \mathcal{L}_b + \lambda_0 \mathcal{L}_0. \tag{13}$$

Detaching the solve removes the backward path through y. The network forward pass and automatic differentiation use FP32, and the $q \times q$ Gram system is solved in FP64 without explicitly forming its inverse.

4.4 Value and detached-update invariance

Fix the dimensionless jet metric and assume that J has full row rank at each evaluation under consideration. Let $\widetilde{r} = Ar$, where $A(z)$ is pointwise invertible and independent of the network parameters and jet state. Then $\widetilde{J} = AJ$ and $\widetilde{G} = AGA^\top$.

Theorem 1 (Value invariance). If J has full row rank,

$$\widetilde{r}^\top \widetilde{G}^{-1} \widetilde{r} = r^\top G^{-1} r, \tag{14}$$

because $\widetilde{G}^{-1} = A^{-\top} G^{-1} A^{-1}$. Hence the pointwise ECI value depends only on the linearized constraint manifold and fixed jet metric.

Theorem 2 (Detached-update invariance). Let $y = \mathrm{stopgrad}(G^{-1}r)$. Under the transformed coordinates, $\widetilde{y} = A^{-\top} y$ and $\partial\widetilde{r}/\partial\theta = A\,\partial r/\partial\theta$. Therefore,

$$\nabla_\theta(\widetilde{r}^\top \widetilde{y}) = \left(\frac{\partial r}{\partial \theta}\right)^\top y = \nabla_\theta(r^\top y). \tag{15}$$

Equivalent residual representations thus produce the same detached update at every parameter state.

Corollary 1 (Deterministic trajectory invariance). In exact arithmetic, consider two runs of the unregularized ECI rule with identical initialization, batches, random state, boundary and initial losses, and optimizer state. If their deterministic update maps use identical hyperparameters and the same invariant inputs at every evaluation, Theorems 1 and 2 give identical parameter sequences by induction.

4.5 TRAINING IMPLEMENTATION

The training configuration combines the full-Gram ECI rule with numerical stabilization and optimizer refinement. Appendix C.2 describes representation whitening under high-severity transformations, a warm-up anchor, safe switching, an optimizer portfolio, and checkpoint selection. The Kovasznay component comparison in Table S4 also includes a variant without L-BFGS refinement. These comparisons evaluate the stabilized configuration built around the unregularized rule analyzed in Section 4.4.

The recorded Gram diagnostics are the minimum eigenvalue, condition number, solve status, and NaN/Inf flags. Each residual block contains two equations for Wave and Reaction-Diffusion or three for Kovasznay, so the corresponding Gram systems are 2×2 or 3×3. The systems are solved in FP64 without explicitly forming an inverse. Completion counts and mean runtimes are given in Table S3.

Algorithm 1: Core ECI residual evaluation.

Input: network u_θ; collocation batch; fixed jet scale S_ξ; canonical residual F; pointwise transformation $A(z)$.
1. Evaluate the network outputs and the derivatives required by the PDE.
2. Construct the dimensionless jet $\widehat{\xi} = S_\xi^{-1}\xi$.
3. Evaluate $r = F(z, \widehat{\xi})$ and $\widetilde{r} = A(z)r$.
4. Compute $\widetilde{J} = \partial\widetilde{r}/\partial\widehat{\xi}$.
5. Form $\widetilde{G} = \widetilde{J}\widetilde{J}^\top$ and solve $\widetilde{G}y = \widetilde{r}$ in FP64.
6. Evaluate $\ell = \widetilde{r}^\top \operatorname{stopgrad}(y)$, detaching the complete solve vector.
7. Average ℓ over the batch and add the boundary and initial-condition losses.
Output: the scalar used for the current residual evaluation and backward pass.

Section 4.5 specifies the surrounding training components; this algorithm gives the core residual evaluation.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP AND EVALUATION PROTOCOL

We evaluate three coupled PDE systems. Wave is posed on $(x,t) \in [0,1]^2$ and predicts displacement u and auxiliary velocity w, with residuals $u_t - w$ and $w_t - u_{xx}$. The Wave reference is $u(x,t) = \sin(\pi x)\cos(\pi t)$ with $w(x,t) = -\pi\sin(\pi x)\sin(\pi t)$, which also determines the boundary and initial values. Reaction-Diffusion is posed on $(x,y,t) \in [0,1]^3$ and predicts coupled fields u and v, with diffusion coefficients $D_u = 0.02$ and $D_v = 0.01$. Its manufactured forcing terms are obtained from the analytical fields specified in Appendix B.2. Kovasznay flow is steady on $(x,y) \in [-0.5,1]\times[-0.5,1.5]$ at $\mathrm{Re} = 40$; the outputs are u, v, and p, and the residual block contains two momentum equations and continuity. The analytical solutions and boundary and initial values are given in Appendices B.1–B.3.

Each task uses a four-layer MLP of width 64, 512 interior collocation points, 128 boundary/initial points, and seeds 0, 1, and 2. Starting from the canonical residual, the representation protocol comprises the identity R_0, diagonal scaling R_1, constant dense mixing R_2, position-dependent mixing R_3, and anisotropic mixing R_4. The transformation acts on the governing-equation residual block, while the predicted physical fields retain their canonical meaning.

For Reaction-Diffusion, the two canonical residuals are the diffusion–reaction operators with manufactured forcings from Appendix B.2. For Kovasznay flow, the residual block contains the two steady momentum equations and continuity, while the network outputs velocity and pressure. All methods are evaluated on the same canonical physical fields and common sampling protocol. Thus, the representation stress test changes the equation-coordinate block used for training, not the target solution, the reported relative L_2 error, or the field panels in Figure 5.

Baselines. Table 1 compares Vanilla PINN (Raissi et al., 2019), MultiAdam (Yao et al., 2023), L-BFGS PINN, AdamW plus L-BFGS, NAdam plus L-BFGS, residual-scale normalization (RMSNorm), and ECI-PINN. ECI-Diag provides the full-versus-diagonal Gram comparison in Table S2. Figure 5 shows seed-0 predictions under R_4.

Table 1: Relative L_2 errors for the five residual representations. Identity, Average, and Worst are defined in Section 5.1. Lower values indicate smaller errors.

Task	Method	Identity ↓	Average ↓	Worst ↓
Wave	Vanilla PINN	5.71e-02	2.12e-01	9.79e-01
	MultiAdam	6.53e-01	8.03e-01	9.66e-01
	L-BFGS PINN	4.00e-02	1.17e-01	9.92e-01
	AdamW + L-BFGS†	1.38e-02	—	—
	NAdam + L-BFGS†	1.44e-02	—	—
	RMSNorm	1.01e+00	1.01e+00	1.06e+00
	ECI-PINN	**1.32e-02**	**1.43e-02**	**1.92e-02**
Reaction-Diffusion	Vanilla PINN	4.42e-02	4.53e-02	2.26e-01
	MultiAdam	3.72e-02	5.42e-02	1.20e-01
	L-BFGS PINN	2.18e-02	2.91e-02	1.64e-01
	AdamW + L-BFGS†	4.42e-03	—	—
	NAdam + L-BFGS	4.02e-03	9.16e-03	1.26e-01
	RMSNorm	2.50e-01	2.52e-01	3.26e-01
	ECI-PINN	**3.48e-03**	**5.20e-03**	**2.35e-02**
Kovasznay	Vanilla PINN	4.36e-02	5.30e-02	1.64e-01
	MultiAdam	6.72e-01	6.05e-01	6.87e-01
	L-BFGS PINN	4.47e-03	8.14e-03	3.22e-02
	AdamW + L-BFGS	2.01e-03	7.72e-03	4.40e-02
	NAdam + L-BFGS	1.17e-03	5.63e-03	4.18e-02
	RMSNorm	6.95e-01	6.95e-01	6.95e-01
	ECI-PINN	**6.07e-04**	**3.74e-03**	**1.92e-02**

† The Average/Worst aggregates for these entries are withheld because they are inconsistent with the stated 15-run aggregation.

