OUTER–INNER LAYER SEPARATION FOR THIRD-ORDER MIXED-DERIVATIVE PINNS WITH BILINEAR LAGGING

Anonymous authors
Paper under double-blind review

ABSTRACT

Higher-order bilinear PDE residuals pose two coupled challenges for physics-informed neural networks (PINNs). First, high-order derivatives entangle parameters across layers. Second, nonlinear products typically destroy the least-squares structure that makes Layer Separation (LySep) computationally attractive. We introduce an outer–inner LySep framework that addresses both. At each outer step, one bilinear factor is frozen, yielding a linear variable-coefficient PDE. An inner third-order LySep solver then reconstructs mixed derivatives through order three using an auxiliary hierarchy $a \to d \to q \to r$, while retaining a conditional least-squares output block. We establish exact auxiliary recovery and state-dependent residual consistency, and characterize the quartic obstruction that motivates outer lagging. Experiments span an intrinsically third-order forced KdV-type equation, hard-solenoidal Navier–Stokes, and resistive MHD. On KdV, the direct third-order route yields lower solution and original-PDE residual errors in all five paired seeds than both a matched low-order LySep route and a MIM-style first-order mixed-residual PINN, with median comparator-to-direct ratios of 1.95–2.97 for solution error and 2.47–3.61 for original-PDE residual. On hard-solenoidal Navier–Stokes, all five paired initializations give lower velocity, nonlinear-momentum, and pressure-gradient errors with LySep than with a matched end-to-end PINN. The MHD comparison likewise yields lower field and nonlinear-residual errors for LySep in all five pairs.

1 INTRODUCTION

Physics-informed neural networks (PINNs) fit a neural representation by minimizing differential-equation residuals (Raissi et al., 2019). Automatic differentiation supplies the required derivatives, but it does not remove the parameter coupling of the resulting objective. This becomes especially restrictive for higher-order bilinear PDEs. High-order derivatives entangle hidden activations, their derivatives, and network parameters across layers, while products of network-dependent quantities can destroy the least-squares structure of an otherwise tractable output block. The question studied here is therefore how to organize higher-order nonlinear physics-informed training into tractable subproblems while retaining a precise connection to the original physical residual.

The published LySep method of Liu & Gu (2025) is our direct foundation. It introduces independent hidden activation and derivative states, couples neighboring layers through mismatch penalties, and alternates network and auxiliary updates. Its detailed construction treats linear PDEs through second order: the hierarchy $a \to d \to q$ supports a least-squares output update for a fixed linear physical operator. Extending this construction to the present setting requires three ingredients not developed there: mixed derivatives through order three, a corresponding third-order auxiliary calculus, and a treatment of bilinear nonlinearities that preserves a tractable output block.

We address these issues with an outer–inner LySep construction. The inner representation extends the auxiliary hierarchy to $a \to d \to q \to r$, including mixed third derivatives and the complete output chain rule. At the outer level, one factor in each bilinear term is frozen, yielding a linear variable-coefficient inner PDE and retaining a conditional W_3 least-squares block. This construction isolates the quartic obstruction that appears when both network-dependent factors remain unfrozen, while the

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.

state-dependent consistency weights introduce a second quartic source. For the stated feedforward network with three affine layers and two hidden layers, we establish exact auxiliary recovery and state-dependent residual consistency.

We evaluate the method on an intrinsically third-order forced Korteweg–de Vries-type equation, hard-solenoidal Navier–Stokes formulations, and coupled hard-solenoidal resistive MHD. The KdV study compares the direct third-order route with both a predecessor-compatible low-order LySep route and a MIM-style first-order mixed-residual PINN. The NS and MHD studies use hard-solenoidal representations and matched lagged end-to-end comparisons.

Our contributions are threefold:

1. We extend LySep to mixed derivatives through order three for the stated three-affine-layer, two-hidden-layer FNN, with permutation symmetry, exact auxiliary recovery, and state-dependent residual consistency.
2. We combine this representation with outer bilinear lagging, retaining a conditional quadratic output block and characterizing quartic obstructions from unfrozen bilinear physics and dynamic-factor consistency weighting.
3. Across five paired KdV seeds, the direct third-order route yields lower solution and original-PDE residual errors in every pair against both matched comparators. The hard-solenoidal NS and resistive MHD comparisons likewise yield lower reported field and nonlinear-residual errors with LySep across all five paired runs. Pressure-gradient error is also lower with LySep in every one of the five paired NS initializations.

2 Related Work

PINN optimization. Residual-based PINNs (Raissi et al., 2019) face gradient imbalance and numerical stiffness (Wang et al., 2021), as well as loss-landscape ill-conditioning linked to differential operators (Rathore et al., 2024). These works diagnose or mitigate training pathologies of the residual objective through optimization- or loss-level interventions. Our focus is complementary: we study the algebraic structure of layer-separated parameter blocks, asking when a tractable output subproblem can be retained and how the separated objective remains connected to the original physical residual.

Layer separation and auxiliary splitting. Self-adaptive weighted auxiliaries (Liu et al., 2025) and LySep (Liu & Gu, 2025) provide the direct training foundation. Published LySep introduces independent hidden activation and derivative states, couples neighboring layers through consistency penalties, and treats linear PDEs through second order with the hierarchy $a \to d \to q$. The present work extends this setting to mixed derivatives through order three and bilinear nonlinearities while retaining a tractable output block. Other splitting strategies operate at a different level: Park et al. (2024) parameterize solution gradients for specified second-order linear PDEs, while FO-PINNs (Gladstone et al., 2023) introduce physical derivative outputs with first-order equations and compatibility losses. Our auxiliaries instead represent hidden preactivations and their derivatives inside trainable-hidden LySep, leading to a different reconstruction and consistency structure.

High-order PDEs and order reduction. MIM (Lyu et al., 2022) addresses high-order PDEs, including KdV examples, by rewriting them as first-order mixed systems with additional auxiliary variables and compatibility equations. Our construction instead retains the derivatives of the original frozen strong operator and represents them directly through the third-order LySep hierarchy, with exact auxiliary recovery, state-dependent consistency, and a conditional output least-squares block under bilinear lagging. The KdV study therefore includes a MIM-style first-order mixed-residual PINN following this order-reduction principle, adapted to the present benchmark (Appendix I.6).

Structure-preserving flow and MHD. Potential formulations and divergence-free representations have established precedents, including MHDnet (Guan et al., 2024), SP-RaNN (Li et al., 2026), and Neural Conservation Laws (Richter-Powell et al., 2022); hard initial and boundary constraints also appear in physics-constrained flow surrogates (Sun et al., 2020). SP-RaNN combines solenoidality, Picard/Newton linearization, and linear least squares with randomized hidden features. In the present

work, the hard-solenoidal representations provide the NS/MHD realization, while mixed-third LySep supplies the derivative and optimization structure for trainable hidden layers.

3 METHOD

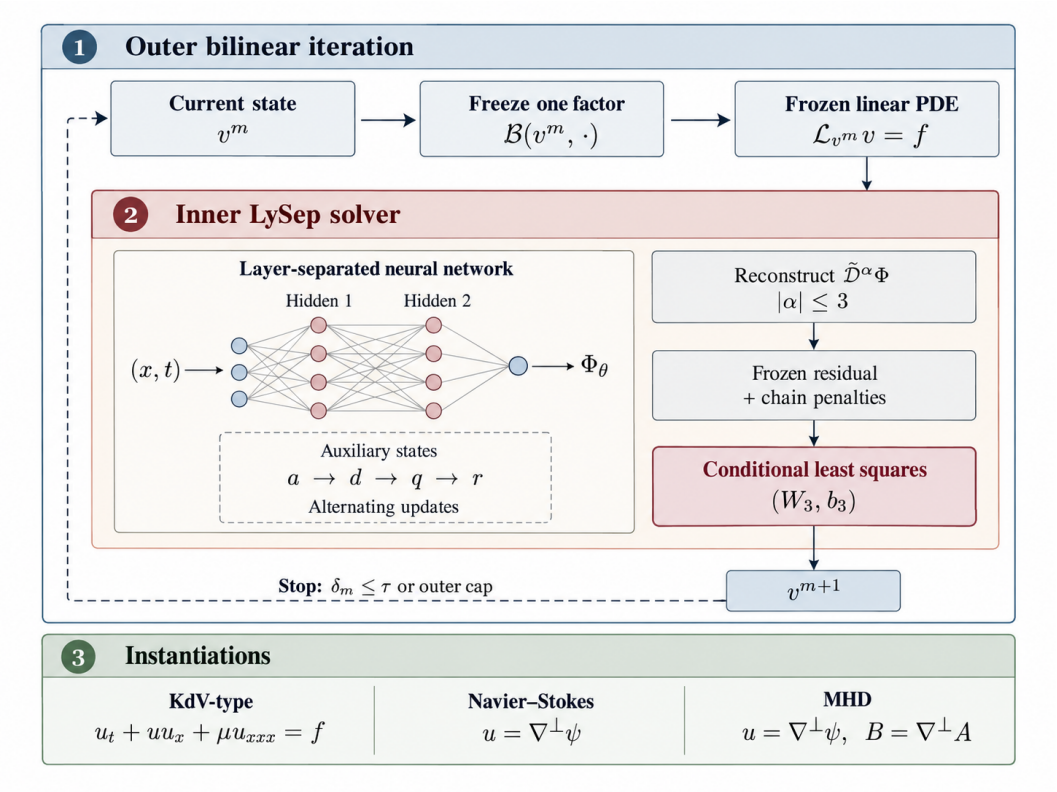


Figure 1: Outer–inner LySep framework for bilinear nonlinear PDEs. One-factor lagging produces a frozen linear variable-coefficient PDE. Inside each outer step, the layer-separated neural solver combines the auxiliary hierarchy, third-order reconstruction, and chain penalties with conditional (W_3, b_3) least squares at fixed hidden, auxiliary, and outer states. The reconstructed physical field supplies the next outer state. The framework is instantiated on KdV-type equations, hard-solenoidal Navier–Stokes flow, and bilinearly coupled resistive MHD.

3.1 FROM PURE-SECOND TO MIXED THIRD-ORDER AUXILIARIES

Let $x \in \mathbb{R}^d$ include time when required. Throughout, the network has exactly three affine maps:

$$a_1 = W_1 x + b_1, \quad a_2 = W_2 \sigma(a_1) + b_2, \quad \Phi_\theta = W_3 \sigma(a_2) + b_3. \tag{1}$$

The output may be scalar or vector valued. LySep treats samplewise hidden quantities as independent optimization variables, collectively z. The progression is $a \to d \to q \to r$: $d_{\ell,i}$ represents $\partial_i a_\ell$, $q_{2,ij}$ represents $\partial_{ij} a_2$, and $r_{2,ijk}$ represents $\partial_{ijk} a_2$. At exact chain values, $d_{1,i} = W_1(:, i)$ and all second and third derivatives of a_1 vanish. During optimization, $d_{1,i}$ is an independent copy tied to the corresponding weight column by a mismatch penalty. Auxiliaries are independent training variables for hidden values and derivatives, rather than additional physical fields in the PDE.

Only the required derivative families are stored. For the unsteady two-dimensional applications, these are $\{t, x, y\}$, $\{tx, ty, xx, xy, yy\}$, and $\{xxx, xxy, xyy, yyy\}$. For the scalar third-order test, only $\{t, x\}$, $\{xx\}$, and $\{xxx\}$ are needed. For mixed derivatives, every permutation of the derivative indices uses the same auxiliary state and hence the same optimization variable.

3.2 THIRD-ORDER RECONSTRUCTION

With $\odot$ denoting componentwise multiplication, the exact second-hidden-layer relations are

$$\begin{aligned} d_{2,i} &= W_2[\sigma'(a_1) \odot d_{1,i}], \\ q_{2,ij} &= W_2[\sigma''(a_1) \odot d_{1,i} \odot d_{1,j}], \\ r_{2,ijk} &= W_2[\sigma'''(a_1) \odot d_{1,i} \odot d_{1,j} \odot d_{1,k}]. \end{aligned} \tag{2}$$

The output third derivative is

$$\begin{aligned} \partial_{ijk}\Phi_\theta = W_3\big[&\sigma'''(a_2) \odot d_{2,i} \odot d_{2,j} \odot d_{2,k} \\ &+ \sigma''(a_2) \odot (q_{2,ij} \odot d_{2,k} + q_{2,ik} \odot d_{2,j} + q_{2,jk} \odot d_{2,i}) \\ &+ \sigma'(a_2) \odot r_{2,ijk}\big]. \end{aligned} \tag{3}$$

All three qd terms are needed, including their multiplicities when indices repeat. Replacing the exact chain quantities by independent auxiliary states defines $\widehat{D}^\alpha\Phi_\theta(z)$. Appendix A gives the lower-order formulas and derivation; Appendix B lists every mismatch row and adaptive weight.

The hard-solenoidal NS formulation requires these third derivatives. With $u = (\psi_y, -\psi_x)$, mixed-derivative symmetry enforces zero divergence, while the viscous term Δu involves $\psi_{xxx}, \psi_{xxy}, \psi_{xyy}, \psi_{yyy}$. Introducing $\chi = \Delta\psi$ as an independent output reduces the derivative order but permits harmonic compensation between χ and pressure in the mixed residual studied in Appendix E. Its L^2 relation penalty controls the channel error without controlling the gradient entering momentum. Direct third-order reconstruction eliminates this independent channel while preserving hard solenoidality (Appendix E, Details 5.1–5.3).

3.3 WHY DYNAMIC NONLINEAR FACTORS BREAK THE OUTPUT BLOCK

There are two distinct sources of quartic dependence. **Unfrozen bilinear physics:** when $u(W_3)$ and $u_x(W_3)$ are linear in W_3 at fixed hidden states, uu_x is generally quadratic, and the squared physical residual is generally quartic even without adaptive weighting. **Dynamic-factor consistency weighting:** an output-dependent squared factor multiplying a squared mismatch can also destroy the quadratic block. For a scalar output coordinate $u(w) = wh$ and a fixed auxiliary U, weighting the squared mismatch by the squared current factor gives

$$|u(w)|^2|u(w) - U|^2 = h^4w^4 - 2h^3Uw^3 + h^2U^2w^2. \tag{4}$$

The nonzero quartic term for $h \neq 0$ is an objective-level obstruction from dynamic-factor consistency weighting. Fixed consistency weights can retain quadraticity of some penalty blocks, with consistency constants depending on current-factor bounds. The unfrozen bilinear physical residual remains a separate source of quartic dependence.

3.4 OUTER LAGGING AND THE SEPARATED INNER LOSS

For $Lv + \mathcal{B}(v, v) = f$ with bilinear $\mathcal{B}$, we use

$$\mathcal{L}_{v^m} v^{m+1} = f, \qquad \mathcal{L}_{v^m} := L + \mathcal{B}(v^m, \cdot) = \sum_{|\alpha| \leq 3} A_\alpha(x; v^m) D^\alpha. \tag{5}$$

Application-specific one-factor orderings are stated in Appendix D (Detail 4.2) and specialized in Appendix E. The frozen fields are data throughout each inner solve. For N samples, its objective is

$$J_S(\theta, z; b) = \frac{1}{N} \sum_n \left\| \sum_\alpha A_{\alpha,n}(b) \widehat{D}^\alpha \Phi_{\theta,n}(z) - f_n \right\|^2 + \frac{1}{N} \sum_{n,s} \lambda_{s,n} \|\rho_{s,n}\|^2. \tag{6}$$

Here b denotes the frozen outer state, ρ the neighboring-layer chain mismatches, and λ the specified state-dependent weights. Any additional fixed linear boundary or gauge rows are handled identically. The reported hard-factor experiments require no such penalty rows (Appendix F).

Outer lagging makes the physical operator linear in the new state and fixes its coefficients throughout the inner solve. With hidden and auxiliary states fixed, the physical residual is then affine in the output parameters. Together with the single squared output norm in the specified adaptive penalties, this gives a conditional quadratic output problem (Appendix D, Details 3.5 and 4.1–4.2). Detaching adaptive weights alone would leave an unfrozen bilinear physical residual nonlinear. The weights are reevaluated at candidate states, and nonlinear block updates include their derivatives.

3.5 Compact training algorithm

Algorithm 1: bilinear-lagged third-order LySep.

```
Initialize the network from the prescribed seed; set frozen physical fields to zero.
for each outer solve up to the stated cap do
    Warm-start the network.
    Refresh all active auxiliaries to exact chain values.
    Assemble the frozen operator and reset trial step sizes.
    for each prescribed inner sweep do
        Update first-layer weights/bias, then a_1 and d_1.
        Update second-layer weights/bias, then a_2 and d_2.
        Update the required q auxiliaries.
        Update the required r auxiliaries.
        Update output weights/bias using the prescribed block rules.
    end for
    Perform the first-frozen-loss check.
    Reconstruct physical fields and compute their relative change δ_m.
    if δ_m ≤ 10^{-3} then
        stop
    end if
    Freeze the new physical fields.
end for
```

Nonlinear blocks use full current-weight gradients with nonincrease backtracking; the r block uses the prescribed quadratic update, and W_3 uses augmented minimum-norm least squares.

Every reported frozen solve uses 1000 sweeps. The exact block ordering, bias handling, and acceptance rule are given in Appendix D (Algorithm D.1), with per-task budgets and settings in Appendix G. The implementation uses an inexact outer iteration; the conditions for exact fixed-point recovery and sufficient local contraction are stated below.