Table 2: Summary of the loss-value diagnostic, metric comparisons, numerical reliability, and runtime.

Evidence	Key result
Loss-value diagnostic	Median absolute loss change under R_1–R_4: Vanilla $2.93\text{e}{+00}$–$4.15\text{e}{+01}$; ECI $\leq 1.11\text{e}{-16}$.
Full versus diagonal metric	Within-table comparison of ECI-PINN and ECI-Diag: Table S2.
Residual normalization	RMSNorm comparisons: Table 1 and Table S2.
Common-grid summary	Table S2: 15 task–method cells with 9 runs each (135 runs).
Numerical reliability	Table S3: Vanilla, MultiAdam, and ECI-PINN complete 45/45 runs each; no NaN/Inf values; no ECI Gram-solve failures.
Runtime	Table S3 mean runtime: ECI 114.17 s; Vanilla 54.73 s; MultiAdam 54.35 s.

Metrics. Every model is evaluated in the canonical physical representation, regardless of its training residual basis. Let $e(m, k, s)$ denote the relative L_2 error for method m, representation R_k, and seed s. Identity summarizes the three seed results under R_0. Average is the arithmetic mean over R_0–R_4 and seeds 0, 1, and 2; Worst is the maximum over the same 15 representation–seed pairs. Solution error and variation across residual representations are considered together in the comparisons.

5.2 Loss-value invariance diagnostic

Figure 3 compares pointwise loss values before and after equivalent residual transformations. The median absolute change in the ECI value is at most 1.11×10^{-16} across R_1–R_4, whereas the standard squared-residual loss changes substantially. The diagnostic examines the value identity in Theorem 1. Detached-update and deterministic trajectory invariance follow analytically under the assumptions in Section 4.4.

5.3 Main Results and Worst-Representation Robustness

Table 1 compares solution errors across the five residual representations. The Wave Worst entries are 1.92×10^{-2} for ECI-PINN and 9.79×10^{-1} for Vanilla PINN. On Reaction-Diffusion, the corresponding entries are 2.35×10^{-2} and 2.26×10^{-1}; MultiAdam has a Worst entry of 1.20×10^{-1}. On Kovasznay, the ECI-PINN and L-BFGS PINN entries are 1.92×10^{-2} and 3.22×10^{-2}, respectively. The separation is therefore larger for the Wave comparison than for Kovasznay against L-BFGS PINN.

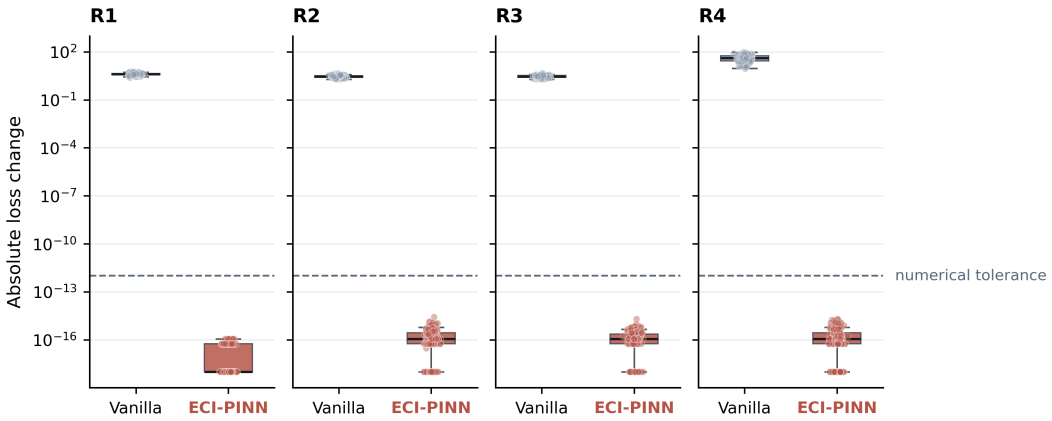

Figure 3: Loss-value invariance diagnostic. The panels compare absolute pointwise loss changes under residual transformations R_1–R_4 for the standard squared-residual loss and the ECI value.

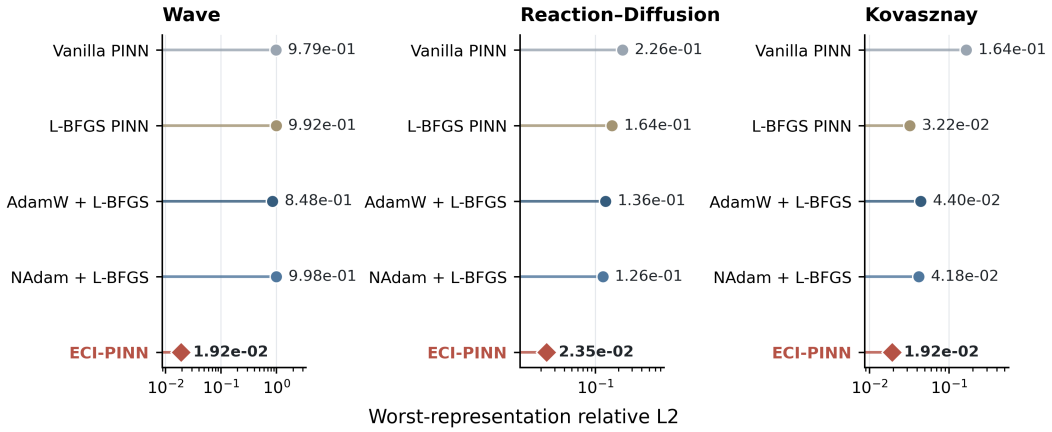

Figure 4: Worst-representation relative L_2 errors for ECI-PINN, Vanilla PINN, and the L-BFGS-based baselines on three PDE tasks. The three task–method entries flagged in Table 1 are excluded from quantitative interpretation.

Figure 4 shows ECI-PINN alongside Vanilla PINN and the L-BFGS-based configurations. The complete method comparison, including MultiAdam and RMSNorm, is given in Table 1. Table S4 examines optimizer refinement and stabilization components within the Kovasznay configuration.

5.4 Field recovery under R_4

Figure 5 shows seed-0 predictions under R_4. For Wave, ECI-PINN reproduces the positive and negative lobes of the reference displacement field, whereas the Vanilla PINN and MultiAdam predictions remain close to zero. Their absolute-error maps consequently retain much of the reference field's structure. The displayed relative L_2 errors are 0.961, 0.966, and 4.85×10^{-3} for Vanilla PINN, MultiAdam, and ECI-PINN, respectively. For Reaction-Diffusion, the u slice at $t = 0.5$ is closer to the reference under ECI-PINN, with smaller visible error regions; the corresponding displayed errors are 0.330, 0.161, and 2.84×10^{-3}.

5.5 Ablation, numerical reliability, and cost

Table 2 collects the metric comparisons and numerical results. Within Table S2, the full metric has lower Mean errors than ECI-Diag on Wave (0.0327 versus 0.158), Reaction-Diffusion (0.0198

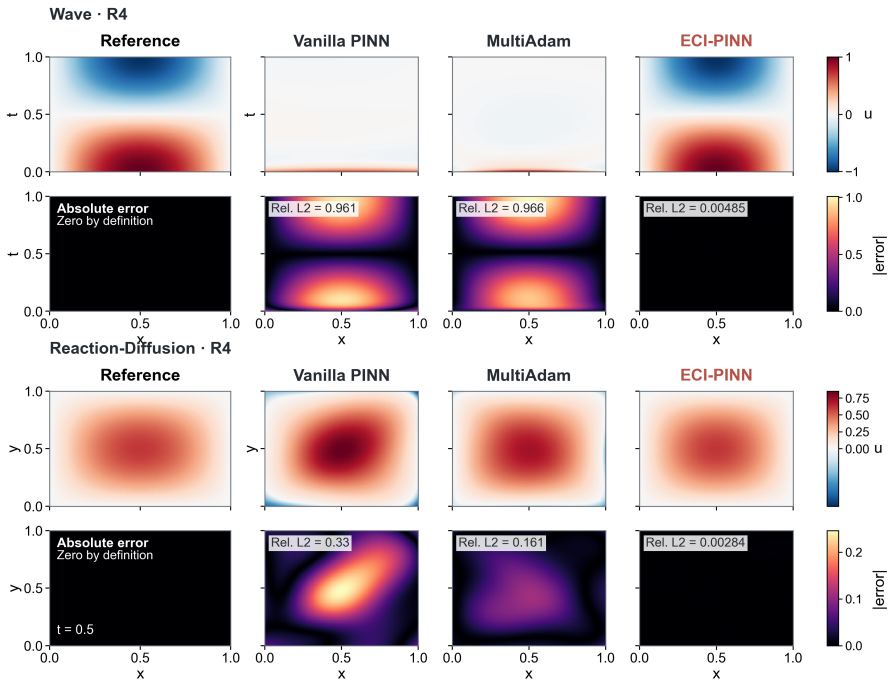

Figure 5: Seed-0 field comparisons under R_4. The upper panels show the Wave displacement field, and the lower panels show the Reaction-Diffusion u field at $t = 0.5$. Reference fields, predictions, and absolute-error maps are shown for Vanilla PINN, MultiAdam, and ECI-PINN.

versus 0.0412), and Kovasznay (0.0136 versus 0.0578). Each task–method cell contains nine runs across three seeds and three shared representations.

The Kovasznay component results in Table S4 show different error patterns across the named removals. Without L-BFGS refinement, the Identity entry is 1.23×10^{-2}, compared with 6.07×10^{-4} for the complete configuration. Removing safe switching or whitening gives Worst entries of 5.29×10^{-2} and 4.22×10^{-2}, respectively, compared with 1.92×10^{-2}. These comparisons describe the effects of the training components alongside the full-versus-diagonal metric comparison in Table S2.

Table S3 records 45 completed runs for each of Vanilla PINN, MultiAdam, and ECI-PINN, spanning three tasks, five representations, and three seeds per method. No NaN/Inf values are recorded, and the ECI runs have no Gram-solve failures. Mean runtimes are 54.73 s for Vanilla PINN, 54.35 s for MultiAdam, and 114.17 s for ECI-PINN.

6 Conclusion

ECI-PINN evaluates the minimum pointwise correction to locally linearized PDE constraints in a fixed dimensionless jet metric. For full-row-rank residual Jacobians, the unregularized full-Gram value and detached update are invariant to pointwise invertible transformations independent of the network parameters and jet state. The experiments examine three coupled PDE systems through loss-value diagnostics, residual-representation comparisons, field recovery, and component ablations. The field comparisons show recovery of the reference structure under strong residual mixing. The Kovasznay ablations also show that removing refinement or stabilization components increases selected Identity, Average, or Worst errors. Removing the optimizer portfolio gives Average and Worst entries of 2.15×10^{-2} and 9.60×10^{-2}, compared with 3.74×10^{-3} and 1.92×10^{-2} for the complete configuration. In the common-grid comparisons, the full metric has lower Mean error than ECI-Diag on all three tasks: 0.0327 versus 0.158 on Wave, 0.0198 versus 0.0412 on Reaction-Diffusion, and 0.0136 versus 0.0578 on Kovasznay.

AI USE STATEMENT

We used Codex to assist with manuscript formatting and LaTeX typesetting. The authors are responsible for the manuscript.

REPRODUCIBILITY STATEMENT

The minimum-correction construction and detached-update interpretation are given in Sections 4.2 and 4.3. Appendix A provides the derivations and proofs; Appendix B specifies the analytical PDE systems and residual-representation categories; and Appendix C.1 records the documented protocol settings. Appendices D–F contain the additional comparisons, completion counts, runtimes, and component results. The accompanying anonymous code package documents its supported evaluation and aggregation entry points, identifies missing checkpoints or external resources, and does not claim a complete end-to-end rerun of every reported experiment.

REFERENCES

Stylianos J. Anagnostopoulos, Juan Diego Toscano, Nikolaos Stergiopulos, and George Em Karniadakis. Residual-based attention in physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 421:116805, 2024. doi: 10.1016/j.cma.2024.116805.

Suyash Arora, Alexander Bihlo, and Francis Valiquette. Invariant physics-informed neural networks for ordinary differential equations. *Journal of Machine Learning Research*, 25(252):1–46, 2024.

M. P. Bento, H. B. Câmara, J. R. Rocha, and J. F. Seabra. Solving stiff dark matter equations via jacobian normalization with physics-informed neural networks. *arXiv preprint arXiv:2602.21988*, 2026.

Michele Benzi. Preconditioning techniques for large linear systems: A survey. *Journal of Computational Physics*, 182(2):418–477, 2002. doi: 10.1006/jcph.2002.7176.

Åke Björck. *Numerical Methods for Least Squares Problems*. SIAM, 1996. doi: 10.1137/1.9781611971484.

Wen Cao and Wei Zhang. An analysis and solution of ill-conditioning in physics-informed neural networks. *Journal of Computational Physics*, 520:113494, 2025. doi: 10.1016/j.jcp.2024.113494.

Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics-informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92:88, 2022. doi: 10.1007/s10915-022-01939-z.

Felix Dangel, Johannes Müller, and Marius Zeinhofer. Kronecker-factored approximate curvature for physics-informed neural networks. In *Advances in Neural Information Processing Systems*, 2024. URL https://arxiv.org/abs/2405.15603.

Zixue Gao, Liang Yan, and Tao Zhou. Failure-informed adaptive sampling for PINNs. *SIAM Journal on Scientific Computing*, 45(4):A1971–A1994, 2023. doi: 10.1137/22M1527763.

Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition, 2004.

Ameya D. Jagtap and George Em Karniadakis. Extended physics-informed neural networks: A generalized space–time domain decomposition framework for nonlinear partial differential equations. *Communications in Computational Physics*, 28(5):2002–2041, 2020. doi: 10.4208/cicp.oa-2020-0164.

Ameya D. Jagtap, Ehsan Kharazmi, and George Em Karniadakis. Conservative physics-informed neural networks on discrete domains for conservation laws. *Computer Methods in Applied Mechanics and Engineering*, 365:113028, 2020. doi: 10.1016/j.cma.2020.113028.

Kenichi Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier, 1996.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021. doi: 10.1038/s42254-021-00314-5.

Ehsan Kharazmi, Zhongqiang Zhang, and George Em Karniadakis. hp-VPINNs: Variational physics-informed neural networks with domain decomposition. *Computer Methods in Applied Mechanics and Engineering*, 374:113547, 2021. doi: 10.1016/j.cma.2020.113547.

Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, Robert M. Kirby, and Michael W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems*, volume 34, pp. 26548–26560, 2021.

Isaac E. Lagaris, Aristidis Likas, and Dimitrios I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, 9(5):987–1000, 1998. doi: 10.1109/72.712178.

Binghang Lu, Jiahao Zhang, and Guang Lin. Muon with spectral guidance: Efficient optimization for scientific machine learning, 2026. URL https://arxiv.org/abs/2602.16167.

Suryanarayana Maddu, Dominik Sturm, Christian L. M"uller, and Ivo F. Sbalzarini. Inverse dirichlet weighting enables reliable training of physics informed neural networks. *Machine Learning: Science and Technology*, 3(1):015026, 2022. doi: 10.1088/2632-2153/ac3712.

Levi D. McClenny and Ulisses M. Braga-Neto. Self-adaptive physics-informed neural networks. *Journal of Computational Physics*, 474:111722, 2023. doi: 10.1016/j.jcp.2022.111722.

Robert I. McLachlan and G. R. W. Quispel. Geometric integrators for ODEs. *Journal of Physics A: Mathematical and General*, 39(19):5251–5285, 2006. doi: 10.1088/0305-4470/39/19/S01.