4 What the training structure preserves

All statements concern (1), derivatives through order three, and the stated bilinear-lagged operators. The relevant activation derivatives are bounded and Lipschitz on the preactivation range; sine satisfies these conditions. Network and true/auxiliary state envelopes are finite, frozen coefficient matrices and included linear constraint rows are bounded, and required mixed derivatives commute.

Theorem 1 (third-order recursion and exact auxiliary recovery). For the network (1), the exact chain satisfies (2)–(3), including mixed-permutation symmetry. If $z(\theta)$ contains those exact hidden values and derivatives, every mismatch vanishes and

$$J_S(\theta, z(\theta); b) = J_{\mathrm{frozen}}(\theta; b). \tag{7}$$

Thus separation exactly preserves the frozen physical objective on the exact-chain manifold. Proofs are in Appendices A and B.

Theorem 2 (state-dependent residual consistency). Under the assumptions above, with positive weights for every mismatch contributing to a physical or included constraint row, there are finite coefficients $M_{s,n}(\theta, z; b)$ such that

$$J_{\text{frozen}}(\theta; b) \le C(\theta, z; b) J_S(\theta, z; b), \quad C = 2 \max\left\{1, \max_{s,n} \frac{M_{s,n}}{\lambda_{s,n}}\right\}. \tag{8}$$

These coefficients bound the true-minus-separated residual by weighted chain mismatches. Matrix-valued operators use Frobenius coefficient envelopes. Zero adaptive coefficients are permissible only when the corresponding physical discrepancy also vanishes, as in the concrete specialization detailed in Appendix C. The result is statewise: it supplies neither a trajectory-uniform constant nor monotone decrease of the true frozen loss. A small separated loss is informative only together with controlled coefficients, not in isolation.

Proposition 1 (conditional quadratic output block). Fix hidden weights, all auxiliaries, and outer coefficients. If every output-dependent adaptive contribution is $\|W_3\|_F^2 c_s \|\rho_s\|^2$, with fixed $c_s \ge 0$ and fixed ρ_s, then the output problem is positive-semidefinite quadratic. In the implemented objective,

$$J_S(W_3, b_3) = N^{-1}\big(\|D\,\mathrm{vec}(W_3) - y(b_3)\|^2 + \Lambda_{\text{ad}} \|W_3\|_F^2\big), \qquad \Lambda_{\text{ad}} \ge 0. \tag{9}$$

The factorization contains exactly one squared output norm and no $\|W_3\|_F^4$ term. A minimum-norm augmented solve selects an output update even when the design is rank deficient. Appendix D gives the derivation and conditions.

Proposition 2 (dynamic-factor quartic obstruction). A squared mismatch weighted by the square of a nonzero current factor linear in output parameters generically contributes a quartic term, as in (4). Such a contribution does not satisfy Proposition 1. Freezing the physical factor in (5) removes this source of output dependence while retaining the specified adaptive LySep weights.

Exact-chain recovery and off-chain residual control connect the separated loss to each frozen physical problem. Conditional quadraticity concerns only the output update with the other blocks fixed; it does not establish convergence of the outer iteration. At an exact fixed point of an exactly solved frozen equation, old and new factors coincide and recover the original nonlinear equation. A sufficient local Picard contraction result additionally requires an invariant complete state set and bounds on the selected inverse frozen solve and operator perturbation whose product is below one. Inexact inner solves contribute an additive error term to this contraction bound (Appendix D, Details 4.3–4.4). These conditions do not establish convergence of the full training procedure. The experiments therefore report the observed outer defects separately.

5 Experiments: testing the mechanism

The KdV, Navier–Stokes, and MHD studies respectively test intrinsic third-order bilinear physics, hard-solenoidal formulation and matched optimization, and coupled hard-solenoidal dynamics with a matched inner-optimization comparison.

We use fixed, non-overlapping Halton training/test points, sine networks and double precision in MATLAB R2024b. Each case has two equal-width hidden layers. Scalar, NS and MHD tests use analytic manufactured forcing, allowing errors and true nonlinear residuals to be measured independently of the separated objective. Field errors are discrete relative L^2 errors; nonlinear residuals are normalized by their source norms with denominator floor 10^{-14}. Pressure is compared only through its spatial gradient: it is determined up to a time-only function $c(t)$. Setting the pressure-output bias to zero removes only the explicit constant bias degree of freedom; the network can still represent time-only pressure, so this does not fully fix the gauge. Appendix G defines all metrics, settings and stopping rules.

5.1 Forced KdV-type benchmark: intrinsic third-order physics

We consider

$$u_t + u u_x + \mu u_{xxx} = f, \qquad (t, x) \in (0, 0.5] \times [-1, 1], \quad \mu = 0.05. \tag{10}$$

The manufactured solution is $u^* = te^{-t} H_K(x)(1 + 0.2\cos \pi x)$, with $H_K = (1 - x)^2(1 + x)$. The ansatz $u = tH_K \phi_\theta$ imposes the initial state and $u(t, -1) = u(t, 1) = u_x(t, 1) = 0$. Its frozen

Table 1: Five-seed KdV accuracy comparison. e_u is solution relative L^2 error; e_{PDE} is the original-PDE residual evaluated directly from the final u branch. Each paired ratio divides comparator error by direct error separately for each seed, then takes the median over the five seeds; it is not a ratio of medians. "Better pairs" counts lower direct errors in both metrics. For each accuracy metric, the lowest median among the three methods is shown in bold.

	Median error		Median paired ratio		Direct
Method	e_u	e_{PDE}	e_u	e_{PDE}	better pairs
Direct third-order LySep	$\mathbf{2.864 \times 10^{-4}}$	$\mathbf{2.517 \times 10^{-3}}$	1.00	1.00	—
Reduced low-order LySep	5.597×10^{-4}	8.703×10^{-3}	1.95	3.61	5/5
MIM-style mixed PINN	8.503×10^{-4}	6.024×10^{-3}	2.97	2.47	5/5

operator acts on ϕ_θ through only $\{0, t, x, xx, xxx\}$. We use width 30, 1000 training points, 400 test points and five paired seeds (1–5).

To ensure the example is neither effectively linear nor effectively lower-order, define

$$R_{\mathrm{nl}} = \frac{\|uu_x\|}{\max(\|u_t + \mu u_{xxx}\|, 10^{-14})}, \qquad R_3 = \frac{\|\mu u_{xxx}\|}{\max(\|u_t\|, 10^{-14})}. \tag{11}$$

Exact training ratios are 0.09384 and 0.1953; the seed-1 learned test ratios are 0.09473 and 0.1922, close to exact test values 0.09477 and 0.1922. Appendix I gives the full-precision values. Both terms remain active; Table 1 reports five-seed accuracy and paired comparisons.

The direct route retains the original third-order strong operator with a, d, q, r and outer one-factor lagging. The reduced low-order LySep route uses $v = u_x$, $w = v_x$ and rows $u_t + u^m v + \mu w_x = f$, $v - u_x = 0$, $w - v_x = 0$, with a, d only. The MIM-style first-order mixed-residual PINN uses the same mixed unknowns and compatibility rows but trains $u_t + uv + \mu w_x = f$ jointly, without outer lagging or LySep auxiliaries.

Across five paired initializations, the direct route gives lower solution error and original-PDE residual than both comparators in every pair. Median paired comparator/direct ratios are 1.95 (solution error) and 3.61 (original-PDE residual) for reduced LySep, and 2.97 (solution error) and 2.47 (original-PDE residual) for the MIM-style PINN (Table 1). Under the common seven-solve cap and 1000 sweeps per frozen problem, direct and reduced LySep reach the prescribed outer tolerance in 4/5 and 0/5 seeds; this criterion is inapplicable to the MIM-style PINN. Seed 1 of the direct route did not reach the outer tolerance within the original five-solve budget and was continued for two post-hoc solves; the seven-solve cap was fixed before seeds 2–5 were run (Appendix I). Appendix I gives the complete continuation details, the distinct lagged E2E result, and full five-seed evidence.

5.2 Hard-solenoidal Navier–Stokes

For two-dimensional NS, set $u = (\psi_y, -\psi_x)$ with $\psi = tP(x)P(y)\phi_\psi$ and $P(s) = (1 - s^2)^2$. Then divergence vanishes by mixed-derivative symmetry, while the boundary factor enforces no slip. Substituting this representation into viscous primitive momentum requires $\psi_{xxx}, \psi_{xxy}, \psi_{xyy}, \psi_{yyy}$. The network outputs only (ϕ_ψ, ϕ_p), with no independent Laplacian field. Appendix E gives the resulting momentum equations and the derivative index set used to assemble the frozen operator.

Table 2(a) shows the formulation progression. Primitive soft incompressibility leaves relative divergence 2.1294×10^{-3}. The second-order route achieves maximum absolute divergence 1.73×10^{-18}, but its independent $\chi = \Delta\psi$ has relation error 1.0810×10^{-1}. The L^2 relation penalty controls values while momentum involves their gradients, permitting harmonic χ–pressure compensation (Appendix E). Removing χ preserves exact divergence and directly represents xxx, xxy, xyy, yyy.

At width 40, 2000/500 points, $T = 0.25$, $\nu = 0.1$, and seed 1, the chi-free route reduces velocity, momentum, and pressure-gradient errors by factors of approximately 332, 60.6, and 27.3, respectively, relative to the mixed route. After two solves its outer defect is 1.3280×10^{-4}, with zero maximum divergence and boundary speed.

Table 2: NS evidence: (a) seed-1 formulation study, with $\chi = \Delta\psi$ denoting the independent Laplacian field in the second-order formulation; (b) five-seed optimization comparison (median errors) within the final chi-free formulation. Lower relative L^2 errors are better; bold marks the lowest values in each panel. "Hard div." marks formulations in which $\nabla\cdot u = 0$ holds exactly by construction rather than through a soft penalty. In (b), better pairs counts lower LySep errors in all three metrics.

	Order	Hard div.	e_u	e_{mom}	$e_{\nabla p}$	LySep better pairs
(a) Formulation study (seed 1)						
Primitive (u, v, p)	2	No (soft)	1.357×10^{-3}	2.492×10^{-3}	7.183×10^{-2}	—
Mixed (ψ, χ, p)	2	Yes	1.256×10^{-2}	4.886×10^{-2}	5.225×10^{-1}	—
Chi-free (ψ, p)	3	Yes	$\mathbf{3.789 \times 10^{-5}}$	$\mathbf{8.063 \times 10^{-4}}$	$\mathbf{1.916 \times 10^{-2}}$	—
(b) Matched five-seed optimization (median errors)						
Third-order LySep	3	Yes	$\mathbf{2.279 \times 10^{-5}}$	$\mathbf{4.775 \times 10^{-4}}$	$\mathbf{1.129 \times 10^{-2}}$	—
Matched E2E PINN	3	Yes	2.299×10^{-3}	1.807×10^{-2}	4.003×10^{-1}	5/5 (all three)

Once the final chi-free formulation is fixed, we compare LySep with an end-to-end PINN using the same network, points, paired initialization, hard representation and Oseen outer protocol. Across five paired seeds, LySep gives lower velocity, nonlinear-momentum and pressure-gradient errors in every pair (Table 2(b)). Median paired E2E/LySep ratios are 100.9, 35.2 and 32.5, respectively. Both methods meet the prescribed outer tolerance in all five seeds and inherit hard solenoidality and boundary conditions from the shared representation. Appendix H reports the protocol, seed-level evidence, and a single-seed width-80 stress test with 8000/2000 train/test points that retains the same ordering in velocity, nonlinear-momentum and pressure-gradient accuracy.

5.3 Bilinearly coupled MHD

MHD uses a shared-trunk output $(\phi_\psi, \phi_A, \phi_p)$, with $B = (A_y, -A_x)$ and $A = t(1-x^2)(1-y^2)\phi_A$. This gives hard $\nabla \cdot B = 0$ and $B \cdot n = 0$ without forcing the tangential magnetic field to vanish. Momentum freezes the advecting velocity and magnetic factors; scalar induction uses $A_t + u^m \cdot \nabla A - \eta \Delta A = g_A$. With smooth fields, solenoidal velocity, spatially constant diffusivity and compatible forcing, its two-dimensional curl recovers nonlinear vector induction at an exact fixed point. Intermediate scalar and primitive-vector Picard linearizations differ in which factor is frozen in the magnetic stretching term (Appendix E, Detail 5.5).

The moderate-coupling manufactured case uses $T = 1$, $\nu = \eta = 0.1$, width 40, and 2000/500 points. Velocity- and magnetic-potential amplitudes are 0.75 and 0.40, and an asymmetric magnetic factor prevents geometric alignment. Exact induction, velocity-advection, and magnetic-tension ratios are 0.05148, 0.21529, and 0.06878 relative to their corresponding linear terms; their definitions and learned values are in Appendix H. These ratios verify active coupling in the manufactured fields.

After four outer solves, the outer defect is 9.0640×10^{-5}. Table 3 includes this seed-1 result alongside the paired runs; Appendix H retains the final-case summary. Maximum velocity divergence is zero and magnetic divergence is 5.55×10^{-17}; both hard boundary errors are zero. This realization satisfies the hard identities and attains the reported error levels.

The reported scalar-induction residual measures the potential equation, while the magnetic-field error measures the reconstructed field. The scalar-to-vector identity differentiates the residual, so a small scalar L^2 norm alone does not control its curl. A vector-induction residual bound would require additional derivative control. The two reported metrics therefore assess magnetic-field accuracy and scalar-equation accuracy separately (Appendix H, Detail 6.4; Appendix J.3).

5.4 Comparison with a matched lagged end-to-end PINN

The MHD comparator shares the physical formulation, hard representation, outer Picard map, zero initial frozen state, stopping criterion, collocation points, architecture, initialization, precision and MATLAB environment, so the comparison isolates the inner optimization procedure under the stated method-specific budgets. The comparison uses five paired initializations: 1000 LySep sweeps per frozen problem versus 2000 full-batch Adam steps and at most 500 native L-BFGS updates. The

Table 3: Paired MHD results for seeds 1–5 and their medians. Lower errors are better; bold marks the lower error within each pair. e_{ind} is the scalar-induction residual. Optimizer budgets differ as specified in the text. Entries are mantissas, scaled by the power of ten shown in each column header.

Seed	Field and pressure-gradient errors						Nonlinear residuals			
	e_u		e_B		$e_{\nabla p}$		e_{mom}		e_{ind}	
	LySep $\times 10^{-4}$	E2E $\times 10^{-4}$	LySep $\times 10^{-4}$	E2E $\times 10^{-3}$	LySep $\times 10^{-2}$	E2E $\times 10^{-2}$	LySep $\times 10^{-4}$	E2E $\times 10^{-3}$	LySep $\times 10^{-3}$	E2E $\times 10^{-3}$
1	**1.671**	4.168	**7.282**	2.010	**5.036**	14.14	**8.313**	2.130	**1.906**	5.082
2	**1.578**	3.409	**5.332**	5.278	**3.191**	13.76	**7.371**	1.633	**1.430**	9.266
3	**1.673**	3.397	**6.016**	3.923	**2.789**	11.61	**7.666**	1.682	**1.825**	6.774
4	**1.768**	4.725	**6.983**	4.431	**6.432**	9.558	**9.089**	2.246	**1.661**	7.866
5	**1.788**	5.758	**7.682**	4.085	**3.941**	13.82	**7.819**	2.492	**1.902**	9.895
Median	**1.673**	4.168	**6.983**	4.085	**3.941**	13.76	**7.819**	2.130	**1.825**	7.866

end-to-end network uses the physical residual alone; Appendices G and H provide the optimizer settings, stopping rules and paired initialization protocol.

With five prescribed paired seeds, the comparison remains a descriptive paired assessment rather than a population-level statistical claim.

LySep gives lower reported velocity, magnetic-field, momentum-residual, scalar-induction-residual and pressure-gradient errors in every one of the five paired seeds and in all five-seed medians (Table 3). Median paired E2E/LySep ratios are approximately 2.49, 6.34, 2.47, 4.73, and 3.51 for velocity, magnetic field, momentum residual, scalar-induction residual, and pressure-gradient error, respectively. All five LySep runs meet the prescribed benchmark tolerances. None of the five E2E runs meets every tolerance: seeds 1–3 and 5 exceed the pressure-gradient threshold of 0.1, whereas seed 4 instead stops at the five-solve cap with outer defect above 10^{-3}.

The shared outer protocol fixes the update rule, while each method supplies its own reconstructed fields to the next frozen solve. Later coefficient fields generally differ, so the final frozen losses measure different linear problems. The end-to-end objective also omits the auxiliary consistency penalties used by LySep. Table 3 therefore compares current-field nonlinear residuals, with both nonlinear factors evaluated from each method's final fields, alongside paired field and pressure-gradient errors. These metrics assess physical accuracy under the stated method-specific optimizer budgets (Appendices H and J.3).