Ben Moseley, Andrew Markham, and Tarje Nissen-Meyer. Finite basis physics-informed neural networks: A scalable domain decomposition approach for solving differential equations. *Advances in Computational Mathematics*, 49:62, 2023. doi: 10.1007/s10444-023-10065-9.

Johannes Müller and Marius Zeinhofer. Achieving high accuracy with PINNs via energy natural gradient descent. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25471–25485. PMLR, 2023. URL https://proceedings.mlr.press/v202/muller23b.html.

Mohammad Amin Nabian, Ryan J. Gladstone, and Hadi Meidani. Efficient training of physics-informed neural networks via importance sampling. *Computer-Aided Civil and Infrastructure Engineering*, 36(8):962–977, 2021. doi: 10.1111/mice.12685.

Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2 edition, 2006. doi: 10.1007/978-0-387-40065-5.

Maziar Raissi, Paris Perdikaris, and George Em 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. doi: 10.1016/j.jcp.2018.10.045.

Pratik Rathore, Weimu Lei, Zachary Frangella, Lu Lu, and Madeleine Udell. Challenges in training PINNs: A loss landscape perspective. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, 2024.

Felix Rydell, Alberto Torres, and Viktor Larsson. Revisiting sampson approximations for geometric estimation problems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Paul D. Sampson. Fitting conic sections to very scattered data: An iterative refinement of the bookstein algorithm. *Computer Graphics and Image Processing*, 18(1):97–108, 1982. doi: 10.1016/0146-664X(82)90101-0.

Yeonjong Shin, Zhongqiang Zhang, and George Em Karniadakis. Error estimates of residual minimization using neural networks for linear PDEs. *Journal of Machine Learning for Modeling and Computing*, 4(4):73–101, 2023. doi: 10.1615/JMachLearnModelComput.2023050411.

Hwijae Son, Sang Woo Cho, and Hyung Ju Hwang. Enhanced physics-informed neural networks with augmented lagrangian relaxation method. *Neurocomputing*, 548:126424, 2023. doi: 10.1016/j.neucom.2023.126424.

N. Sukumar and Ankit Srivastava. Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks. *Computer Methods in Applied Mechanics and Engineering*, 389:114333, 2022. doi: 10.1016/j.cma.2021.114333.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7537–7547, 2020.

Gabriel Taubin. Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11):1115–1138, 1991. doi: 10.1109/34.103273.

Abraham van der Sluis. Condition numbers and equilibration of matrices. *Numerische Mathematik*, 14:14–23, 1969. doi: 10.1007/BF02165096.

Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021. doi: 10.1137/20M1318043.

Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why PINNs fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2022. doi: 10.1016/j.jcp.2021.110768.

Chenxi Wu, Min Zhu, Qinyang Tan, Yadhu Kartha, and Lu Lu. A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 403:115671, 2023. doi: 10.1016/j.cma.2022.115671.

Jiawei Yao, Chen Su, Zhongkai Hao, Songming Liu, Huan Su, and Jun Zhu. MultiAdam: Parameter-wise scale-invariant optimizer for multiscale training of physics-informed neural networks. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 39702–39721, 2023.

Zhiqiang Zhao, Xueying Ding, and B. Aditya Prakash. PINNsFormer: A transformer-based framework for physics-informed neural networks. In *International Conference on Learning Representations*, 2024.

APPENDIX

A THEORY AND PROOFS

A.1 CLOSED FORM OF THE MINIMUM-CORRECTION METRIC

At a fixed sample z, omit the explicit coordinate dependence. Let $r \in \mathbb{R}^q$ be the current residual vector and $J \in \mathbb{R}^{q\times p}$ the Jacobian of the residual with respect to the dimensionless jet $\widehat{\xi}$. Assume that J has full row rank. Consider

$$\min_{\delta\widehat{\xi}\in\mathbb{R}^p} \|\delta\widehat{\xi}\|_2^2 \quad \text{s.t.} \quad r + J\delta\widehat{\xi} = 0. \tag{16}$$

With Lagrange multiplier $\lambda \in \mathbb{R}^q$, the Lagrangian is

$$Q(\delta\widehat{\xi}, \lambda) = \delta\widehat{\xi}^\top\delta\widehat{\xi} + 2\lambda^\top(r + J\delta\widehat{\xi}). \tag{17}$$

First-order optimality gives $\delta\widehat{\xi}^* = -J^\top\lambda$ and $r - JJ^\top\lambda = 0$. Since J has full row rank, $G = JJ^\top$ is positive definite, and

$$\lambda = G^{-1}r, \qquad \delta\widehat{\xi}^* = -J^\top(JJ^\top)^{-1}r, \tag{18}$$

$$\|\delta\widehat{\xi}^*\|_2^2 = r^\top(JJ^\top)^{-1}r. \tag{19}$$

The value is the minimum squared correction to the affine constraint $r + J\delta\widehat{\xi} = 0$ at a fixed sample. It is defined in the selected dimensionless jet metric, with the correction minimized over the jet coordinates at that sample.

A.2 LOSS-VALUE INVARIANCE

Let $\widetilde{r} = Ar$, where $A(z) \in GL(q)$ is pointwise invertible and independent of the network parameters and jet state. The transformed Jacobian and Gram matrix are $\widetilde{J} = AJ$ and $\widetilde{G} = AGA^\top$. Therefore, $\widetilde{G}^{-1} = A^{-\top}G^{-1}A^{-1}$ and

$$\widetilde{r}^\top\widetilde{G}^{-1}\widetilde{r} = r^\top A^\top A^{-\top}G^{-1}A^{-1}Ar = r^\top G^{-1}r. \tag{20}$$

The equality holds at every position and parameter state under a full-row-rank residual Jacobian, a fixed jet metric, an exact solve, and transformations independent of the parameters and jet state.

A.3 DETACHED-UPDATE AND TRAJECTORY INVARIANCE

Let $y = \mathrm{stopgrad}(G^{-1}r)$ and $\ell = r^\top y$. In transformed coordinates, $\widetilde{y} = \mathrm{stopgrad}(\widetilde{G}^{-1}\widetilde{r}) = A^{-\top}y$. Since A is independent of θ, $\partial\widetilde{r}/\partial\theta = A\,\partial r/\partial\theta$, yielding

$$\nabla_\theta\widetilde{\ell} = \left(\frac{\partial\widetilde{r}}{\partial\theta}\right)^\top\widetilde{y} = \left(\frac{\partial r}{\partial\theta}\right)^\top A^\top A^{-\top}y = \nabla_\theta\ell. \tag{21}$$

If two runs additionally share the initialization, batch order, random state, boundary and initial losses, optimizer state, and deterministic update map, induction gives identical parameter sequences for the detached update under the same assumptions.

The detached residual term in Section 4.3 of the main paper returns the derivative of $s_t(\theta) = r(\theta)^\top y_t$, with the complete solve vector y_t held fixed during that evaluation.

Define $h(\theta) = r(\theta)^\top G(\theta)^{-1}r(\theta)$ and let $y = G^{-1}r$. For a parameter coordinate θ_a, ordinary differentiation gives

$$\partial_a h = 2(\partial_a r)^\top y - y^\top(\partial_a G)y.$$

The local surrogate has component $(g_t)_a = (\partial_a r)^\top y_t$. The invariance proof concerns this refreshed surrogate update and establishes equality of the surrogate updates across equivalent residual bases.

A.4 Conditions and Degenerate Cases

The closed form requires a full-row-rank residual Jacobian, making $JJ^\top$ invertible. For a pointwise invertible transformation that is independent of the network parameters and jet state, the relation $\widetilde{J} = AJ$ gives the transformed Gram matrix used in the proofs. State-dependent transformations introduce additional chain-rule terms. The recorded minimum-singular-value and condition-number diagnostics evaluate R_0–R_4 at the domain sample points.