6 Conclusion and limitations

The extension from $a \to d \to q$ to $a \to d \to q \to r$, combined with outer bilinear lagging, retains conditional W_3 least squares for nonlinear mixed-third physics. Exact auxiliary recovery and state-dependent consistency connect the separated problem to the original frozen residual. The experiments connect this construction to accuracy across three PDE families: intrinsic third-order KdV-type physics and hard-solenoidal NS/MHD flow. In KdV, the direct route improves on both the predecessor-compatible low-order LySep route and the MIM-style mixed-residual baseline in solution and original-PDE residual errors for every one of the five paired initializations. For NS, the formulation study supports the chi-free representation, within which LySep has lower velocity, nonlinear-momentum and pressure-gradient errors than matched E2E in all five pairs. Lower reported continuous errors also hold for MHD across all five paired seeds and their medians.

Scope and outlook. The analysis covers the fixed three-affine-layer, two-hidden-layer architecture, mixed derivatives through order three, and the stated bilinear operators. The layerwise recursion suggests a route to deeper networks; the additional auxiliary states, block updates, and consistency constants for deeper variants remain to be established. The experiments cover smooth manufactured KdV/NS/MHD benchmarks, including moderate MHD coupling. Extending this derivative calculus and block structure to broader architectures and physical regimes is a natural next step. In the current MATLAB implementation, LySep incurs higher wall-clock cost than the matched end-to-end baseline (Appendix H); improving the accuracy–cost tradeoff is left for future work.

AI USE STATEMENT

Generative AI tools were used to assist with writing and editing, retrieval and discovery of related work, and research ideation and execution, including methodological discussion, implementation, mathematical checking, and experiment-workflow support. All AI-assisted outputs were reviewed by the authors, who take responsibility for the final scientific content and reported results.

REPRODUCIBILITY STATEMENT

Complete derivations, mismatch rows, adaptive weights, physical formulations, manufactured forcings, metric definitions, hyperparameters, and training protocols are given in Appendices A–I, with seed-level raw results in Appendices H and I. All experiments use fixed non-overlapping Halton point sets, prescribed seeds, and double precision in MATLAB R2024b as documented there; the training-free implementation checks are listed in Appendices I and J.

REFERENCES

Rini J. Gladstone, Mohammad A. Nabian, N. Sukumar, Ankit Srivastava, and Hadi Meidani. FO-PINNs: A first-order formulation for physics-informed neural networks, 2023. URL https://arxiv.org/abs/2210.14320v2. Preprint.

X. Guan, B. Hu, S. Mao, X. Wang, and Z. Yang. MHDnet: Physics-preserving learning for solving magnetohydrodynamics problems. *Communications in Computational Physics*, 36(4):943–976, 2024. doi: 10.4208/cicp.OA-2024-0002.

Y. Li, F. Wang, and L. Li. Structure-preserving randomized neural networks for incompressible magnetohydrodynamics equations, 2026. URL https://arxiv.org/abs/2603.01102. Preprint.

Y. Liu and Y. Gu. Layer separation deep learning model with auxiliary variables for partial differential equations. *Journal of Computational Physics*, 543:114414, 2025. doi: 10.1016/j.jcp.2025.114414.

Y. Liu, Y. Gu, and M. K. Ng. Deep learning optimization using self-adaptive weighted auxiliary variables, 2025. URL https://arxiv.org/abs/2504.21501. Preprint.

Liyao Lyu, Zhen Zhang, Minxin Chen, and Jingrun Chen. MIM: A deep mixed residual method for solving high-order partial differential equations. *Journal of Computational Physics*, 452:110930, 2022. doi: 10.1016/j.jcp.2021.110930.

Y. Park, C. Song, and M. Kang. Beyond derivative pathology of PINNs: Variable splitting strategy with convergence analysis, 2024. URL https://arxiv.org/abs/2409.20383. Preprint.

M. Raissi, P. Perdikaris, and G. 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. 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 of *Proceedings of Machine Learning Research*, pp. 42159–42191, 2024. URL https://proceedings.mlr.press/v235/rathore24a.html.

Jack Richter-Powell, Yaron Lipman, and Ricky T. Q. Chen. Neural conservation laws: A divergence-free perspective. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/f8d39584f87944e5dbe46ec76f19e20a-Abstract-Conference.html.

Luning Sun, Han Gao, Shaowu Pan, and Jian-Xun Wang. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. *Computer Methods in Applied Mechanics and Engineering*, 361:112732, 2020. doi: 10.1016/j.cma.2019.112732.

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.

This appendix supplies full mathematics, implemented weights, physical formulations, protocols, raw numerical results, and development disclosures. The main text presents the primary scientific results. All references are the same as in the main text.

Notation: the detailed architecture is (2.1). The physical unknown and auxiliary state both use z in some detailed formulas; context distinguishes them, while main text uses v for the physical unknown. Norms in the consistency proof are pointwise Euclidean/Frobenius norms and summed sample norms as stated. The implemented output least-squares regularization is the adaptive consistency contribution, not an additional physical or pressure regularizer.

Detailed equations use the identifiers (2.1)–(6.3e) and (A.1)–(A.13). These are appendix equation identifiers, not references to main-text equations (1)–(11). The detailed theorem labels 3.1/3.2 correspond jointly to main Theorem 1, 3.3 to main Theorem 2, 3.4 to main Proposition 1, and 4.1 to main Proposition 2. Their statements are unchanged in mathematical scope. The appendix sections below are indexed A–J. Appendices A–D develop the auxiliary calculus and optimization theory; E–F give the NS/MHD formulations and concrete operators; G–I record protocols and accepted numerical evidence; J documents development-stage qualifications.

A Architecture, complete recursion and permutation symmetry

Detail 2.1 Second-order LySep foundation

Let $x \in \Omega \subset \mathbb{R}^d$, where time may be included among the coordinates, and let the network output dimension be m. We use the three-layer fully connected form

$$a_1(x) = W_1 x + b_1, \qquad a_2(x) = W_2\sigma(a_1(x)) + b_2, \qquad \Phi_\theta(x) = W_3\sigma(a_2(x)) + b_3, \tag{2.1}$$

with componentwise activation σ. For a multi-index α, $D^\alpha\Phi_\theta$ denotes the corresponding input derivative. The published second-order LySep construction (Liu & Gu, 2025) introduces activation, first-derivative, and pure second-derivative auxiliaries. Denoting componentwise multiplication by $\odot$, its exact second-layer derivative relations include

$$d_{2,i} = W_2[\sigma'(a_1) \odot d_{1,i}], \qquad q_{2,i} = W_2[\sigma''(a_1) \odot d_{1,i}^{\odot 2}], \tag{2.2}$$

where $q_{2,i} = \partial_{ii} a_2$ and $d_{1,i} = W_1(:, i)$ because a_1 is affine. The general mixed family $q_{2,ij}$ is part of the present extension, followed by the third-order family $r_{2,ijk}$ in Appendix A; it is not attributed to the published pure-second construction. Replacing exact hidden quantities by independent states breaks the multilayer computational graph into shallow neighboring-layer relations. Their mismatches are penalized in a separated loss, denoted J_S, while the corresponding unseparated physical loss is denoted J_{frozen}.

For a fixed linear differential operator, exact auxiliary values make all consistency penalties vanish and recover the original network derivatives. Conversely, weighted consistency residuals control the discrepancy between separated and true derivatives. The output-layer update is particularly valuable: once hidden features and auxiliaries are fixed, every network value or derivative is affine or linear in (W_3, b_3), so the physical rows form a least-squares block. The present work preserves this foundation but extends its derivative algebra and consistency propagation through order three.

We assume throughout that the activation derivatives needed below are bounded and Lipschitz on the relevant preactivation range, all network and auxiliary states entering a bound are finite, and the frozen coefficient fields are essentially bounded. A convenient sufficient global condition is $\sigma \in C^4(\mathbb{R})$ with bounded derivatives through order four. The sine activation used in the experiments satisfies these assumptions globally.

Detail 3.1 Auxiliary variables

For coordinate indices $i, j, k \in \{1, \ldots, d\}$, define hidden derivative states

$$d_{\ell,i} = \partial_i a_\ell, \qquad q_{\ell,ij} = \partial_{ij} a_\ell, \qquad r_{\ell,ijk} = \partial_{ijk} a_\ell. \tag{3.1}$$

Since a_1 is affine,

$$d_{1,i} = W_1(:, i), \qquad q_{1,ij} = 0, \qquad r_{1,ijk} = 0. \tag{3.2}$$

Equation (3.2) describes the exact chain. During separated optimization, sample-wise copies of $d_{1,i}$ are independent states and are tied to $W_1(:, i)$ by the $\rho_{d_{1,i}}$ rows in (3.10b).

Only derivative indices actually present in the PDE operator need to be stored. For the two-dimensional unsteady applications below, the minimal set is

$$\begin{aligned} \mathcal{I}_1 &= \{t, x, y\}, \\ \mathcal{I}_2 &= \{tx, ty, xx, xy, yy\}, \\ \mathcal{I}_3 &= \{xxx, xxy, xyy, yyy\}. \end{aligned} \tag{3.3}$$

Mixed-derivative permutations share one state; for example, $r_{xxy} = r_{xyx} = r_{yxx}$. Neither third-order time derivatives nor a full tensor of unused permutations is introduced.

DETAIL 3.2 THIRD-ORDER NETWORK DERIVATIVE RECURSION

Theorem 3.1 (third-order auxiliary recursion). Suppose the network (2.1) has classical input derivatives through order three. Then

$$\begin{aligned} d_{2,i} &= W_2[\sigma'(a_1) \odot d_{1,i}], \\ q_{2,ij} &= W_2[\sigma''(a_1) \odot d_{1,i} \odot d_{1,j}], \\ r_{2,ijk} &= W_2[\sigma'''(a_1) \odot d_{1,i} \odot d_{1,j} \odot d_{1,k}], \end{aligned} \tag{3.4}$$

and the output derivatives are

$$\partial_i \Phi_\theta = W_3[\sigma'(a_2) \odot d_{2,i}], \tag{3.5}$$

$$\partial_{ij} \Phi_\theta = W_3[\sigma''(a_2) \odot d_{2,i} \odot d_{2,j} + \sigma'(a_2) \odot q_{2,ij}], \tag{3.6}$$

and

$$\begin{aligned} \partial_{ijk} \Phi_\theta = W_3 \big[& \sigma'''(a_2) \odot d_{2,i} \odot d_{2,j} \odot d_{2,k} \\ & + \sigma''(a_2) \odot (q_{2,ij} \odot d_{2,k} + q_{2,ik} \odot d_{2,j} + q_{2,jk} \odot d_{2,i}) \\ & + \sigma'(a_2) \odot r_{2,ijk} \big]. \end{aligned} \tag{3.7}$$

The q and r states, and the reconstructed output derivatives, are invariant under permutations of their derivative indices.

Proof. Differentiating $a_1 = W_1 x + b_1$ yields (3.2). One differentiation of $a_2 = W_2 \sigma(a_1) + b_2$ gives the first relation in (3.4). A second differentiation would include a term proportional to $\partial_{ij} a_1$, but this term vanishes, leaving the $\sigma'' d_{1,i} d_{1,j}$ product. At third order, all terms containing second or third derivatives of a_1 vanish, leaving only the $\sigma''' d_{1,i} d_{1,j} d_{1,k}$ term.

For the output, the first derivative gives (3.5). Applying the product rule once gives (3.6). Differentiating the first term of (3.6) with respect to x_k produces the cubic $d_{2,i} d_{2,j} d_{2,k}$ term and two of the q, d terms. Differentiating the second term produces the remaining $q_{2,ij} d_{2,k}$ term and $r_{2,ijk}$, which yields (3.7). Classical mixed derivatives commute, and all displayed Hadamard products are commutative, establishing permutation symmetry. □

B FULL MISMATCH ROWS, ADAPTIVE WEIGHTS AND EXACT RECOVERY

DETAIL 3.3 EXACT AUXILIARY RECOVERY

For a separated state $z = (a, d, q, r)$, let $\widehat{D}^\alpha \Phi_\theta(z)$ denote the value or derivative obtained by substituting z into (2.1) and (3.5)–(3.7). At a fixed outer state b, write the linear operator

$$\mathcal{L}_b \Phi = \sum_{|\alpha| \le 3} A_\alpha(x; b) D^\alpha \Phi, \tag{3.8}$$

where A_α may be matrix-valued and sample dependent. On collocation samples x_n, $n = 1, \dots, N$, the original frozen loss is normalized per sample as

$$J_{\text{frozen}}(\theta; b) = \frac{1}{N} \sum_{n=1}^{N} \|\mathcal{L}_b \Phi_\theta(x_n) - f_n\|_2^2 + J_{\text{bdry/gauge}}(\theta), \tag{3.9}$$

and the separated loss is

$$\begin{aligned} J_S(\theta, z; b) =& \frac{1}{N} \sum_{n=1}^{N} \left\| \sum_{|\alpha| \le 3} A_{\alpha,n}(b) \widehat{D}^\alpha \Phi_{\theta,n}(z) - f_n \right\|_2^2 + \widehat{J}_{\text{bdry/gauge}}(\theta, z) \\ &+ \frac{1}{N} \sum_{n=1}^{N} \sum_{s \in \mathcal{S}} \lambda_{s,n}(\theta, z; b) \|\rho_{s,n}(\theta, z)\|_2^2. \end{aligned} \tag{3.10}$$

For the actual three-layer implementation, the complete mismatch set is

$$\mathcal{S} = \{a_1, a_2\} \cup \{d_{1,i}, d_{2,i} : i \in \mathcal{I}_1\} \cup \{q_{ij} : ij \in \mathcal{I}_2\} \cup \{r_{ijk} : ijk \in \mathcal{I}_3\}, \tag{3.10a}$$

with sample-wise residuals

$$\begin{aligned} \rho_{a_1,n} &= W_1 x_n + b_1 - a_{1,n}, \\ \rho_{a_2,n} &= W_2 \sigma(a_{1,n}) + b_2 - a_{2,n}, \\ \rho_{d_1,i,n} &= W_1(:, i) - d_{1,i,n}, \\ \rho_{d_2,i,n} &= W_2[\sigma'(a_{1,n}) \odot d_{1,i,n}] - d_{2,i,n}, \\ \rho_{q,ij,n} &= W_2[\sigma''(a_{1,n}) \odot d_{1,i,n} \odot d_{1,j,n}] - q_{ij,n}, \\ \rho_{r,ijk,n} &= W_2[\sigma'''(a_{1,n}) \odot d_{1,i,n} \odot d_{1,j,n} \odot d_{1,k,n}] - r_{ijk,n}. \end{aligned} \tag{3.10b}$$

Thus the implementation stores both layer-activation mismatches, the two first-derivative mismatch families for t, x, y, the five mixed/pure second-derivative mismatches in (3.3), and the four spatial third-derivative mismatches in (3.3). No unused derivative or permutation duplicate is hidden in $\mathcal{S}$.

The concrete adaptive weights used for the third-order Navier–Stokes and MHD calculations can be written compactly as follows. At sample n, let

$$\begin{aligned} n_3 = \|W_3\|_F^2, \qquad n_2 = \|W_2\|_F^2, \qquad n_{1i} = \|W_1(:, i)\|_2^2, \\ u_{i,n} = \|d_{1,i,n}\|_2^2, \qquad v_{i,n} = \|d_{2,i,n}\|_2^2, \qquad z_{i,n} = \sum_{ij \in \mathcal{I}_2 :\, i \in ij} \|q_{ij,n}\|_2^2. \end{aligned} \tag{3.10c}$$

For the frozen operator, set $\kappa_{0,n} = \|A_{0,n}\|_F$, $\kappa_{i,n} = \|A_{i,n}\|_F$, $h_{ij,n} = \|A_{ij,n}\|_F$, and $g_{ijk,n} = \|A_{ijk,n}\|_F$. Define

$$H_{i,n} = \sum_{ij \in \mathcal{I}_2 :\, i \in ij} h_{ij,n}^2, \quad H_i^\infty = \max_n H_{i,n}, \quad H_{ij}^\infty = \max_n h_{ij,n}^2, \tag{3.10d}$$

where repeated indices are counted once in $H_{i,n}$, exactly as in the stored unordered derivative list. The sample envelopes used in the lower-order part are

$$\begin{aligned}
D_{a_2,n} &= \kappa_{0,n}^2 + \sum_i (\kappa_{i,n}^2 v_{i,n} + H_{i,n} z_{i,n}),\\
D_{a_1,n} &= D_{a_2,n} + \sum_i (\kappa_{i,n}^2 u_{i,n} + H_{i,n} u_{i,n} v_{i,n}),\\
D^{(1)}_{a_1 i,n} &= H_{i,n} u_{i,n},\\
D^{(2)}_{a_1 i,n} &= H_{i,n}(u_{i,n} + v_{i,n}),\\
D^{(1)}_{a_2 i,n} &= H_{i,n} v_{i,n},\\
D_{d_1 i,n} &= \kappa_{i,n}^2 + D^{(2)}_{a_1 i,n},\\
D_{d_2 i,n} &= \kappa_{i,n}^2 + D^{(1)}_{a_2 i,n}.
\end{aligned} \tag{3.10e}$$

Writing $E_{s,n} = \|\rho_{s,n}\|_2^2$, the implemented activation/first/mixed-second contribution is

$$\begin{aligned}
P_{\le 2} = \sum_n \Bigg\{ & \lambda^{\le 2}_{a_1,n} E_{a_1,n} + \lambda^{\le 2}_{a_2,n} E_{a_2,n}\\
& + \sum_i \left(\lambda^{\le 2}_{d_1 i,n} E_{d_1 i,n} + \lambda^{\le 2}_{d_2 i,n} E_{d_2 i,n} \right)\\
& + \sum_{ij \in \mathcal{I}_2} \lambda^{\le 2}_{q_{ij}} E_{q_{ij},n} \Bigg\},
\end{aligned} \tag{3.10f}$$

where

$$\begin{aligned}
\lambda^{\le 2}_{a_1,n} &= w_{a11} D_{a_1,n} + \sum_i (w_{a12,i} D^{(1)}_{a_1 i,n} + w_{a13,i} D^{(2)}_{a_1 i,n}),\\
\lambda^{\le 2}_{a_2,n} &= w_{a21} D_{a_2,n} + \sum_i w_{a22,i} D^{(1)}_{a_2 i,n},\\
\lambda^{\le 2}_{d_1 i,n} &= w_{d11,i} + w_{d12} D_{d_1 i,n},\\
\lambda^{\le 2}_{d_2 i,n} &= w_{d21,i} D_{d_2 i,n} + w_{d22,i},\\
\lambda^{\le 2}_{q_{ij}} &= n_3 H^{\infty}_{ij},
\end{aligned}$$

and

$$\begin{aligned}
w_{a11} &= n_3 n_2, & w_{a12,i} &= n_3 n_2 n_{1i},\\
w_{a13,i} &= n_3 n_2^2 n_{1i}, & w_{a21} &= n_3,\\
w_{a22,i} &= n_3 n_2 n_{1i}, & w_{d11,i} &= H_i^{\infty}(w_{a12,i} + w_{a13,i}),\\
w_{d12} &= w_{a11}, & w_{d21,i} &= n_3,\\
w_{d22,i} &= H_i^{\infty} w_{a12,i}.
\end{aligned} \tag{3.10g}$$

The third-order contribution is also explicit. Put

$$L_{i,n} = u_{i,n} + n_{1i}, \quad V_{i,n} = v_{i,n} + n_2 n_{1i}, \quad Z_{ij,n} = \|q_{ij,n}\|_2^2 + n_2 n_{1i} n_{1j}, \quad c_{ijk,n} = 15 g^2_{ijk,n}.$$

For a stored triple $\tau = (i,j,k)$, let $\mathcal{C}_\tau$ be the three position choices $(i;j,k),(j;i,k),(k;i,j)$ and let $\mathcal{P}_\tau$ be the three pair–remainder choices $(i,j;k),(i,k;j),(j,k;i)$, both as multisets so repeated indices retain their multiplicity. Define

$$
\begin{aligned}
\Xi_{\tau,n} &= \Gamma_{\tau,n}(E_{a_2,n} + n_2 E_{a_1,n}) \\
&\quad + \sum_{(a;b,c)\in\mathcal{C}_\tau} (V_{b,n}V_{c,n} + Z_{bc,n}) \\
&\qquad (E_{d_2 a,n} + n_2 E_{d_1 a,n} + n_2 L_{a,n} E_{a_1,n}) \\
&\quad + \sum_{(a,b;c)\in\mathcal{P}_\tau} V_{c,n}\{E_{q_{ab},n} + n_2 L_{a,n}L_{b,n}E_{a_1,n} \\
&\qquad + n_2 L_{b,n}E_{d_1 a,n} + n_2 L_{a,n}E_{d_1 b,n}\} \\
&\quad + E_{r_\tau,n} + n_2 L_{i,n}L_{j,n}L_{k,n}E_{a_1,n} \\
&\qquad + n_2\{L_{j,n}L_{k,n}E_{d_1 i,n} + L_{i,n}L_{k,n}E_{d_1 j,n} + L_{i,n}L_{j,n}E_{d_1 k,n}\}, \\
\Gamma_{\tau,n} &= V_{i,n}V_{j,n}V_{k,n} + Z_{ij,n}V_{k,n} + Z_{ik,n}V_{j,n} \\
&\quad + Z_{jk,n}V_{i,n} + n_2 n_{1i}n_{1j}n_{1k}.
\end{aligned}
\tag{3.10h}
$$

The implemented third-order penalty is

$$
P_3 = n_3 \sum_{n=1}^{N} \sum_{\tau\in\mathcal{I}_3} c_{\tau,n}\Xi_{\tau,n}, \qquad J_S = \frac{1}{N}\left(\sum_n \|\widehat{R}_n\|_2^2 + P_{\le 2} + P_3\right). \tag{3.10i}
$$

Equations (3.10a)–(3.10i) are the objective used in the final moderate-coupling MHD qualification and in the LySep side of the matched MHD comparison, with $\sigma(s) = \sin s$. The hard factors make boundary and initial rows identically satisfied, so no boundary term is added; fixing the pressure-output bias to zero removes only the explicit constant-bias degree of freedom; it does not fully fix the gauge $p(t,x) \mapsto p(t,x) + c(t)$, and it is not a pressure-loss term.

Theorem 3.2 (exact auxiliary recovery). Let $z(\theta)$ contain the actual forward preactivations and network derivatives through order three. Then

$$
J_S(\theta, z(\theta)) = J_{\mathrm{frozen}}(\theta). \tag{3.11}
$$

Proof. The forward definitions and Theorem 3.1 make every $\rho_s(\theta, z(\theta))$ zero. The same theorem gives $\widehat{D}^\alpha \Phi_\theta(z(\theta)) = D^\alpha \Phi_\theta$ for every $|\alpha| \le 3$. Hence the separated and original physical residuals agree pointwise. Any included fixed linear boundary or gauge row receives the same value and derivatives, so its contribution also agrees. Equality holds term by term; it is not an asymptotic or one-sided statement. □

C Full consistency argument and constants

Detail 3.4 Third-order consistency

Exact recovery identifies the chain manifold. The next result controls states away from it. The constant is deliberately state dependent.

At one sample, use bars for true network quantities and no bars for auxiliary quantities. Let $e_a = \bar{a}_2 - a_2$, $e_i = \bar{d}_{2,i} - d_{2,i}$, and similarly define $e_{q,ij}$ and $e_{r,ijk}$. Subtracting the recursion relations and using Lipschitz activation derivatives gives finite coefficients $g_{\beta s}$ such that

$$
\|e_\beta\| \le \sum_{s\in\mathcal{S}} g_{\beta s}\|\rho_s\|, \qquad \beta \in \{a_2, d_i, q_{ij}, r_{ijk}\}. \tag{3.12}
$$

For the third-order output, define

$$
\mathcal{T}(d_i, d_j, d_k) = d_i \odot d_j \odot d_k, \qquad \mathcal{M}(q, d) = q_{ij} \odot d_k + q_{ik} \odot d_j + q_{jk} \odot d_i. \tag{3.13}
$$

The true-minus-separated derivative error E_{ijk} admits the exact six-term decomposition

$$\begin{aligned} E_{ijk} = W_3\big[&(\sigma'''(\bar{a}_2) - \sigma'''(a_2)) \odot \mathcal{T}(\bar{d}_i, \bar{d}_j, \bar{d}_k) \\ &+ \sigma'''(a_2) \odot \{\mathcal{T}(\bar{d}_i, \bar{d}_j, \bar{d}_k) - \mathcal{T}(d_i, d_j, d_k)\} \\ &+ (\sigma''(\bar{a}_2) - \sigma''(a_2)) \odot \mathcal{M}(\bar{q}, \bar{d}) \\ &+ \sigma''(a_2) \odot \{\mathcal{M}(\bar{q}, \bar{d}) - \mathcal{M}(q, d)\} \\ &+ (\sigma'(\bar{a}_2) - \sigma'(a_2)) \odot \bar{r}_{ijk} + \sigma'(a_2) \odot e_{r,ijk}\big]. \end{aligned} \tag{3.14}$$

The two product differences are expanded by

$$abc - ABC = (a - A)bc + A(b - B)c + AB(c - C), \tag{3.15}$$

and

$$qd - QD = (q - Q)d + Q(d - D). \tag{3.16}$$

These identities expose every activation, d, q, and r mismatch without introducing a product auxiliary.

Theorem 3.3 (third-order LySep consistency). Assume bounded and Lipschitz activation derivatives on the relevant range, finite true and auxiliary state envelopes, bounded frozen coefficient matrices and included linear boundary/gauge rows, classical derivatives through order three, and positive penalties for every mismatch entering those rows. Then a finite, generally state-dependent coefficient constant $C_{\mathcal{L}}(\theta, z; b)$ exists such that

$$J_{\text{frozen}}(\theta) \le C_{\mathcal{L}}(\theta, z; b) J_S(\theta, z). \tag{3.17}$$

One admissible form is

$$C_{\mathcal{L}} = 2 \max\left\{1, \max_{s \in \mathcal{S}} \frac{M_s(\theta, z; b)}{\lambda_s}\right\}, \tag{3.18}$$

where the finite M_s satisfy

$$\|\mathcal{R}_{\text{true}} - \mathcal{R}_{\text{sep}}\|^2 \le \sum_{s \in \mathcal{S}} M_s \|\rho_s\|^2. \tag{3.19}$$

The result remains valid for vector outputs and matrix-valued A_α by using Frobenius coefficient envelopes.

Proof. The first- and second-order analogues of (3.14), together with (3.12) and the telescoping identities (3.15)–(3.16), yield finite statewise coefficients $h_{\alpha s}$ satisfying

$$\|D^\alpha \Phi_\theta - \widehat{D}^\alpha \Phi_\theta(z)\| \le \|W_3\|_F \sum_s h_{\alpha s}(\theta, z) \|\rho_s\|, \qquad |\alpha| \le 3. \tag{3.20}$$

The value case follows from Lipschitz continuity of σ. For a matrix coefficient,

$$\|A_\alpha E_\alpha\| \le \|A_\alpha\|_F \|E_\alpha\|. \tag{3.21}$$

Applying (3.20)–(3.21), the finite-sum Cauchy inequality, and the bounded coefficient envelopes to

$$\mathcal{R}_{\text{true}} - \mathcal{R}_{\text{sep}} = \sum_{|\alpha| \le 3} A_\alpha E_\alpha \tag{3.22}$$

gives (3.19); fixed linear boundary and gauge rows are handled identically. Finally,

$$\|\mathcal{R}_{\mathrm{true}}\|^2 \le 2\|\mathcal{R}_{\mathrm{sep}}\|^2 + 2\sum_s M_s\|\rho_s\|^2, \tag{3.23}$$

and $\sum_s M_s\|\rho_s\|^2 \le \max_s(M_s/\lambda_s)\sum_s \lambda_s\|\rho_s\|^2$ proves (3.17)–(3.18). □

The dependencies of $C_{\mathcal{L}}$ are explicit: frozen coefficient and boundary/gauge envelopes, activation bounds and Lipschitz constants, network layer norms, finite true and auxiliary d, q, r envelopes, derivative combinatorics, and penalty weights. It is not a universal constant independent of the state. Theorem 3.3 is a statewise residual-control result. By itself it gives neither a trajectory-uniform constant nor monotone decrease of the original frozen loss along the alternating updates; either conclusion would require controlled state envelopes and weights throughout the trajectory. Detail A.1 records the detailed lower-order mismatch decompositions.

For the concrete sine-activation objective (3.10), the activation and Lipschitz envelopes are at most one. The retained telescoping majorant has at most 18 squared terms per reconstructed derivative, while the coupled MHD residual has the 13 derivative families in (5.9). Consequently the actual nonnegative weights in (3.10f)–(3.10i) give the statewise specialization

$$\frac{1}{N}\sum_n \|\mathcal{R}_{\mathrm{true},n} - \mathcal{R}_{\mathrm{sep},n}\|_2^2 \le 13\cdot 18\,\frac{P_{\le 2}+P_3}{N},$$

and hence

$$J_{\mathrm{frozen}} \le 2\cdot 13\cdot 18\,J_S = 468 J_S. \tag{3.23a}$$

This is the bound evaluated in the final moderate-coupling MHD qualification. If an envelope or layer multiplier vanishes, the affected physical discrepancy vanishes with it; otherwise the corresponding displayed weight is positive. Thus zero adaptive coefficients do not leave a residual contribution uncontrolled. The same construction applies to the Navier–Stokes subset of the operator rows.

Detail A.1 Auxiliary mismatch propagation

For completeness, let bars denote exact network-chain values and unbarred symbols denote independent auxiliaries. The complete residual families at one sample, matching (3.10b), are

$$\begin{aligned}
\rho_{a_1} &= W_1 x + b_1 - a_1,\\
\rho_{a_2} &= W_2\sigma(a_1) + b_2 - a_2,\\
\rho_{d_1,i} &= W_1(:,i) - d_{1,i},\\
\rho_{d_2,i} &= W_2[\sigma'(a_1)\odot d_{1,i}] - d_{2,i},\\
\rho_{q_{ij}} &= W_2[\sigma''(a_1)\odot d_{1,i}\odot d_{1,j}] - q_{2,ij},\\
\rho_{r_{ijk}} &= W_2[\sigma'''(a_1)\odot d_{1,i}\odot d_{1,j}\odot d_{1,k}] - r_{2,ijk}.
\end{aligned} \tag{A.1}$$

For the true chain, $\bar a_1 = W_1 x + b_1$ and $\bar d_{1,i} = W_1(:,i)$. Subtraction gives

$$\begin{aligned}
\bar a_1 - a_1 &= \rho_{a_1},\\
\bar d_{1,i} - d_{1,i} &= \rho_{d_1,i},\\
\bar a_2 - a_2 &= W_2[\sigma'(\xi_1)\odot(\bar a_1 - a_1)] + \rho_{a_2},\\
\bar d_{2,i} - d_{2,i} &= W_2[(\sigma'(\bar a_1)-\sigma'(a_1))\odot \bar d_{1,i} + \sigma'(a_1)\odot(\bar d_{1,i} - d_{1,i})] + \rho_{d_2,i},\\
\bar q_{2,ij} - q_{2,ij} &= W_2[(\sigma''(\bar a_1)-\sigma''(a_1))\odot \bar d_{1,i}\odot \bar d_{1,j}\\
&\qquad + \sigma''(a_1)\odot(\bar d_{1,i}\odot \bar d_{1,j} - d_{1,i}\odot d_{1,j})] + \rho_{q_{ij}},\\
\bar r_{2,ijk} - r_{2,ijk} &= W_2[(\sigma'''(\bar a_1)-\sigma'''(a_1))\odot \bar d_{1,i}\odot \bar d_{1,j}\odot \bar d_{1,k}\\
&\qquad + \sigma'''(a_1)\odot(\bar d_{1,i}\odot \bar d_{1,j}\odot \bar d_{1,k} - d_{1,i}\odot d_{1,j}\odot d_{1,k})] + \rho_{r_{ijk}},
\end{aligned} \tag{A.2}$$

where ξ_1 is a componentwise mean-value point. Applying the two- and three-factor telescoping identities to the last two rows, together with bounded layer norms, finite d_1 factors, and Lipschitz activation derivatives, yields (3.12).

For the first output derivative, the exact difference is

$$E_i = W_3\big[(\sigma'(\bar{a}_2) - \sigma'(a_2)) \odot \bar{d}_i + \sigma'(a_2) \odot (\bar{d}_i - d_i)\big]. \tag{A.3}$$

For the second output derivative,

$$\begin{aligned} E_{ij} = W_3\big[&(\sigma''(\bar{a}_2) - \sigma''(a_2)) \odot \bar{d}_i \odot \bar{d}_j \\ &+ \sigma''(a_2) \odot \{(\bar{d}_i - d_i) \odot \bar{d}_j + d_i \odot (\bar{d}_j - d_j)\} \\ &+ (\sigma'(\bar{a}_2) - \sigma'(a_2)) \odot \bar{q}_{ij} + \sigma'(a_2) \odot (\bar{q}_{ij} - q_{ij})\big]. \end{aligned} \tag{A.4}$$

Equations (3.14)–(3.16) give the third-order counterpart. Applying componentwise product bounds to (A.2)–(A.4) and (3.14) yields (3.20). For a finite set of derivative rows, bounded matrix coefficient envelopes then produce finite M_s in (3.19). Equations (3.10c)–(3.10i) instantiate this construction for the implemented sine network; the 13-row, 18-term specialization gives (3.23a). This remains a statewise bound, not a uniform assertion over an unbounded parameter space.

D Output least squares, quartic obstruction and outer correctness

Detail 2.2 Bilinear nonlinear PDEs

Consider the abstract nonlinear equation

$$Lz + \mathcal{B}(z, z) = f, \tag{2.3}$$

where L is linear and $\mathcal{B}$ is bilinear. If z and the required derivative of z are both affine in the output-layer parameters, then $\mathcal{B}(z, z)$ is quadratic and its squared residual is generically quartic. Introducing direct product-factor auxiliaries does not remove this issue if their consistency weights depend honestly on current unknown factors.

We instead freeze one bilinear argument at outer step n:

$$Lz^{n+1} + \mathcal{B}(z^n, z^{n+1}) = f. \tag{2.4}$$

The old state z^n is data throughout the inner optimization. Equation (2.4) is therefore a linear variable-coefficient PDE for z^{n+1}, compatible with LySep. Different applications may use another explicitly stated one-factor ordering; the essential property is linear dependence on the new state. This separation creates two logically distinct levels: an outer nonlinear fixed-point map and an inner separated approximation of its frozen linear problem.