B PDE-Rep Benchmarks and Equivalent Representations

B.1 Wave

The Wave benchmark is defined on $(x, t) \in [0, 1]^2$. The network predicts displacement u and auxiliary velocity w. The canonical first-order coupled system is

$$r_1 = u_t - w, \qquad r_2 = w_t - u_{xx}, \tag{22}$$

with analytical solution

$$u(x, t) = \sin(\pi x)\cos(\pi t), \tag{23}$$

$$w(x, t) = -\pi \sin(\pi x)\sin(\pi t). \tag{24}$$

The analytical solution gives $u(0, t) = u(1, t) = w(0, t) = w(1, t) = 0$. At $t = 0$, the values are $u(x, 0) = \sin(\pi x)$ and $w(x, 0) = 0$.

B.2 Reaction-Diffusion

The Reaction-Diffusion benchmark is defined on $(x, y, t) \in [0, 1]^3$ and predicts coupled states u and v. With $D_u = 0.02$ and $D_v = 0.01$, the canonical operators are

$$r_1 = u_t - D_u \Delta u - (u - u^3 - 0.5v) - f_u, \tag{25}$$

$$r_2 = v_t - D_v \Delta v - (0.75u - v) - f_v. \tag{26}$$

The manufactured forcing terms f_u and f_v are generated exactly from

$$u = \sin(\pi x)\sin(\pi y)e^{-t}, \qquad v = \cos(\pi x)\sin(\pi y)e^{-0.5t}. \tag{27}$$

$$f_u = (2D_u\pi^2 - 2)u + u^3 + \tfrac{1}{2}v, \qquad f_v = (\tfrac{1}{2} + 2D_v\pi^2)v - \tfrac{3}{4}u.$$

Substituting Equation (12) into the canonical operators gives the forcing terms above, since $u_t = -u$, $v_t = -v/2$, $\Delta u = -2\pi^2 u$, and $\Delta v = -2\pi^2 v$. Boundary and initial values are obtained by restricting the analytical fields to the four spatial boundaries and the initial-time slice. The main-paper field figure reports the u slice at $t = 0.5$.

B.3 Kovasznay Flow

The Kovasznay benchmark solves the steady incompressible Navier–Stokes equations on $x \in [-0.5, 1]$, $y \in [-0.5, 1.5]$, with Reynolds number $Re = 40$ and $\nu = 1/Re$. The network predicts u, v, and p; the residual block contains the two momentum equations and continuity. Let

$$\lambda = \frac{Re}{2} - \sqrt{\left(\frac{Re}{2}\right)^2 + 4\pi^2}. \tag{28}$$

The analytical solution is

$$u = 1 - e^{\lambda x}\cos(2\pi y), \tag{29}$$

$$v = \frac{\lambda}{2\pi} e^{\lambda x}\sin(2\pi y), \tag{30}$$

$$p = \frac{1}{2}(1 - e^{2\lambda x}). \tag{31}$$

Boundary values and the reference pressure field follow the analytical solution above. The residual block contains the two momentum equations and continuity.

Table S3: Documented network, sampling, precision, and evaluation settings.

Protocol item	Shared setting	Note
Network	Four-layer MLP, width 64	Identical capacity
Sampling	512 collocation, 128 boundary/initial	Identical coordinates and targets
Randomness	Seeds 0, 1, 2; matched initialization/batches	Shared across tasks
Representations	R_0–R_4	Only the training residual basis changes
Precision	FP32 network/AD; FP64 ECI Gram solve	Training computation
Training budget	2000 epochs; evaluation every 100; checkpoint every 500	Shared baseline registry
Optimizer base	Initial learning rate 10^{-3}; boundary weight $\lambda_{\mathrm{BC}} = 10$	Method-specific refinements are reported separately
Evaluation	64-point dense grid in the canonical representation	Relative L_2 is computed after flattening all outputs

B.4 Equivalent Residual Representations R_0–R_4

For the canonical residual r_0, the representation protocol uses $r_k(z) = A_k(z)r_0(z)$. With severity parameter σ, the five labels denote identity (R_0); fixed diagonal scaling with entries linearly spaced from 0.25 to $2.5+\sigma$ (R_1); QR-based dense mixing with a diagonal similarity transform and an upper-triangular perturbation (R_2); the same mixing with a smooth coordinate-dependent skew term (R_3); and dense mixing at 1.5σ followed by diagonal scaling from 0.15 to $3.5+\sigma$ (R_4). The transformation acts on the PDE residual block; physical outputs and boundary and initial data retain their canonical meaning.

C Implementation Details

C.1 Unified Training Protocol

The reported network and sampling settings are collected in Table S1. The representation comparisons use the same physical tasks and canonical evaluation space. Figure 5 displays seed-0 fields under R_4. The common-grid results and their sample counts are reported separately in Table S2.

The archived common-grid registry uses the budget shown in Table S1. The final task–representation portfolio registry documents 5000 epochs, evaluation every 25 steps, and the same sampling, initial learning rate, and boundary weight; its entries enable warm-up fractions of 0.15–0.20, conditioned ridges of 10^{-6}–10^{-5}, and condition thresholds of 10^3–10^4. Where L-BFGS refinement is enabled, the implementation records one iteration per call, history size 50, and strong-Wolfe line search. These are registry-specific settings, not a claim of one uniform budget for every table.

C.2 Numerical Solve and Stabilization

The ECI computation solves $Gy = r$ without explicitly forming G^{-1}. The residual blocks contain two equations for Wave and Reaction-Diffusion and three for Kovasznay. The implementation records conditioning diagnostics and numerical failures. Whitening is applied before the Gram solve; the warm-up anchor blends the vanilla and ECI terms for its configured fraction; safe switching falls back according to the recorded condition gate; the optimizer portfolio selects a task–representation configuration; and checkpoint selection retains the lowest recorded relative L_2 among evaluation checkpoints. The component comparisons on Kovasznay are given in Table S4.

C.3 Evaluation Statistics

For each task and method, Average is the arithmetic mean of the relative L_2 errors over R_0–R_4 and seeds 0, 1, and 2; Worst is the maximum over the same 15 representation–seed pairs. Table S2 reports nine-run task–method summaries, with Std describing variation across the nine observations in each cell. Comparisons based on Table S2 are made within that table; its summaries are not pooled with Table 1.

Table S4: Common-grid relative L_2 statistics. Each task–method cell contains nine runs across three seeds and three shared representations. Std summarizes the variation within that cell; the 15 cells contain 135 runs in total.

Task	Method	N	Mean ↓	Std	Identity ↓	Worst ↓
Kovasznay	ECI-Diag	9	5.78e-02	1.44e-02	5.56e-02	7.74e-02
	ECI-PINN	9	**1.36e-02**	5.85e-03	**1.19e-02**	**2.32e-02**
	MultiAdam	9	6.70e-01	1.06e-02	6.72e-01	6.87e-01
	RMSNorm	9	6.95e-01	1.42e-06	6.95e-01	6.95e-01
	Vanilla PINN	9	5.84e-02	1.62e-02	5.28e-02	8.22e-02
Reaction-Diffusion	ECI-Diag	9	4.12e-02	2.41e-02	4.56e-02	9.07e-02
	ECI-PINN	9	**1.98e-02**	8.28e-03	**1.79e-02**	**3.84e-02**
	MultiAdam	9	5.02e-02	1.55e-02	3.72e-02	7.72e-02
	RMSNorm	9	2.47e-01	1.90e-02	2.50e-01	2.73e-01
	Vanilla PINN	9	4.69e-02	2.72e-02	5.47e-02	1.15e-01
Wave	ECI-Diag	9	1.58e-01	1.18e-01	6.62e-02	3.33e-01
	ECI-PINN	9	**3.27e-02**	7.62e-03	**3.31e-02**	**4.38e-02**
	MultiAdam	9	8.06e-01	2.14e-01	6.53e-01	9.46e-01
	RMSNorm	9	1.02e+00	1.94e-02	1.01e+00	1.06e+00
	Vanilla PINN	9	2.13e-01	1.19e-01	6.27e-02	3.50e-01

Table S5: Completion counts, numerical failures, and mean runtime for Vanilla PINN, MultiAdam, and ECI-PINN. Each method has 45 reported runs across three tasks, five representations, and three seeds.