Detail 3.5 Conditional quadratic output-layer block

Proposition 3.4 (conditional quadratic output block). Fix the outer coefficients, hidden-layer variables, and all auxiliary states a, d, q, r. If each self-adaptive consistency contribution involving W_3 has the form

$$\|W_3\|_F^2 c_s \|\rho_s\|^2, \qquad c_s \geq 0, \tag{3.24}$$

with c_s and ρ_s fixed in this subproblem, then the subproblem in (W_3, b_3) is a positive-semidefinite quadratic least-squares-type problem.

Proof. With hidden states fixed, $\Phi_\theta = W_3\sigma(a_2) + b_3$ is affine in (W_3, b_3). Every positive-order derivative in Theorem 3.1 is W_3 multiplying a fixed feature and is therefore linear in W_3. The frozen

physical residual and each fixed linear boundary or gauge row consequently have the form $Hv - y$, where $v = \mathrm{vec}(W_3, b_3)$. Their squared norms are positive-semidefinite quadratics. Each term in (3.24) is a nonnegative ridge-like quadratic because its final two factors are fixed. Their sum remains quadratic. Rank deficiency from pressure gauge freedom may make the minimizer nonunique, in which case a minimum-norm least-squares representative is valid. Simultaneous variation of hidden states, auxiliaries, or outer coefficients invalidates the fixed-design argument; the complete LySep objective is not asserted to be convex. □

The implemented schedule realizes the proposition rather than merely assuming it. The envelopes and weights in (3.10c)–(3.10h) are recomputed from the current state at every objective, gradient, and backtracking-candidate evaluation. Therefore the nonlinear block gradients include the dependence of the adaptive weights on the block being varied. For one r_τ update, its coefficients are independent of that r_τ and the block is quadratic. For the output block, every term in $P_{\le 2} + P_3$ contains exactly one factor $n_3 = \|W_3\|_F^2$ and no other W_3 dependence. With all other states fixed,

$$J_S(W_3, b_3) = \frac{1}{N}\left(\|D\,\mathrm{vec}(W_3) - y(b_3)\|_2^2 + \Lambda_{\mathrm{ad}}\|W_3\|_F^2\right), \qquad \Lambda_{\mathrm{ad}} \ge 0, \tag{3.25}$$

where (3.10f)–(3.10i) give the algebraic factorization $P_{\le 2} + P_3 = n_3\Lambda_{\mathrm{ad}}$; Λ_{ad} is assembled after factoring out n_3, so no division by $\|W_3\|_F^2$ is required, and it is fixed in this block. The update is therefore the minimum-norm solution of the augmented design $[D; \sqrt{\Lambda_{\mathrm{ad}}}I]$. No $\|W_3\|_F^4$ term is present.

Detail 4.1 Direct-split quartic obstruction

The obstacle can be seen in one scalar output coordinate. Let $u(w) = wh$, $h \neq 0$, where w is an output-layer parameter, and let U be a fixed auxiliary. If an honest current-factor weight is

$$\omega(w) = |u(w)|^2 = w^2h^2, \tag{4.1}$$

then the weighted consistency term is

$$\omega(w)|u(w) - U|^2 = h^4w^4 - 2h^3Uw^3 + h^2U^2w^2. \tag{4.2}$$

Proposition 4.1 (generic dynamic-factor obstruction). If a quadratic consistency mismatch is weighted by the square of a nonzero factor that is linear in the current output-layer variables, its contribution is generically quartic in those variables and the conditional quadratic structure of Proposition 3.4 is lost.

Proof. Equation (4.2) has nonzero leading coefficient h^4. The same calculation applies to a derivative factor whenever its output-layer feature is nonzero. □

This proposition is deliberately limited to generic dynamic-factor weighting. It does not state that every direct nonlinear formulation is necessarily quartic. Fixed penalties retain a quadratic output block but yield consistency constants that depend on current factor bounds. Hard a priori bounds could also retain the block, but enforcing them would add machinery not used here. Lagging evaluates the factor at the preceding outer state and therefore makes it genuine fixed PDE data during the inner solve.

Detail 4.2 Frozen linear inner problem

For (2.3), the adopted iteration is (2.4). At outer step n, the operator

$$\mathcal{L}_{z^n} w = Lw + \mathcal{B}(z^n, w) \tag{4.3}$$

is linear in w. Its coefficient fields define matrices $A_\alpha(x; z^n)$ in (3.8). The inner loss is exactly the third-order separated loss $J_S(\theta, z; z^n)$, using only the derivative auxiliaries required by the frozen operator. After an accepted inner solve, the reconstructed physical state becomes the next frozen state. The old factor is never jointly optimized or silently detached inside the inner objective.

For viscous Burgers,

$$(u^{n+1})_t + u^n(u^{n+1})_x - \nu(u^{n+1})_{xx} = f. \tag{4.4}$$

For Navier–Stokes, the inertial term becomes the Oseen term $(u^n \cdot \nabla)u^{n+1}$. For MHD, the chosen momentum ordering freezes the first factors of velocity advection and magnetic tension, while the selected scalar induction equation freezes the velocity in $u^n \cdot \nabla A^{n+1}$. These inner problems are linear variable-coefficient PDEs, although the outer map remains nonlinear.

Algorithm D.1 (implemented lagged third-order LySep iteration). The following is the update contract used for the final moderate-coupling MHD qualification and for the LySep branch of the matched MHD comparison.

1. Initialize the network from the prescribed seed and set the four frozen training fields $(u_1^0, u_2^0, B_1^0, B_2^0)$ to zero.
2. For outer index $m = 0, \ldots, 4$, retain the current network parameters as a warm start, refresh every a, d, q, r auxiliary to its exact chain value for that network, assemble the 13 frozen coefficient matrices from (u^m, B^m), and reset all stored trial step sizes to one.
3. Perform 1000 complete inner sweeps unless a nonfinite state is encountered. One sweep updates, in this order,

   $$W_1(:,t), W_1(:,x), W_1(:,y); \quad b_1^{\mathrm{net}}; \quad a_1; \quad d_{1,t}, d_{1,x}, d_{1,y};$$
   $$W_2; \quad b_2^{\mathrm{net}}; \quad a_2; \quad d_{2,t}, d_{2,x}, d_{2,y};$$
   $$q_{tx}, q_{ty}, q_{xx}, q_{xy}, q_{yy};$$
   $$r_{xxx}, r_{xxy}, r_{xyy}, r_{yyy}; \quad W_3; \quad b_3^{\mathrm{net}}.$$

   The $W_1, a_1, d_1, W_2, a_2, d_2, q$ blocks use the full current-weight gradient of (3.10) and nonincrease backtracking. A block starts from $\min(1, 2s_{\mathrm{previous}})$, halves at most 100 times, and accepts when

   $$J_{\mathrm{candidate}} \le J_{\mathrm{current}} + 100\ \mathrm{eps}(\max\{|J_{\mathrm{current}}|, \mathrm{realmin}\}).$$

   The two hidden biases use their weighted least-squares updates. Each r block uses its exact quadratic line minimizer and is accepted under the same roundoff-scaled nonincrease rule. The W_3 block uses the minimum-norm augmented least-squares solve in (3.25). The final bias update is a minimum-norm physical-residual correction, after which the pressure-output bias is set to zero to remove only the explicit constant bias degree of freedom. This does not fully fix the pressure gauge $p \mapsto p + c(t)$, since the remaining network can represent time-only functions; pressure accuracy is evaluated through its spatial gradient.
4. At the end of the sweep budget, evaluate the last frozen inner physical loss and reconstruct the new training fields. The first LySep outer solve must reduce that frozen loss by at least 1%; a nonfinite state or failure of this qualification check terminates the run. Otherwise compute the outer defect defined in (6.3e). Stop when it is at most 10^{-3}; if it is larger, freeze the newly reconstructed fields and continue. Thus auxiliaries are reinitialized exactly at each new frozen problem, whereas the trained network is warm-started.

The reported final moderate-coupling MHD qualification run completed all 1000 sweeps in each of four outer solves. No unreported inner convergence tolerance or alternative block order was used.

Detail 4.3 Fixed-point recovery

Proposition 4.2 (nonlinear fixed-point recovery). Any exact fixed point $z^{n+1} = z^n = z^*$ of (2.4) satisfies the original nonlinear equation (2.3). The same conclusion holds for an application-specific one-factor ordering whenever its old and new factors recover the original bilinear term when they coincide.

Proof. Substitute $z^{n+1} = z^n = z^*$ into (2.4) to obtain $Lz^* + \mathcal{B}(z^*, z^*) = f$. □

This statement separates fixed-point correctness from iteration convergence. In Burgers, $u^n(u^{n+1})_x$ becomes $u^* u_x^*$. In Navier–Stokes, the Oseen term recovers $(u^* \cdot \nabla)u^*$. In MHD, the selected

momentum and scalar induction equations recover the nonlinear momentum and scalar potential induction equations; taking the two-dimensional curl of the latter gives the vector induction equation under the assumptions of Appendix E (scalar induction).

Detail 4.4 Conditional local contraction

Let $S(b)$ denote a uniquely selected solution of the frozen linear problem on a closed, complete, bounded set K of states.

Proposition 4.3 (sufficient local contraction). Suppose $S(K) \subseteq K$ and

$$\|S(b_1) - S(b_2)\|_X \le C_{\mathrm{lin}} C_{\mathrm{bil}} \|b_1 - b_2\|_X \tag{4.5}$$

for all $b_1, b_2 \in K$, where C_{lin} is a uniform bound for the selected inverse frozen solve and C_{bil} bounds the operator perturbation induced by the frozen bilinear factor, including the bounded new-state factor. If

$$q = C_{\mathrm{lin}} C_{\mathrm{bil}} < 1, \tag{4.6}$$

then S is contractive on K, has a unique fixed point there, and exact outer iterates initialized in K converge to it.

Proof. Writing the frozen equation as $\mathcal{A}(b)S(b) = f$ and subtracting two instances gives

$$\mathcal{A}(b_1)\{S(b_1) - S(b_2)\} = \{\mathcal{A}(b_2) - \mathcal{A}(b_1)\}S(b_2). \tag{4.7}$$

The inverse and coefficient-perturbation bounds give (4.5). Equation (4.6), invariance of K, and the Banach fixed-point theorem give the conclusion. □

For inexact inner solves with error ε_n, the corresponding conditional recurrence is

$$\|z^{n+1} - z^*\|_X \le q\|z^n - z^*\|_X + \varepsilon_n. \tag{4.8}$$

No PDE-specific global value of C_{lin} or C_{bil} is established here. Proposition 4.3 is a sufficient local result, not a theorem of global Picard, Navier–Stokes, or MHD convergence.

E Complete NS/MHD formulations and harmonic compensation

Detail 2.3 Incompressible NS and resistive MHD

On $Q = (0, T] \times \Omega$, $\Omega = [-1, 1]^2$, the incompressible Navier–Stokes equations are

$$u_t + (u \cdot \nabla)u + \nabla p - \nu \Delta u = f_u, \qquad \nabla \cdot u = 0, \tag{2.5}$$

with viscosity $\nu > 0$. The resistive incompressible MHD system considered here is

$$\begin{aligned} u_t + (u \cdot \nabla)u - (B \cdot \nabla)B + \nabla p - \nu \Delta u &= f_u, \\ B_t + (u \cdot \nabla)B - (B \cdot \nabla)u - \eta \Delta B &= f_B, \\ \nabla \cdot u = 0, \qquad \nabla \cdot B &= 0, \end{aligned} \tag{2.6}$$

where $\eta > 0$. Pressure is determined only modulo a spatial constant $c(t)$; all pressure comparisons below use its gradient. The two-dimensional hard-solenoidal representations

$$u = \mathrm{curl}_{2D}\, \psi = (\psi_y, -\psi_x), \qquad B = \mathrm{curl}_{2D}\, A = (A_y, -A_x) \tag{2.7}$$

make both divergence identities exact whenever mixed derivatives commute. Appendices A–D give the general third-order nonlinear LySep theory, and Appendix E specializes it to (2.5)–(2.6).

Detail 5.1 Streamfunction Navier–Stokes

Let

$$u = \operatorname{curl}_{2D} \psi = (\psi_y, -\psi_x). \tag{5.1}$$

Lemma 5.1 (hard velocity solenoidality and no-slip boundary). If mixed derivatives commute, then $\nabla \cdot u = 0$ pointwise. On $\Omega = [-1, 1]^2$, set

$$P(s) = (1 - s^2)^2, \qquad H(x, y) = P(x)P(y), \qquad \psi(t, x, y) = tH(x, y)\phi_\psi(t, x, y). \tag{5.2}$$

Then $u = 0$ on $\partial\Omega$.

Proof. Equation (5.1) gives $\nabla \cdot u = \psi_{yx} - \psi_{xy} = 0$. At $s = \pm 1$, both $P(s)$ and $P'(s)$ vanish. Hence $H = H_x = H_y = 0$ on the spatial boundary. Both components in (5.1) therefore vanish there. □

Hard solenoidality and no slip require no divergence or boundary penalty. The representation does, however, increase derivative order: the primitive viscous term Δu contains derivatives of ψ one order higher than the velocity derivatives.

Detail 5.2 Chi-pressure harmonic defect

A seemingly economical second-order formulation introduces

$$\chi = \Delta\psi \tag{5.3}$$

as an independent output and enforces (5.3) in L^2. For frozen velocity $b = (b_1, b_2)$, its residuals are

$$\begin{aligned} R_1 &= \psi_{yt} + b_1\psi_{xy} + b_2\psi_{yy} - \nu\chi_y + p_x - f_1, \\ R_2 &= -\psi_{xt} - b_1\psi_{xx} - b_2\psi_{xy} + \nu\chi_x + p_y - f_2, \\ R_3 &= \chi - \Delta\psi. \end{aligned} \tag{5.4}$$

Lemma 5.2 (chi-pressure harmonic compensation on a time slice). For each fixed t, hold $\psi(t, \cdot)$ and $b(t, \cdot)$ fixed and let

$$\widehat{\chi}(t, \cdot) = \chi(t, \cdot) + h(t, \cdot), \qquad \widehat{p}(t, \cdot) = p(t, \cdot) + \pi(t, \cdot), \tag{5.5}$$

where

$$\Delta h = 0, \qquad \pi_x = \nu h_y, \qquad \pi_y = -\nu h_x. \tag{5.6}$$

Then $\widehat{R}_1 = R_1$, $\widehat{R}_2 = R_2$, and $\widehat{R}_3 = R_3 + h$. On each spatial time slice, $\pi(t, \cdot)$ satisfying (5.6) exists locally precisely for harmonic $h(t, \cdot)$; on a simply connected domain it exists globally and is determined up to the pressure gauge $c(t)$.

Proof. Fix t. The momentum changes are $-\nu h_y + \pi_x = 0$ and $\nu h_x + \pi_y = 0$, while the coupling row changes by h. Compatibility of the two spatial equations for π requires

$$(\pi_x)_y = (\pi_y)_x \iff \nu h_{yy} = -\nu h_{xx} \iff \Delta h = 0. \tag{5.7}$$

The local potential criterion and, globally, simple connectedness complete the argument on that time slice; the undetermined spatial constant may depend on t. □

The coupling row prevents an exact nullspace of the complete loss, so the precise designation is a **chi-pressure harmonic near-nullspace**. The order mismatch is important: R_3 directly controls $e = \chi - \Delta\psi$ in L^2, while momentum contains ∇e. No uniform constant satisfies $\|\nabla e\|_{L^2} \le C\|e\|_{L^2}$

over all smooth states; oscillatory or harmonic sequences can have arbitrarily large derivative-to-value ratios. Increasing a finite L^2 weight does not change this Sobolev order. This diagnosis applies to (5.4), not to every possible second-order mixed, weak, stabilized, or vorticity formulation.

Detail 5.3 Third-order chi-free NS formulation

Substituting $\chi = \Delta\psi$ into momentum removes the independent compensation variable and gives

$$\boxed{\begin{aligned} R_1^{(3)} &= \psi_{yt} + b_1\psi_{xy} + b_2\psi_{yy} - \nu(\psi_{xxy} + \psi_{yyy}) + p_x - f_1, \\ R_2^{(3)} &= -\psi_{xt} - b_1\psi_{xx} - b_2\psi_{xy} + \nu(\psi_{xxx} + \psi_{xyy}) + p_y - f_2. \end{aligned}} \tag{5.8}$$

The signs follow from $u = (\psi_y, -\psi_x)$: $\Delta u_1 = \psi_{xxy} + \psi_{yyy}$ and $\Delta u_2 = -(\psi_{xxx} + \psi_{xyy})$. Thus the strong primitive formulation requires the four third spatial derivatives in (3.3). The increase from order two to order three is not introduced merely to raise the formal order of the method. It is the direct mathematical consequence of combining hard divergence-free primitive velocity, viscous momentum, and elimination of the diagnosed χ channel.

The network output is only $\Phi_\theta = (\phi_\psi, \phi_p)^T$, with $\psi = tH\phi_\psi$ and $p = \phi_p$. For fixed b, (5.8) is a 2×2 matrix-valued linear operator over the derivative index set

$$\mathcal{I} = \{0, t, x, y, tx, ty, xx, xy, yy, xxx, xxy, xyy, yyy\}. \tag{5.9}$$

All coefficient matrices depend only on the sample, H and its analytic derivatives, ν, and the frozen velocity. Theorem 3.3 therefore applies directly. Pressure enters only through p_x, p_y, and no pressure boundary condition or pressure-gradient regularizer is added.

Detail 5.4 Magnetic potential formulation

For MHD, define

$$B = \operatorname{curl}_{2D} A = (A_y, -A_x), \qquad A = tK(x,y)\phi_A, \qquad K = (1-x^2)(1-y^2). \tag{5.10}$$

Mixed-derivative symmetry gives $\nabla \cdot B = 0$ identically. Since $K = 0$ on each boundary edge, $A = 0$ there. Its tangential derivative is also zero along the constant boundary trace, so $B \cdot n = 0$. The normal derivative of A need not vanish; the tangential magnetic field is not artificially set to zero. The scalar Dirichlet condition $A = 0$ is the sole magnetic boundary condition in the adopted potential equation and fixes the gauge $A \mapsto A + C(t)$.

With $\psi = tH\phi_\psi$, $A = tK\phi_A$, and $p = \phi_p$, the shared-trunk vector output is

$$\Phi_\theta = (\phi_\psi, \phi_A, \phi_p)^T. \tag{5.11}$$

Velocity and magnetic components are derived quantities rather than independent outputs. The velocity branch uses third derivatives through viscosity; the magnetic branch uses at most second derivatives under the scalar induction equation; pressure uses first derivatives. Consequently the coupled system remains within the third-order theory of Appendices A–C.

Detail 5.5 Scalar magnetic-potential induction

Proposition 5.3 (scalar-to-vector induction identity). Let $B = \operatorname{curl}_{2D} A$, let $\nabla \cdot u = 0$, and let η be spatially constant. Then

$$\operatorname{curl}_{2D}(A_t + u \cdot \nabla A - \eta\Delta A) = B_t + (u \cdot \nabla)B - (B \cdot \nabla)u - \eta\Delta B. \tag{5.12}$$

Proof. The two-dimensional curl commutes with ∂_t and Δ. Write $u = (u_1, u_2)$ and $B = (A_y, -A_x)$. The first transport component is

$$\partial_y(u_1 A_x + u_2 A_y) = u_1 A_{xy} + u_2 A_{yy} + u_{1,y} A_x + u_{2,y} A_y. \tag{5.13}$$

Using $u_{1,x} + u_{2,y} = 0$, (5.13) equals

$$u \cdot \nabla B_1 - (B \cdot \nabla) u_1. \tag{5.14}$$

Similarly,

$$-\partial_x(u_1 A_x + u_2 A_y) = u \cdot \nabla B_2 - (B \cdot \nabla) u_2. \tag{5.15}$$

Adding the commuting time and diffusion terms proves (5.12). □

We select the scalar equation

$$A_t + u \cdot \nabla A - \eta \Delta A = g_A, \qquad f_B = \operatorname{curl}_{2D} g_A, \tag{5.16}$$

with a fixed scalar-source representative. The vector source determines g_A only modulo a function of time; manufactured data below define a compatible representative directly. At outer step n, the MHD inner residuals are

$$\begin{aligned} R_x &= \psi_{yt} + c_1 \psi_{xy} + c_2 \psi_{yy} - d_1 A_{xy} - d_2 A_{yy} + p_x - \nu(\psi_{xxy} + \psi_{yyy}) - f_{u,1}, \\ R_y &= -\psi_{xt} - c_1 \psi_{xx} - c_2 \psi_{xy} + d_1 A_{xx} + d_2 A_{xy} + p_y + \nu(\psi_{xxx} + \psi_{xyy}) - f_{u,2}, \\ R_A &= A_t + c_1 A_x + c_2 A_y - \eta(A_{xx} + A_{yy}) - g_A, \end{aligned} \tag{5.17}$$

where $c = u^n$ and $d = B^n$ are fixed throughout the inner solve. Curling the intermediate scalar induction equation produces

$$B_t^{n+1} + (u^n \cdot \nabla) B^{n+1} - (B^{n+1} \cdot \nabla) u^n - \eta \Delta B^{n+1} = f_B. \tag{5.18}$$

This differs at intermediate Picard states from a primitive vector linearization that uses $-(B^n \cdot \nabla) u^{n+1}$. Both recover the same nonlinear induction equation when old and new states coincide. The distinction is retained in the interpretation of the numerical results.

F HARD-FACTOR EXPANSIONS AND COMPLETE MATRIX OPERATOR

DETAIL A.2 HARD-FACTOR DERIVATIVE EXPANSIONS

The clamped factor satisfies

$$P'(s) = 4s(s^2 - 1), \qquad P''(s) = 12s^2 - 4, \qquad P'''(s) = 24s. \tag{A.5}$$

Writing $\psi = tH\phi$ with $H = P(x)P(y)$, the third derivatives in (5.8) are

$$\begin{aligned} \psi_{xxx} &= t(H_{xxx}\phi + 3H_{xx}\phi_x + 3H_x\phi_{xx} + H\phi_{xxx}), \\ \psi_{xxy} &= t(H_{xxy}\phi + H_{xx}\phi_y + 2H_{xy}\phi_x + 2H_x\phi_{xy} + H_y\phi_{xx} + H\phi_{xxy}), \\ \psi_{xyy} &= t(H_{xyy}\phi + H_{yy}\phi_x + 2H_{xy}\phi_y + 2H_y\phi_{xy} + H_x\phi_{yy} + H\phi_{xyy}), \\ \psi_{yyy} &= t(H_{yyy}\phi + 3H_{yy}\phi_y + 3H_y\phi_{yy} + H\phi_{yyy}). \end{aligned} \tag{A.6}$$

For $A = tK\varphi$ and $K = (1 - x^2)(1 - y^2)$,

$$
\begin{aligned}
A_t &= K\varphi + tK\varphi_t, \\
A_x &= t(K_x\varphi + K\varphi_x), \qquad A_y = t(K_y\varphi + K\varphi_y), \\
A_{xx} &= t(K_{xx}\varphi + 2K_x\varphi_x + K\varphi_{xx}), \\
A_{xy} &= t(K_{xy}\varphi + K_x\varphi_y + K_y\varphi_x + K\varphi_{xy}), \\
A_{yy} &= t(K_{yy}\varphi + 2K_y\varphi_y + K\varphi_{yy}).
\end{aligned} \tag{A.7}
$$

DETAIL A.3 CONCRETE FROZEN MHD OPERATOR

For reproducibility, this subsection lists the operator used by both the final moderate-coupling MHD qualification and the matched MHD comparison. Let $c = (c_1, c_2) = u^m$ and $d = (d_1, d_2) = B^m$ denote the frozen sample-wise fields, and let the network output be $\Phi = (\phi_\psi, \phi_A, \phi_p)^T$. In addition to the derivatives of H in (A.5)–(A.6),

$$
\begin{aligned}
K_x = -2x(1-y^2), \quad K_y = -2y(1-x^2), \\
K_{xx} = -2(1-y^2), \quad K_{xy} = 4xy, \quad K_{yy} = -2(1-x^2).
\end{aligned} \tag{A.8}
$$

With omitted entries equal to zero, the value and first-derivative matrices in

$$
R = \sum_{\alpha \in \mathcal{I}} A_\alpha D^\alpha \Phi - F
$$

are

$$
A_0 = \begin{bmatrix}
H_y + t(c_1 H_{xy} + c_2 H_{yy}) & -t(d_1 K_{xy} + d_2 K_{yy}) & 0 \\
-\nu t(H_{xxy} + H_{yyy}) & & \\
-H_x - t(c_1 H_{xx} + c_2 H_{xy}) & t(d_1 K_{xx} + d_2 K_{xy}) & 0 \\
+\nu t(H_{xxx} + H_{xyy}) & & \\
0 & K + t(c_1 K_x + c_2 K_y) - \eta t(K_{xx} + K_{yy}) & 0
\end{bmatrix}, \tag{A.9}
$$

$$
A_t = \begin{bmatrix} tH_y & 0 & 0 \\ -tH_x & 0 & 0 \\ 0 & tK & 0 \end{bmatrix},
$$

$$
A_x = \begin{bmatrix}
tc_1 H_y - 2\nu t H_{xy} & -td_1 K_y & 1 \\
-H - 2tc_1 H_x - tc_2 H_y + \nu t(3H_{xx} + H_{yy}) & t(2d_1 K_x + d_2 K_y) & 0 \\
0 & t(c_1 K - 2\eta K_x) & 0
\end{bmatrix},
$$

$$
A_y = \begin{bmatrix}
H + tc_1 H_x + 2tc_2 H_y - \nu t(H_{xx} + 3H_{yy}) & -t(d_1 K_x + 2d_2 K_y) & 0 \\
-tc_2 H_x + 2\nu t H_{xy} & td_2 K_x & 1 \\
0 & t(c_2 K - 2\eta K_y) & 0
\end{bmatrix}. \tag{A.10}
$$

The mixed/pure second-derivative matrices are

$$
A_{tx} = \begin{bmatrix} 0 & 0 & 0 \\ -tH & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad A_{ty} = \begin{bmatrix} tH & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},
$$

$$
A_{xx} = \begin{bmatrix} -\nu t H_y & 0 & 0 \\ -tc_1 H + 3\nu t H_x & td_1 K & 0 \\ 0 & -\eta t K & 0 \end{bmatrix},
$$

$$
A_{xy} = \begin{bmatrix} tc_1 H - 2\nu t H_x & -td_1 K & 0 \\ -tc_2 H + 2\nu t H_y & td_2 K & 0 \\ 0 & 0 & 0 \end{bmatrix},
$$

$$A_{yy} = \begin{bmatrix} tc_2H - 3\nu tH_y & -td_2K & 0 \\ \nu tH_x & 0 & 0 \\ 0 & -\eta tK & 0 \end{bmatrix}. \tag{A.11}$$

The four third-derivative matrices are

$$\begin{aligned} A_{xxx} &= \begin{bmatrix} 0 & 0 & 0 \\ \nu tH & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad A_{xxy} = \begin{bmatrix} -\nu tH & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\ A_{xyy} &= \begin{bmatrix} 0 & 0 & 0 \\ \nu tH & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad A_{yyy} = \begin{bmatrix} -\nu tH & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned} \tag{A.12}$$

Direct expansion of (A.6)–(A.7) in (5.17) gives (A.9)–(A.12), so these matrices and the sources (6.3b) uniquely specify the physical objective used in the MHD qualification and matched comparison. For derivative reconstruction errors e_α,

$$\left\| \sum_{\alpha \in \mathcal{I}} A_\alpha e_\alpha \right\|_2^2 \le 13 \sum_{\alpha \in \mathcal{I}} \|A_\alpha\|_F^2 \|e_\alpha\|_2^2, \tag{A.13}$$

which is the coupled MHD specialization used in Theorem 3.3.

G Manufactured forcing, metrics, hyperparameters and training protocol

Detail 6.1 Experimental setup

The experiments are qualifications of the derived mechanism rather than a claim of broad benchmark coverage. All reported calculations used double precision in MATLAB R2024b. Training and test points were fixed, non-overlapping Halton samples in the relevant space–time box; no result was selected from a parameter sweep. Every network had exactly the three affine maps in (2.1), equal hidden widths, and sine activation. Algorithm D.1 gives the complete LySep update order. The LySep runs in the final MHD qualification and matched MHD comparison used 1000 full inner sweeps per accepted frozen problem, at most five outer solves, and outer tolerance 10^{-3}.

The one-dimensional Burgers benchmark used $Q = (0, 0.5] \times [-1, 1]$, $\nu = 0.1$, and

$$u^*(t, x) = 0.1\, te^{-t}(1 - x^2) \sin x. \tag{6.1}$$

It used width $M = 30$, 1000 training points, 350 test points, and seed 1. The Navier–Stokes formulation study used $Q = (0, 0.25] \times [-1, 1]^2$, $\nu = 0.1$, width $M = 40$, 2000 training points, 500 test points, seed 1, and

$$\psi^* = 0.05\, te^{-t} P(x)P(y), \qquad p^* = 0.02\, te^{-t} xy. \tag{6.2}$$

For the flagship MHD case, $T = 1$, $\nu = \eta = 0.1$, $M = 40$, $N_{\text{train}} = 2000$, $N_{\text{test}} = 500$, and

$$\begin{aligned} \psi^* &= 0.75\, te^{-t} P(x)P(y), \\ A^* &= 0.40\, te^{-t}(1 - x^2)(1 - y^2)(1 + 0.5x), \\ p^* &= 0.02\, te^{-t} xy. \end{aligned} \tag{6.3}$$

The sources are defined without numerical differentiation. For Burgers,

$$f_{\text{Burg}} = u_t^* + u^* u_x^* - \nu u_{xx}^*. \tag{6.3a}$$

For Navier–Stokes, $u^* = (\psi_y^*, -\psi_x^*)$ is inserted into (2.5) to define both components of f_u. For MHD, set $u^* = (\psi_y^*, -\psi_x^*)$ and $B^* = (A_y^*, -A_x^*)$ and evaluate

$$\begin{aligned} F_1 &= u_{1,t}^* + u_1^* u_{1,x}^* + u_2^* u_{1,y}^* - B_1^* B_{1,x}^* - B_2^* B_{1,y}^* + p_x^* - \nu(u_{1,xx}^* + u_{1,yy}^*), \\ F_2 &= u_{2,t}^* + u_1^* u_{2,x}^* + u_2^* u_{2,y}^* - B_1^* B_{2,x}^* - B_2^* B_{2,y}^* + p_y^* - \nu(u_{2,xx}^* + u_{2,yy}^*), \\ F_3 &= A_t^* + u_1^* A_x^* + u_2^* A_y^* - \eta(A_{xx}^* + A_{yy}^*). \end{aligned} \tag{6.3b}$$

Equations (6.1)–(6.3), the polynomial hard factors, and their analytic derivatives therefore determine every source value. The asymmetric factor $1 + 0.5x$ prevents geometric alignment of the velocity and magnetic potentials.

For the final MHD qualification and matched comparison, the frozen residual vector is exactly (5.17), or equivalently

$$\widehat{R}_n = \sum_{\alpha \in \mathcal{I}} A_{\alpha,n}(u^m, B^m) \widehat{D}^\alpha \Phi_n - F_n, \qquad J_{\text{last frozen}} = \frac{1}{N} \sum_{n=1}^{N} \|R_n\|_2^2, \tag{6.3c}$$

with the 13 concrete matrices listed in Appendix F. The LySep objective reconstructs the derivatives and adds (3.10f)–(3.10i); the end-to-end objective uses the true network derivatives and no auxiliary penalty. Hard initial/boundary factors make their corresponding conditions identities. Neither objective adds a divergence, boundary, pressure-boundary, pressure-gradient, or gauge penalty. The reported scalar $J_{\text{last frozen}}$ is the last frozen inner physical loss at each method's own final outer state.

All relative metrics use the fixed test samples and the discrete norm $\|v\|_Q^2 = \sum_n \|v_n\|_2^2$. Thus

$$\begin{aligned} e_u &= \|u - u^*\|_Q / \|u^*\|_Q, & e_B &= \|B - B^*\|_Q / \|B^*\|_Q, \\ e_{\text{mom}} &= \|(R_1^{\text{NL}}, R_2^{\text{NL}})\|_Q / \max(\|(F_1, F_2)\|_Q, 10^{-14}), & e_{\text{ind}} &= \|R_A^{\text{NL}}\|_Q / \max(\|F_3\|_Q, 10^{-14}), \\ e_{\nabla p} &= \|\nabla p - \nabla p^*\|_Q / \|\nabla p^*\|_Q. \end{aligned} \tag{6.3d}$$

Here R^{NL} is evaluated with the learned current fields in every bilinear factor, not with frozen coefficients. The Burgers metrics use the same discrete norm:

$$e_{u,\text{Burg}} = \|u - u^*\|_Q / \|u^*\|_Q, \qquad e_{R,\text{Burg}} = \|u_t + u u_x - \nu u_{xx} - f_{\text{Burg}}\|_Q / \max(\|f_{\text{Burg}}\|_Q, 10^{-14}).$$

For all three Navier–Stokes formulations, e_u, e_{mom}, and $e_{\nabla p}$ are the restrictions of (6.3d) with the magnetic terms omitted. The primitive soft-divergence metric is

$$\|u_x + v_y\|_Q / \max\{(\|u_x\|_Q^2 + \|u_y\|_Q^2 + \|v_x\|_Q^2 + \|v_y\|_Q^2)^{1/2}, 10^{-14}\},$$

whereas the two hard-solenoidal formulations report $\max |u_x + v_y|$. The mixed-formulation relation metric is

$$\|\chi - \Delta\psi\|_Q / \max\{\|\chi\|_Q, \|\Delta\psi\|_Q, 10^{-14}\}.$$

The velocity boundary metric is maximum speed, and the magnetic boundary metric is maximum $|B \cdot n|$, on 64 deterministic points per spatial edge. The outer defect is

$$\delta_m = \frac{\left(\|u^{m+1} - u^m\|_Q^2 + \|B^{m+1} - B^m\|_Q^2\right)^{1/2}}{\max\{(\|u^{m+1}\|_Q^2 + \|B^{m+1}\|_Q^2)^{1/2}, 10^{-14}\}}. \tag{6.3e}$$

The reported observed outer contraction is $\rho_m = \delta_m / \delta_{m-1}$ for $m \geq 1$.