Method	Complete	NaN/Inf	Time (s)	Solve fail
Vanilla PINN	45/45	0	54.73	0
MultiAdam	45/45	0	54.35	0
ECI-PINN	45/45	0	114.17	0

For prediction $\widehat{u}$ and reference u, the reported multi-output error is $\|\operatorname{vec}(\widehat{u} - u)\|_2 / \max(\|\operatorname{vec}(u)\|_2, 10^{-12})$ on the common dense grid. Identity is the arithmetic mean of the three R_0 seed values.

Table S2 contains 15 task–method cells, with nine runs per cell across three seeds and three shared representations. Its Std summarizes variation across all nine observations in a cell. Table S3 covers three methods with 45 runs each, corresponding to three tasks, five representations, and three seeds per method. Runtime includes the forward pass, automatic differentiation, backward pass, and optimizer step, and excludes one-time environment startup and plotting.

D Additional representation comparisons

Table S2 gives the additional common-grid comparisons. Figure S1 shows Reaction-Diffusion and Kovasznay fields under R_0 and R_4, together with absolute-error and residual-norm maps. Figure S2 gives the representation-wise relative errors for Vanilla PINN, MultiAdam, and ECI-PINN. Figures S3 and S4 summarize Average error and the Worst-minus-Average gap for these three methods.

E Stability, Numerical Reliability, and Efficiency

E.1 Error distributions

Figure S5 shows the relative L_2 error distributions for the three methods on each task. The ECI-PINN distributions lie below the baseline distributions in these panels. Table S2 gives the numerical summary of the nine-run common-grid cells.

E.2 Runtime and Numerical Reliability

Table S3 reports completion of 45 runs for each of Vanilla PINN, MultiAdam, and ECI-PINN, with no NaN/Inf values. The ECI runs record no Gram-solve failures. Mean runtimes are 54.73 s, 54.35

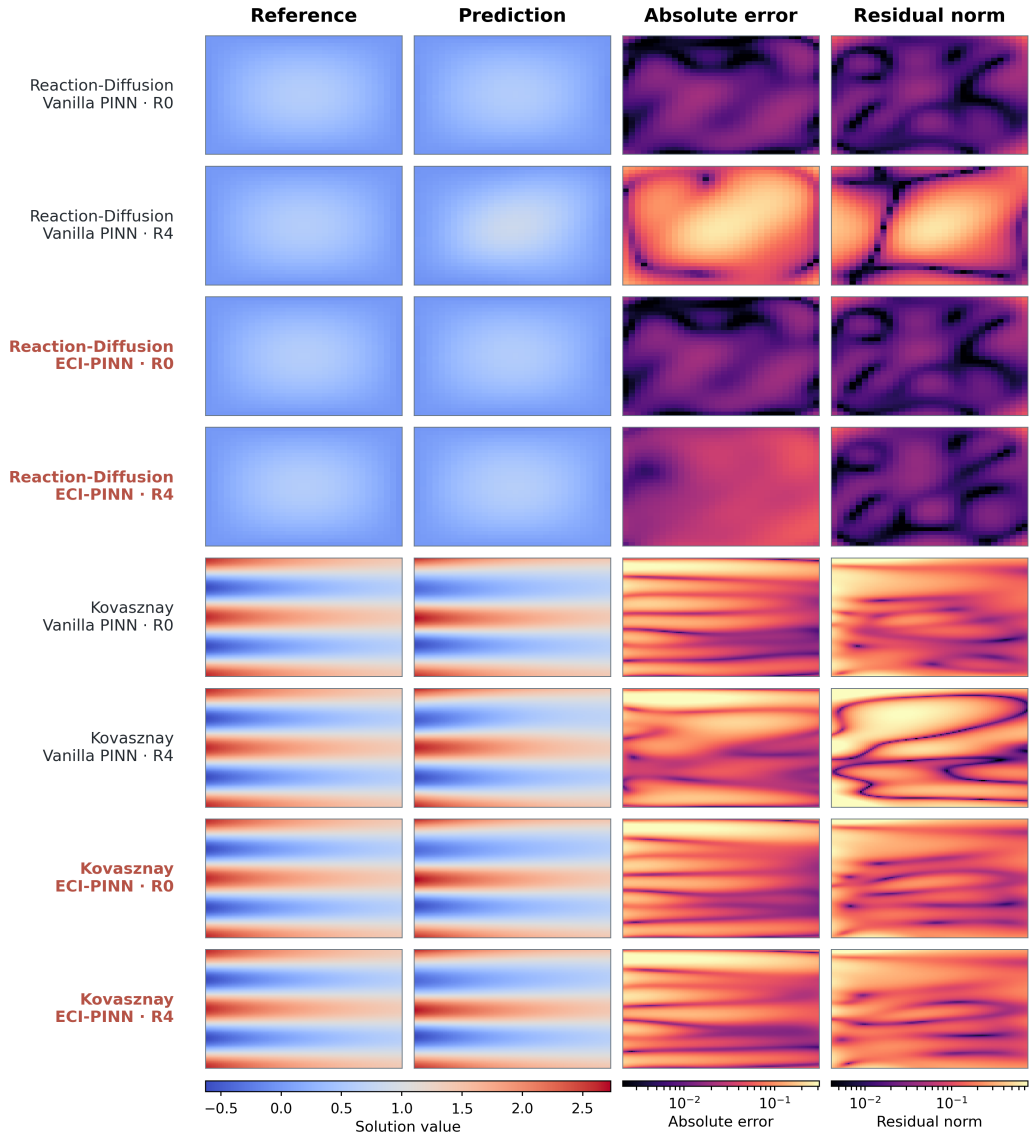

Figure S6: Field comparisons for Reaction-Diffusion and Kovasznay under R_0 and R_4. Columns show reference fields, predictions, absolute errors, and residual norms; the color bars specify the displayed scales.

s, and 114.17 s, respectively. Figure S6 plots runtime against solution error. Figure S7 plots solution error against the condition-number diagnostic.

F Strong Baselines and Component Ablations

F.1 External Optimizer Baselines

Figure S8 and Panel A of Table S4 compare the Kovasznay configurations using AdamW, NAdam, RMSProp, and momentum SGD with L-BFGS refinement. They supplement the optimizer comparisons in the main paper.

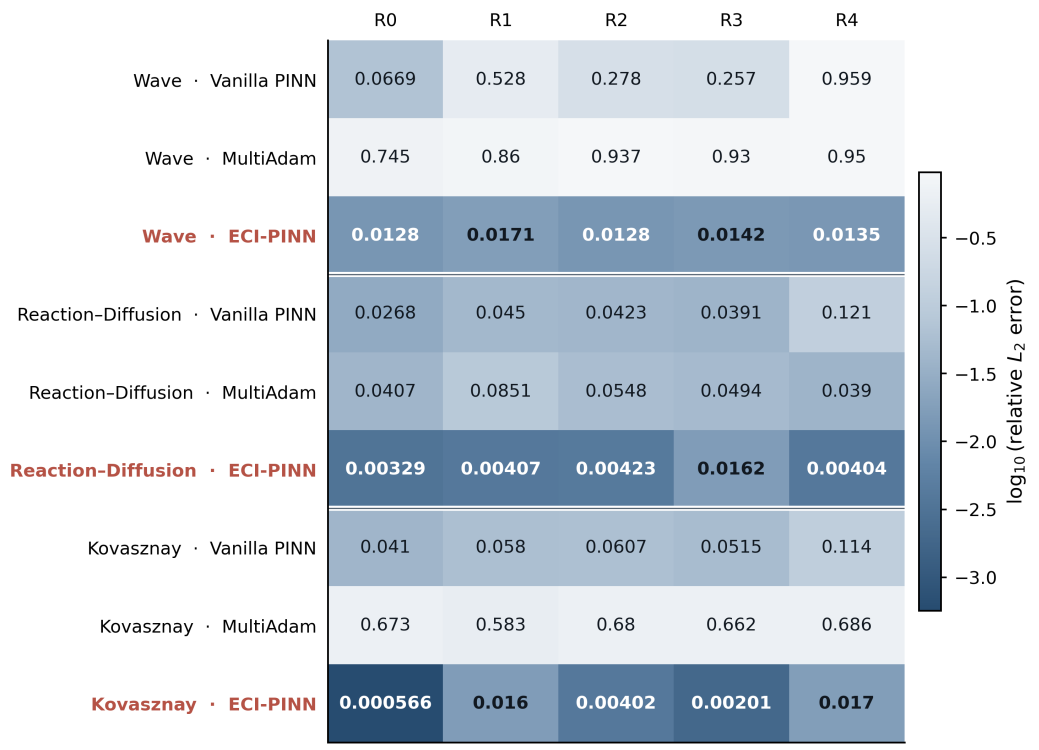

Figure S7: Displayed relative L_2 errors across R_0–R_4 for Vanilla PINN, MultiAdam, and ECI-PINN. The cell annotations give error values, and the color scale is logarithmic.