Table 4: Main physical evidence. Residuals use the learned current fields in both nonlinear factors. The first field is scalar u for the forced KdV-type case and velocity for NS/MHD. MHD induction is the scalar-potential residual, not a measured vector-induction residual. Dashes denote inapplicable quantities.

Test	Field error	Magnetic error	PDE / mom. residual	Scalar ind. residual	Pressure grad. error	Outer solves
Forced KdV-type case	3.1123×10^{-4}	—	2.4099×10^{-3}	—	—	7
Hard-solenoidal NS	3.7893×10^{-5}	—	8.0630×10^{-4}	—	1.9159×10^{-2}	2
Moderate-coupling MHD	1.6712×10^{-4}	7.2821×10^{-4}	8.3131×10^{-4}	1.9064×10^{-3}	5.0359×10^{-2}	4

The predeclared benchmark qualification tolerances for MHD were $\delta_m \leq 10^{-3}$; $e_u, e_B < 5 \times 10^{-2}$; $e_{\mathrm{mom}}, e_{\mathrm{ind}}, e_{\nabla p} < 10^{-1}$; and both maximum divergences and both maximum hard-boundary errors at most 10^{-12}. The pressure-gradient value 0.1 is therefore one benchmark tolerance, not a claim that larger values make a finite field solution scientifically meaningless.

Both methods used the same outer Picard map, zero initial frozen state, stopping criterion, physical formulation, collocation sets, $L = 3$, $M = 40$ architecture, paired network initialization, precision, and MATLAB environment; subsequent frozen fields were generated by the respective method's preceding iterate. LySep alone used the independent a, d, q, r states and Algorithm D.1. At each frozen problem the end-to-end PINN warm-started its network, reset Adam moments, and used 2000 full-batch Adam steps with learning rate 10^{-3}, $(\beta_1, \beta_2) = (0.9, 0.999)$, and $\epsilon = 10^{-8}$. It then called MATLAB R2024b's Deep Learning Toolbox `lbfgsState` and `lbfgsupdate` with their native weak-Wolfe line search and default state options; no L-BFGS name–value option was overridden. The second stage used at most 500 updates and stopped when `GradientsNorm` or `StepNorm` was at most 10^{-5} or when `LineSearchStatus` was `failed`. A loss-identity check had relative difference zero, and a directional-gradient check had relative error 8.14×10^{-10}; both checks were excluded from the runtimes. Because later frozen states are method dependent, cross-method conclusions below are led by the true nonlinear metrics in (6.3d), not by $J_{\text{last frozen}}$.

H Detailed accepted results and seed-level comparison

Detail 6.2 Burgers nonlinear qualification

The Burgers experiment isolates the nonlinear bridge. Three outer solves reduced the relative outer defect from its initialized value to 1.5284×10^{-3} and then 2.8177×10^{-6}; the final observed contraction ratio was 1.8436×10^{-3}. The test relative L^2 solution error was 6.8119×10^{-7}, and the true nonlinear residual relative L^2 was 4.0849×10^{-6}. Thus, in this smooth viscous case, the frozen inner problems and outer factor update recovered an accurate nonlinear solution. This experiment is evidence for the lagged bridge only, not a primary performance benchmark or a global convergence test.

Detail 6.3 Navier–Stokes formulation study

Table 5 compares three formulations under the same manufactured solution, network width, point counts, and seed. The primitive velocity-pressure formulation obtained good velocity, momentum, and pressure-gradient values, but its least-squares incompressibility equation left a divergence relative error above the predeclared 10^{-3} tolerance. The second-order streamfunction formulation made divergence arithmetic-precision small and removed the observed inner stagnation, but the independent χ relation and pressure gradient remained inaccurate. The chi-free third-order formulation retained exact divergence and reduced the velocity, momentum, and pressure-gradient errors substantially.

Table 5: Navier–Stokes formulation study. A dash indicates that the quantity is not present in that formulation; the solenoidality column states each reported metric explicitly.

(a) Numerical comparison

Formulation	Highest order	velocity rel. L^2	nonlinear momentum rel. L^2	pressure-gradient rel. L^2
Primitive u, v, p, soft incompressibility	2	1.3573×10^{-3}	2.4922×10^{-3}	7.1831×10^{-2}
Hard-solenoidal ψ, χ, p	2	1.2565×10^{-2}	4.8861×10^{-2}	5.2250×10^{-1}
Hard-solenoidal chi-free ψ, p	3	3.7893×10^{-5}	8.0630×10^{-4}	1.9159×10^{-2}

(b) Structural and formulation quantities

Formulation	Solenoidality statement	$\chi - \Delta\psi$ rel. L^2	Interpretation
Primitive u, v, p, soft incompressibility	2.1294×10^{-3} rel. L^2	–	Velocity and momentum meet their tolerances; soft divergence does not
Hard-solenoidal ψ, χ, p	Exact by construction; reported max abs. 1.73×10^{-18}	1.0810×10^{-1}	Exact divergence; chi-pressure defect remains
Hard-solenoidal chi-free ψ, p	Exact by construction; reported max abs. 0	–	Pressure-gradient defect removed while exact divergence is retained

The third-order run completed two outer solves with final defect 1.3280×10^{-4}, zero maximum boundary speed, and finite state. This formulation study supports the compensation mechanism in Lemma 5.2 and the benefit of removing the independent Laplacian field.

Detail 6.3a Matched five-seed NS optimization comparison

This comparison fixes the final chi-free (ψ, p) formulation selected above and varies the inner optimization mechanism. Both methods use the same width-40 shared trunk with two sine hidden layers, outputs (ϕ_ψ, ϕ_p), the analytic manufactured forcing, $T = 0.25$, $\nu = 0.1$, and the original fixed 2000/500 non-overlapping Halton split. For each prescribed seed 1–5, all initial network parameters match between methods. Seed 1 reuses the existing formal results; seeds 2–5 use the frozen protocol. The hard representation and zero pressure-output bias convention are identical; pressure accuracy is evaluated only through its gradient, with no added pressure supervision or gauge loss.

The common Oseen outer map starts from zero velocity, warm-starts the network between solves, uses the original velocity-relative defect, tolerance 10^{-3} and a cap of five solves. The existing first-frozen-loss stagnation rule is retained. LySep uses its frozen 1000-sweep separated block schedule with adaptive consistency weights. The E2E PINN minimizes only the mean sum of squared true frozen momentum residuals using 2000 full-batch Adam steps (learning rate 10^{-3}, moments $0.9, 0.999$, epsilon 10^{-8}), followed by at most 500 native MATLAB L-BFGS updates per solve, stopping on gradient/step norm 10^{-5} or failed line search. Optimizer state resets per solve; there are no extra physical penalties. Thus, the comparison holds formulation and outer protocol fixed while comparing the stated inner optimization procedures and budgets.

The final nonlinear momentum metric uses the learned current velocity in all nonlinear factors, rather than the last frozen Oseen coefficient. Table 6 reports the five paired results. LySep has lower velocity, nonlinear-momentum and pressure-gradient errors in every pair. Median paired E2E/LySep ratios are 100.901, 35.199 and 32.502, respectively. Each ratio is computed separately for each seed before taking its median; these are not ratios of medians. Both methods meet outer tolerance in 5/5 seeds. Hard solenoidality and boundary conditions are shared representation properties: both retain arithmetic-precision divergence and zero recorded maximum boundary speed. The evidence concerns these five paired initializations under the matched final formulation, separately from the formulation study.

Table 6: Five paired NS initializations under the final hard-solenoidal formulation. Ratios are E2E/LySep; bold marks the lower error. All errors use the final physical network fields.

(a) Velocity and nonlinear momentum

	e_u			e_{mom}		
Seed	LySep	E2E	Ratio	LySep	E2E	Ratio
1	$\mathbf{3.789 \times 10^{-5}}$	2.248×10^{-3}	59.317	$\mathbf{8.063 \times 10^{-4}}$	1.655×10^{-2}	20.521
2	$\mathbf{1.791 \times 10^{-5}}$	2.791×10^{-3}	155.865	$\mathbf{4.775 \times 10^{-4}}$	1.807×10^{-2}	37.855
3	$\mathbf{1.938 \times 10^{-5}}$	2.208×10^{-3}	113.928	$\mathbf{4.358 \times 10^{-4}}$	1.876×10^{-2}	43.036
4	$\mathbf{3.703 \times 10^{-5}}$	2.314×10^{-3}	62.481	$\mathbf{9.343 \times 10^{-4}}$	1.901×10^{-2}	20.344
5	$\mathbf{2.279 \times 10^{-5}}$	2.299×10^{-3}	100.901	$\mathbf{4.470 \times 10^{-4}}$	1.573×10^{-2}	35.199

(b) Pressure gradient and outer tolerance

	$e_{\nabla p}$			Outer tolerance met	
Seed	LySep	E2E	Ratio	LySep	E2E
1	$\mathbf{1.916 \times 10^{-2}}$	3.640×10^{-1}	18.998	Yes	Yes
2	$\mathbf{1.129 \times 10^{-2}}$	4.003×10^{-1}	35.448	Yes	Yes
3	$\mathbf{1.032 \times 10^{-2}}$	4.190×10^{-1}	40.584	Yes	Yes
4	$\mathbf{2.212 \times 10^{-2}}$	4.261×10^{-1}	19.265	Yes	Yes
5	$\mathbf{1.053 \times 10^{-2}}$	3.424×10^{-1}	32.502	Yes	Yes

Table 7: NS five-seed aggregate accuracy statistics. Std. is sample standard deviation (denominator $5 - 1$). The paired ratios are summarized separately in the text.

Method	Mean	Std.	Median	Min.	Max.
e_u					
Third-order LySep	2.700×10^{-5}	9.719×10^{-6}	2.279×10^{-5}	1.791×10^{-5}	3.789×10^{-5}
Matched E2E PINN	2.372×10^{-3}	2.380×10^{-4}	2.299×10^{-3}	2.208×10^{-3}	2.791×10^{-3}
e_{mom}					
Third-order LySep	6.202×10^{-4}	2.333×10^{-4}	4.775×10^{-4}	4.358×10^{-4}	9.343×10^{-4}
Matched E2E PINN	1.762×10^{-2}	1.426×10^{-3}	1.807×10^{-2}	1.573×10^{-2}	1.901×10^{-2}
$e_{\nabla p}$					
Third-order LySep	1.469×10^{-2}	5.546×10^{-3}	1.129×10^{-2}	1.032×10^{-2}	2.212×10^{-2}
Matched E2E PINN	3.904×10^{-1}	3.605×10^{-2}	4.003×10^{-1}	3.424×10^{-1}	4.261×10^{-1}

Detail 6.3b Larger-scale Navier–Stokes stress test

This single-seed larger-scale stress test uses the same manufactured NS problem and final chi-free hard-solenoidal (ψ, p) formulation as Detail 6.3a, with seed 1, width 80, 8000 training points and 2000 test points. It retains the same paired initialization principle, LySep and matched E2E definitions, and Oseen outer protocol, including zero initial frozen velocity, outer tolerance 10^{-3} and the five-solve cap.

Table 8: Single-seed larger-scale NS stress test at width 80 and 8000/2000 train/test points. Accuracy metrics use the final physical fields, with current-velocity nonlinear momentum residuals.

Metric	Third-order LySep	Matched E2E PINN
Outer solves	2	2
Final outer defect	7.0088×10^{-5}	6.8069×10^{-4}
Velocity relative L^2 error	1.6490×10^{-6}	2.5346×10^{-3}
Nonlinear momentum relative L^2 residual	7.3436×10^{-5}	1.9556×10^{-2}
Pressure-gradient relative L^2 error	1.7362×10^{-3}	4.3607×10^{-1}
Maximum absolute divergence	0	1.7347×10^{-18}
Maximum boundary speed	0	0

Both methods remain finite, meet the outer tolerance and preserve the hard structural conditions, with arithmetic-precision divergence and zero boundary speed. LySep retains lower errors in all three reported accuracy metrics for this paired stress-test run. This is a single-seed larger-scale stress test, not a multi-seed statistical robustness or asymptotic scaling claim.

Detail 6.4 Moderate-coupling MHD

The MHD benchmark was designed to avoid a nearly linear or geometrically aligned regime. With all norms evaluated on the fixed training geometry, define

$$R_{\mathrm{ind}} = \frac{\|u \cdot \nabla A\|_Q}{\|A_t - \eta \Delta A\|_Q}, \qquad R_{\mathrm{adv}} = \frac{\|(u \cdot \nabla)u\|_Q}{\|u_t + \nabla p - \nu \Delta u\|_Q}, \qquad R_L = \frac{\|(B \cdot \nabla)B\|_Q}{\|u_t + \nabla p - \nu \Delta u\|_Q}.$$

For the exact manufactured fields, these nonlinear-to-linear scale ratios were

$$\begin{aligned} R_{\mathrm{ind}} &= 0.051482987707379108, \\ R_{\mathrm{adv}} &= 0.21528963507791876, \\ R_L &= 0.068776811095094537. \end{aligned} \tag{6.4}$$

Here R_{ind} compares nonlinear magnetic advection with the linear time-diffusion induction scale, R_{adv} compares inertial advection with the linear momentum scale, and R_L measures magnetic tension relative to the same momentum scale. The corresponding learned-state ratios were

$$0.051772776340638572, \qquad 0.21619830653557304, \qquad 0.068504742731582358, \tag{6.5}$$

respectively. Their proximity to (6.4), together with the non-negligible magnitudes of all three exact ratios, confirms that the reported solution resolves the intended moderate dynamic coupling rather than collapsing to a decoupled or near-linear state.

Four outer solves reduced the relative defect to 9.0640×10^{-5}; the last observed contraction ratio was 7.2875×10^{-2}. Table 9 gives the final physical quantities. The velocity and magnetic errors are in the 10^{-4} range, the nonlinear momentum and scalar induction residuals are in the 10^{-3} range, and both divergence identities and both hard boundary conditions hold exactly or to arithmetic precision.

The scalar induction residual in Table 9 is the true nonlinear residual of (5.16). Proposition 5.3 links an exact smooth scalar fixed point to vector induction. The finite scalar residual reported here is not a direct vector-residual measurement or a bound on its curl. It should not be interpreted as evidence that the scalar and primitive-vector linearizations have identical intermediate iterates.

Table 9: Moderate-coupling MHD result.

Quantity	Value
Outer iterations	4
Final outer defect	$9.0639946035263484 \times 10^{-5}$
Last frozen inner physical loss at the final outer state	$2.8712514474747504 \times 10^{-7}$
Velocity relative L^2 error	$1.6711862936803873 \times 10^{-4}$
Magnetic relative L^2 error	$7.2821280877405738 \times 10^{-4}$
Nonlinear momentum relative L^2 residual	$8.3131084500949069 \times 10^{-4}$
Nonlinear scalar-induction relative L^2 residual	$1.9063856563993299 \times 10^{-3}$
Pressure-gradient relative L^2 error	$5.0359057776301057 \times 10^{-2}$
$\max\lvert\nabla \cdot u\rvert$	0
$\max\lvert\nabla \cdot B\rvert$	$5.5511151231257827 \times 10^{-17}$
Maximum velocity boundary error	0
Maximum magnetic normal-boundary error	0
Wall-clock runtime	5643.349100 s

DETAIL 6.5 MATCHED END-TO-END PINN COMPARISON

The primary continuous comparison is the pressure-gradient relative error: its five-seed median was 3.9408×10^{-2} for LySep and 1.3760×10^{-1} for the matched end-to-end PINN. Table 10 retains all ten method runs across five paired seeds. All five LySep runs, and none of the five end-to-end runs, fell below every predeclared benchmark qualification tolerance. E2E seeds 1–3 and 5 exceed $e_{\nabla p} < 0.1$. Seed 4 satisfies the continuous-error thresholds but reaches the five-solve cap with outer defect $0.001332115341958439 > 10^{-3}$. All baseline solutions remain finite and preserve the hard divergence and boundary conditions.

The five-seed LySep medians were 1.6734×10^{-4} for velocity error, 6.9830×10^{-4} for magnetic error, 7.8195×10^{-4} for momentum residual, 1.8253×10^{-3} for induction residual, and 3.9408×10^{-2} for pressure-gradient error. The corresponding end-to-end medians were 4.1681×10^{-4}, 4.0850×10^{-3}, 2.1301×10^{-3}, 7.8659×10^{-3}, and 1.3760×10^{-1}. The method-specific last frozen inner physical losses are not evaluations at common final frozen coefficient fields. The $5/5$ versus $0/5$ tolerance count is therefore a secondary benchmark qualification summary rather than the headline statistic.

Table 3 displays the paired continuous metrics and their medians. Section 5.4 gives the matched physical-accuracy interpretation and the five-pair sampling qualification.

I INTRINSIC THIRD-ORDER NONLINEAR SCALAR EXPERIMENT AND CONTINUATION DISCLOSURE

I.1. COMPLETE MANUFACTURED EQUATION, HARD ANSATZ AND FROZEN COEFFICIENTS

On $Q = (0, 0.5] \times [-1, 1]$, take $\mu = 0.05$, $\epsilon = 1$, $\alpha = 0.2$, and

$$H_K = (1-x)^2(1+x) = 1 - x - x^2 + x^3, \quad S = 1 + \alpha\cos(\pi x), \quad F = H_K S, \quad u^* = \epsilon t e^{-t} F.$$

The boundary/initial conditions are $u(0, x) = 0$, $u(t, -1) = u(t, 1) = 0$, and $u_x(t, 1) = 0$. These follow identically from $u = tH_K\phi_\theta$. They do not prescribe an extra derivative trace at $x = -1$. All manufactured derivatives and forcing are analytic:

Table 10: Paired LySep and matched lagged end-to-end PINN results.