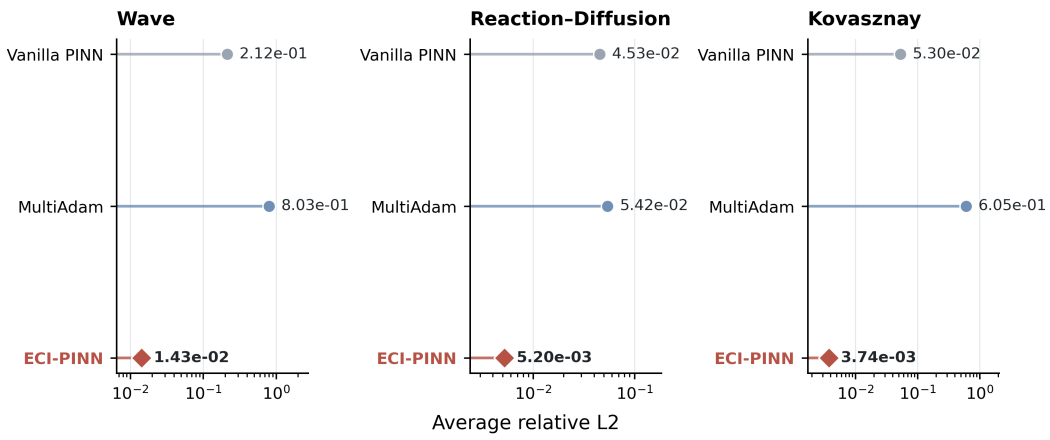

Figure S8: Average-error comparison for Vanilla PINN, MultiAdam, and ECI-PINN on the three PDE tasks.

F.2 Full Component Ablation

Figure S9 and Panel B of Table S4 compare the complete Kovasznay configuration with variants removing L-BFGS refinement, safe switching, whitening, the optimizer portfolio, or best-checkpoint selection.

The variant without best-checkpoint selection has Identity, Average, and Worst entries of 4.87×10^{-3}, 6.74×10^{-3}, and 3.88×10^{-2}, respectively.

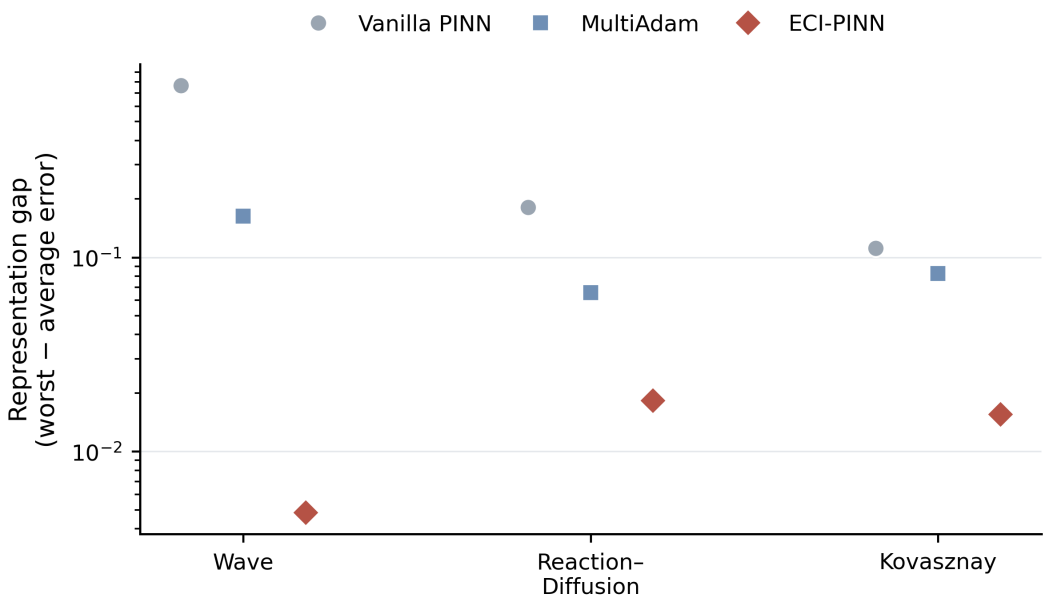

Figure S9: Representation-gap comparison for the three methods. The plotted gap is Worst minus Average error.

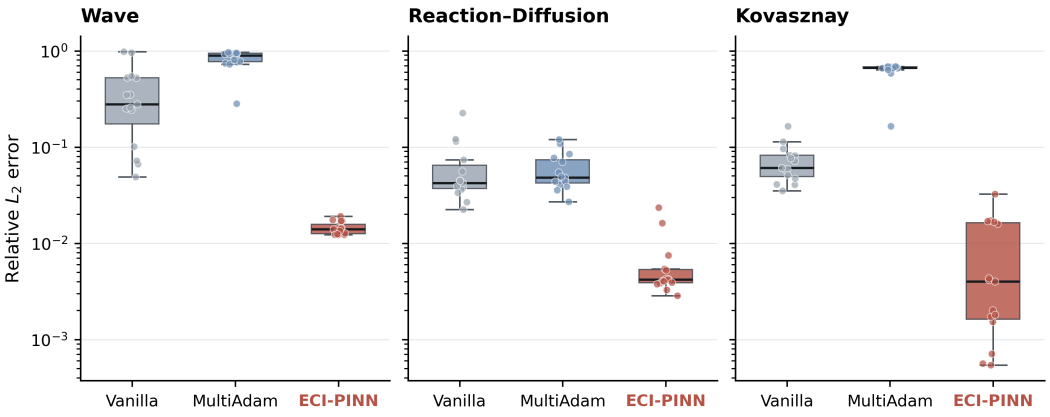

Figure S10: Relative L_2 error distributions for the three displayed methods on each task. Boxes and individual observations summarize the plotted values.

G Scope and Reproducibility

G.1 Scope

The analysis concerns a fixed dimensionless jet metric, full-row-rank residual Jacobians, and pointwise invertible residual transformations independent of parameters and jet state. The experimental systems contain two or three coupled residual components.

G.2 Documented settings

Appendix A provides the derivations and proofs. Appendix B defines the analytical PDE systems and residual-representation categories. Table S1 lists the documented network, sampling, and precision settings. Tables S2–S4 contain the additional comparisons, completion counts, runtimes, and component results.

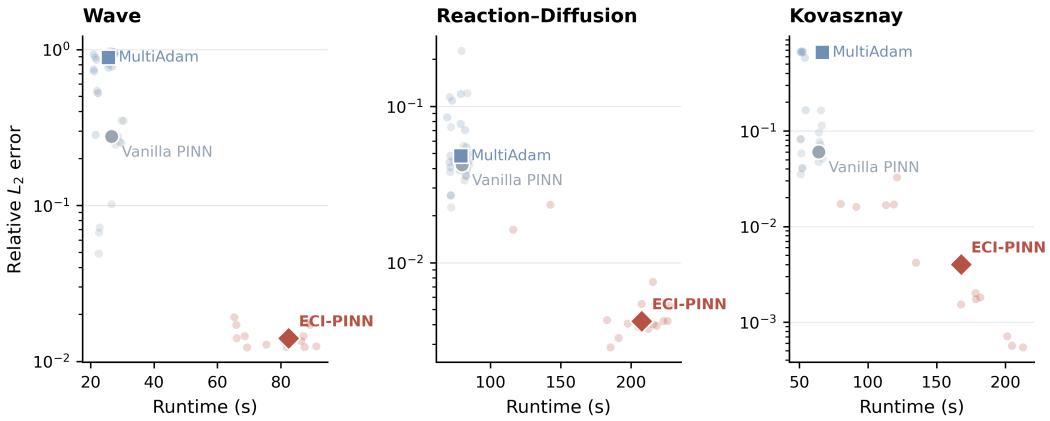

Figure S11: Runtime and solution-error observations for Vanilla PINN, MultiAdam, and ECI-PINN. Each panel corresponds to one PDE task.

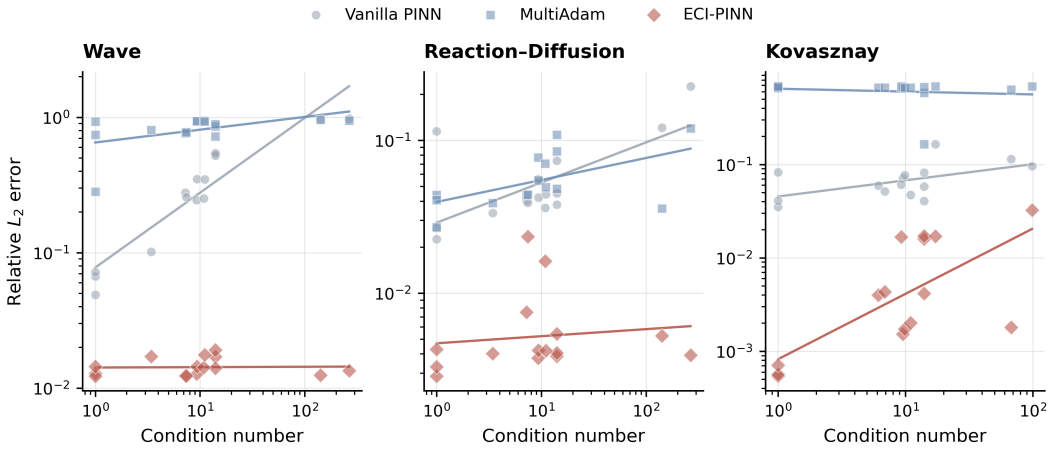

Figure S12: Solution error against the recorded condition-number diagnostic for the three displayed methods. Curves summarize the plotted observations within each task.

Exact transformation matrices, numerical jet scales, complete optimization and stabilization settings, and some evaluation conventions are not documented.

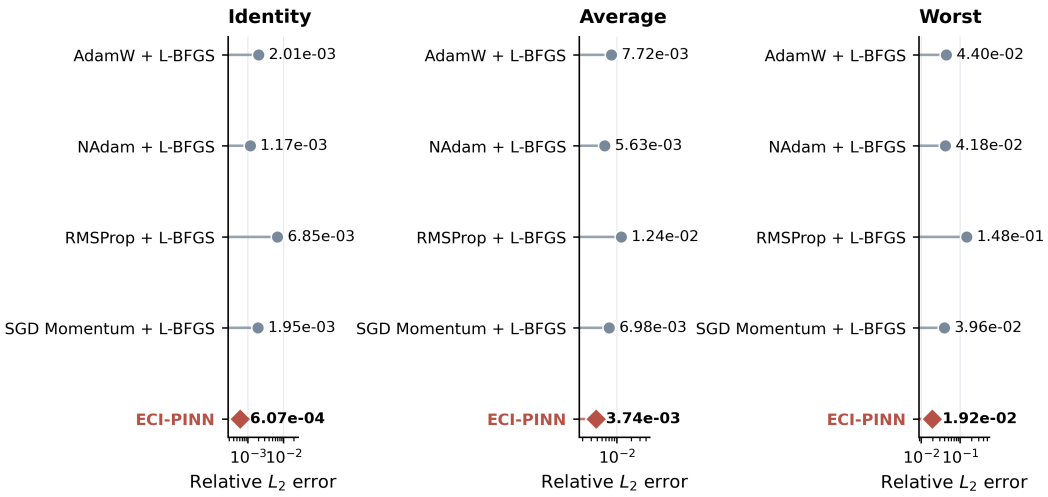

Figure S13: Kovasznay optimizer comparisons corresponding to Panel A of Table S4. Panels show Identity, Average, and Worst errors.

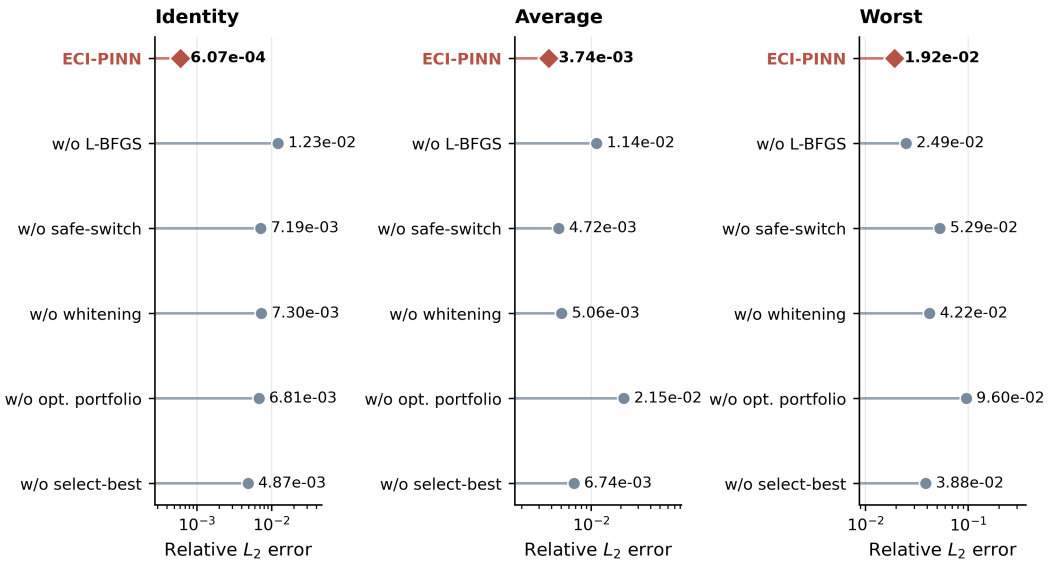

Figure S14: Kovasznay component comparisons corresponding to Panel B of Table S4. The complete configuration is compared with five named component removals.

Table S6: Kovasznay optimizer configurations and component ablations. Panel A lists optimizer baselines; Panel B lists the complete ECI-PINN configuration and five component removals. Identity, Average, and Worst follow the main comparison's definitions.

Panel	Method	Identity ↓	Average ↓	Worst ↓
External baseline	AdamW + L-BFGS	2.01e-03	7.72e-03	4.40e-02
	NAdam + L-BFGS	1.17e-03	5.63e-03	4.18e-02
	RMSProp + L-BFGS	6.85e-03	1.24e-02	1.48e-01
	SGD Momentum + L-BFGS	1.95e-03	6.98e-03	3.96e-02
	ECI-PINN	**6.07e-04**	**3.74e-03**	**1.92e-02**
Component ablation	ECI-PINN	**6.07e-04**	**3.74e-03**	**1.92e-02**
	w/o L-BFGS	1.23e-02	1.14e-02	2.49e-02
	w/o safe switch	7.19e-03	4.72e-03	5.29e-02
	w/o whitening	7.30e-03	5.06e-03	4.22e-02
	w/o optimizer portfolio	6.81e-03	2.15e-02	9.60e-02
	w/o select-best	4.87e-03	6.74e-03	3.88e-02