(a) Seed 1

Method	LySep	End-to-end PINN
Below all predeclared tolerances	YES	NO: pressure gradient only
Outer steps	4	4
Last frozen inner physical loss at method-specific final outer state	$2.8712514474747504 \times 10^{-7}$	$1.9567551613648221 \times 10^{-6}$
Velocity error	$1.6711862936803873 \times 10^{-4}$	$4.1681063435869338 \times 10^{-4}$
Magnetic error	$7.2821280877405738 \times 10^{-4}$	$2.0096428828289461 \times 10^{-3}$
Momentum residual	$8.3131084500949069 \times 10^{-4}$	$2.1300755460340521 \times 10^{-3}$
Induction residual	$1.9063856563993299 \times 10^{-3}$	$5.0820689219923908 \times 10^{-3}$
Pressure-gradient error	$5.0359057776301057 \times 10^{-2}$	$1.4138072562168136 \times 10^{-1}$
Runtime (s)	5643.349100	528.672700

(b) Seed 2

Method	LySep	End-to-end PINN
Below all predeclared tolerances	YES	NO: pressure gradient only
Outer steps	4	5
Last frozen inner physical loss at method-specific final outer state	$2.1407154158015347 \times 10^{-7}$	$2.1635498161883479 \times 10^{-6}$
Velocity error	$1.5776736050220644 \times 10^{-4}$	$3.4089577261936749 \times 10^{-4}$
Magnetic error	$5.3315675406263982 \times 10^{-4}$	$5.2780299651903236 \times 10^{-3}$
Momentum residual	$7.3705162641295868 \times 10^{-4}$	$1.6332092897578532 \times 10^{-3}$
Induction residual	$1.4304414674486926 \times 10^{-3}$	$9.2658565318975775 \times 10^{-3}$
Pressure-gradient error	$3.1913884193479077 \times 10^{-2}$	$1.3759529389332525 \times 10^{-1}$
Runtime (s)	5579.552616	568.204507

(c) Seed 3

Method	LySep	End-to-end PINN
Below all predeclared tolerances	YES	NO: pressure gradient only
Outer steps	4	4
Last frozen inner physical loss at method-specific final outer state	$2.4701827651110523 \times 10^{-7}$	$1.6071430243299655 \times 10^{-6}$
Velocity error	$1.6734218614122595 \times 10^{-4}$	$3.3969011526145372 \times 10^{-4}$
Magnetic error	$6.0162181001025450 \times 10^{-4}$	$3.9230744691930910 \times 10^{-3}$
Momentum residual	$7.6662835402299472 \times 10^{-4}$	$1.6815669159182234 \times 10^{-3}$
Induction residual	$1.8253453653016948 \times 10^{-3}$	$6.7737667442180488 \times 10^{-3}$
Pressure-gradient error	$2.7886317095959345 \times 10^{-2}$	$1.1607282397326739 \times 10^{-1}$
Runtime (s)	5389.259546	494.831132

Table 10: Paired LySep and matched lagged end-to-end PINN results (continued).

(d) Seed 4 Method	LySep	End-to-end PINN
Below all predeclared tolerances	YES	NO: outer tolerance only
Outer steps	4	5
Last frozen inner physical loss at method-specific final outer state	$3.2424520406735821 \times 10^{-7}$	$2.5848511282050338 \times 10^{-6}$
Velocity error	$1.7680225193693656 \times 10^{-4}$	$4.7245770742902966 \times 10^{-4}$
Magnetic error	$6.9830424768389126 \times 10^{-4}$	$4.4306554618415931 \times 10^{-3}$
Momentum residual	$9.0891206121482791 \times 10^{-4}$	$2.2456164065601412 \times 10^{-3}$
Induction residual	$1.6613794990195385 \times 10^{-3}$	$7.8658821414187147 \times 10^{-3}$
Pressure-gradient error	$6.4323550093554879 \times 10^{-2}$	$9.5578611679157155 \times 10^{-2}$
Runtime (s)	5526.507563	557.247970

(e) Seed 5 Method	LySep	End-to-end PINN
Below all predeclared tolerances	YES	NO: pressure gradient only
Outer steps	4	5
Last frozen inner physical loss at method-specific final outer state	$2.5747036967847514 \times 10^{-7}$	$3.4962185135243043 \times 10^{-6}$
Velocity error	$1.7883431376156190 \times 10^{-4}$	$5.7584482811253899 \times 10^{-4}$
Magnetic error	$7.6819594658419053 \times 10^{-4}$	$4.0849628737138089 \times 10^{-3}$
Momentum residual	$7.8194725729829363 \times 10^{-4}$	$2.4924865021124856 \times 10^{-3}$
Induction residual	$1.9019327616405719 \times 10^{-3}$	$9.8948494861560172 \times 10^{-3}$
Pressure-gradient error	$3.9408433000827012 \times 10^{-2}$	$1.3824921436805579 \times 10^{-1}$
Runtime (s)	5278.091699	540.361212

$$
\begin{gathered}
H_K' = -1 - 2x + 3x^2, \quad H_K'' = -2 + 6x, \quad H_K''' = 6, \\
S' = -\alpha\pi \sin \pi x, \quad S'' = -\alpha\pi^2 \cos \pi x, \quad S''' = \alpha\pi^3 \sin \pi x, \\
F' = H_K' S + H_K S', \quad F'' = H_K'' S + 2H_K' S' + H_K S'', \\
F''' = H_K''' S + 3H_K'' S' + 3H_K' S'' + H_K S''', \\
u_t^* = \epsilon e^{-t}(1-t)F, \quad u_x^* = \epsilon t e^{-t} F', \quad u_{xxx}^* = \epsilon t e^{-t} F''', \quad f = u_t^* + u^* u_x^* + \mu u_{xxx}^*.
\end{gathered}
\tag{I.1}
$$

For arbitrary network output ϕ,

$$
\begin{aligned}
u_t &= H_K \phi + t H_K \phi_t, \\
u_x &= t(H_K' \phi + H_K \phi_x), \\
u_{xx} &= t(H_K'' \phi + 2H_K' \phi_x + H_K \phi_{xx}), \\
u_{xxx} &= t(H_K''' \phi + 3H_K'' \phi_x + 3H_K' \phi_{xx} + H_K \phi_{xxx}).
\end{aligned}
\tag{I.2}
$$

Freezing $b = u^m$ gives $(u^{m+1})_t + b(u^{m+1})_x + \mu(u^{m+1})_{xxx} = f$. The coefficients acting on ϕ are

$$
\begin{aligned}
A_0 &= H_K + tbH_K' + \mu t H_K''', & A_t &= tH_K, \\
A_x &= tbH_K + 3\mu t H_K'', & A_{xx} &= 3\mu t H_K', \quad A_{xxx} = \mu t H_K.
\end{aligned}
\tag{I.3}
$$

The active state is exactly $a_1, a_2, d_{1,t}, d_{1,x}, d_{2,t}, d_{2,x}, q_{xx}, r_{xxx}$. There are no stored inactive derivative auxiliaries. The adaptive-weight formulas of Appendix B are restricted to these active coefficient families. The coefficient 468 used for the recorded consistency check is a conservative inherited specialization, not an optimized universal bound for arbitrary states and objectives.

I.2. Sampling and solver settings

The network has input dimension 2, two hidden layers of width 30, scalar output, sine activation, and seed 1. Parameters are initialized independently as $2(U - 1/2)/\sqrt{M}$ with uniform U in the prescribed seeded sequence. The first 1400 unscrambled Halton points in two coordinates are mapped by $(s_1, s_2) \mapsto (0.5s_1, 2s_2 - 1)$; the first 1000 are training points and the remaining 400 test points. This is a physics-residual fit, not supervised fitting to sampled solution values.

All computations use double precision. The scalar solver follows the active-index restriction of the block schedule in Appendix D: first-layer coordinate columns, first bias, first activation and derivatives; second-layer weights, second bias, second activation and derivatives; q_{xx}; quadratic r_{xxx} update; output LS and output bias. Each frozen solve uses 1000 sweeps. Current adaptive weights are differentiated in nonlinear blocks, and the same roundoff-scaled nonincrease backtracking is used. Forced KdV-type case stopping is based on finite states, the stated fixed inner budget and $\delta_m \le 10^{-3}$; the MHD-specific first-frozen-loss qualification is not an added scalar stopping condition.

The outer defect is $\|u^{m+1} - u^m\|_{\text{train}} / \max(\|u^{m+1}\|_{\text{train}}, 10^{-14})$. Test solution and true nonlinear residual thresholds are 0.05 and 0.1, respectively, in addition to the recorded consistency and finite-state checks. No parameter or tolerance was adjusted to turn a failed numerical criterion into a pass.

I.3. Training-free checks

Gate0 additionally recorded $\|u^* u_x^*\| = 1.697437256865$ and $\|u_{xxx}^*\| = 66.79347962983$, both nonzero. Ratios use the denominators in main-text equation (11). Gate0 uses training points;

Check	Scientific purpose	Recorded value / status
Gate0	Both nonlinear transport and third-order dispersion are active	Exact training $R_{nl} = 0.0938401155585$, $R_3 = 0.195276218896$; each exceeds 0.05; PASS
A	Analytic manufactured substitution	Maximum residual 0; PASS
B	Independent network derivative chain	Relative errors for first/second/third spatial derivatives: 0, 1.5870×10^{-16}, 1.5091×10^{-16}; PASS
C	Hard-factor reconstruction	Relative errors for u_t, u_x, u_{xxx}: 0, 1.2691×10^{-16}, 7.8711×10^{-17}; PASS
D	Frozen operator versus direct physical expression	2.0817×10^{-16}; PASS
E	Exact auxiliary recovery	Both losses 0.5462425526934; auxiliary residual and penalty 0; relative difference 0; PASS
F	Conditional output-block stationarity	Relative gradient 9.9820×10^{-18}; PASS

Outer index (0-based)	Sweeps	Separated loss	Frozen physical loss	δ_m	ρ_m
0	1000	0.002383191010	0.007205706128	1	—
1	1000	0.001255795074	0.003516381399	0.07603843728	0.07603843728
2	1000	0.001008003025	0.002671861514	0.008443885276	0.1110475909
3	1000	$6.201955560 \times 10^{-5}$	$1.740473198 \times 10^{-4}$	0.04209469697	4.985228434
4	1000	$5.869212660 \times 10^{-5}$	$1.606280035 \times 10^{-4}$	0.002379841998	0.05653543486
5	1000	$1.805231559 \times 10^{-6}$	$3.718876473 \times 10^{-6}$	0.005423153391	2.278787161
6	1000	$6.776786167 \times 10^{-7}$	$1.905548120 \times 10^{-6}$	0.0004763563102	0.08783751368

final learned ratios use test points. The exact test values are $R_{nl} = 0.0947689783239$ and $R_3 = 0.192193641970$.

I.4. Outer-budget disclosure and complete history

Post-hoc outer-budget extension. The original five-outer-solve budget did not reach the prescribed outer tolerance: its final δ was 0.00237984199811. All other recorded acceptance checks passed, but the original five-solve result was not accepted. A continuation resumed from the saved final state for at most two additional frozen solves. Scientific parameters, sample sets, network, initialization history, objective, inner solver, 1000-sweep budget, and tolerances were unchanged. The total cap became 7. The previously recorded training-free checks were retained and not rerun. The seed-1 result in the five-seed study is this continued state, not a reclassification of the original five-solve outcome. Seeds 2–5 were run under the seven-solve protocol fixed before execution.

The sequence is not monotonically contractive: indices 3 and 5 have $\rho_m > 1$. No estimate of a global Picard contraction constant is inferred.

Original solve time was 566.6815598 s; continuation time was 227.2535704 s. These are separately recorded optimization intervals, not an independently rerun seven-solve timing or a speed benchmark. The original five-solve solution and nonlinear residual errors were 0.005772240936 and 0.022270060266. The reported final state follows the continuation protocol described above.

I.5. Matched order-reduced mixed PINN

This matched mixed PINN is an order-reduction alternative, not a faithful MIM reproduction.

The shared-trunk network outputs (ϕ_u, ϕ_v, ϕ_w) with hard factors $u = tH_K\phi_u$, $v = t(1-x)\phi_v$, and $w = t\phi_w$; every output bias is trainable. The three frozen rows in Section 5.1 are scaled by fixed training-exact RMS values of f, u_x^*, u_{xx}^* (floored at 10^{-14}). The objective sums their mean squared

Final continued quantity	Value
Solution relative L^2 error	0.00031122819685447547
True nonlinear residual relative L^2	0.0024099232555148749
Learned test R_{nl}	0.0947259432763574
Learned test R_3	0.192186009225745
Separated loss	$6.7767861671593154 \times 10^{-7}$
Last frozen physical loss	$1.9055481197800972 \times 10^{-6}$
Recorded $468J_S$	0.00031715359262305597
Finite state / consistency / final acceptance	PASS / PASS / PASS

scaled residuals. Each frozen solve uses 2000 Adam steps (learning rate 10^{-3}, moments $0.9, 0.999$, numerical epsilon 10^{-8}), followed by up to 500 native L-BFGS updates, stopping on gradient/step norm 10^{-5} or failed line search. Optimizer state resets per solve; the network warm-starts from the previous solve. The 1000/400 points, sine activation, width 30, and seed 1 match the direct case; three output branches change the output-parameter count. Starting at $u^0 = 0$, the predeclared cap is seven solves with outer tolerance 10^{-3}.

The saved run converged in four solves, with final defect $2.2924478934 \times 10^{-4}$ and optimization time 77.4618221 s. Its solution relative L^2 error is $1.9017098657228679 \times 10^{-3}$, and its original-PDE residual is $1.2562982780521401 \times 10^{-2}$. Its mixed-PDE residual is $5.0236301012 \times 10^{-3}$; $\|v - u_x\|/\|u_x\| = 3.6232902970 \times 10^{-3}$ and $\|w - v_x\|/\|u_{xx}\| = 2.7052976179 \times 10^{-3}$. The headline residual uses $u_t + uu_x + \mu u_{xxx} - f$ evaluated from the final u branch, independently of v, w. Learned test ratios are $R_{\mathrm{nl}} = 0.09474652617$ and $R_3 = 0.19748394113$. Both methods meet solution-error 0.05 and original-residual 0.1 thresholds; The mixed-PINN result provides additional optimization-cost context. These are the existing matched mixed-PINN results, with end-to-end training rather than LySep auxiliary updates.

I.6. Five-seed accuracy and mixed-residual baseline evidence

Tables 11–13 report the paired prescribed seeds 1–5. All routes use the same manufactured equation, deterministic 1000/400 train/test split, width 30, two sine hidden layers and double precision. Errors and ratios are rounded only for display; statistics use the recorded values. Every headline original-PDE residual differentiates the final u branch directly through u_{xxx}, independently of learned v, w.

The two LySep routes use the common seven-solve cap, 1000 sweeps per frozen problem and outer tolerance 10^{-3}. The direct route uses a, d, q_{xx}, r_{xxx}; the reduced route uses only a, d for the first-order mixed system. Seed 1 reuses the formal direct continuation in I.4 and the formal reduced result; seeds 2–5 use the protocol fixed before execution. Direct meets outer tolerance in 4/5 seeds and reduced in 0/5.

The MIM-style baseline follows the first-order mixed principle with a shared trunk and outputs (ϕ_u, ϕ_v, ϕ_w), using the same hard representations as I.5: $u = tH_K\phi_u$, $v = t(1-x)\phi_v$, $w = t\phi_w$. It trains the full nonlinear mixed residual with no outer lagging or LySep auxiliaries, unlike the lagged E2E comparator retained in I.5. The fixed scales are the training-exact RMS values of f, u_x^*, u_{xx}^*, floored at 10^{-14}, and the objective sums the three mean squared scaled residuals. Each prescribed seed uses one continuous run of 8000 full-batch Adam steps (learning rate 10^{-3}, moments $0.9, 0.999$, epsilon 10^{-8}), followed by up to 2000 native MATLAB L-BFGS updates, stopping on gradient/step norm 10^{-5} or failed line search, with no restart or additional budget. Seed 1 reuses the formal pilot. The benchmark-specific adaptation is described in Related Work.

Direct has lower solution and original-PDE residual errors in all five pairs against each comparator. Median paired ratios (comparator/direct) are 1.95 and 3.61 for reduced LySep, and 2.97 and 2.47 for the MIM-style PINN. The outer-tolerance metric applies only to the two lagged LySep routes. These results establish the reported accuracy ordering under this matched benchmark and these five paired initializations.

Table 11: Direct versus reduced LySep across five paired KdV seeds. Ratios are reduced/direct; bold marks lower errors. Panel (b) reports the final outer defect and whether $\delta \le 10^{-3}$, under the common seven-solve cap.

(a) Solution and original-PDE residual

	e_u			e_{PDE}		
Seed	Direct	Reduced	Ratio	Direct	Reduced	Ratio
1	$\mathbf{3.112 \times 10^{-4}}$	5.580×10^{-4}	1.79	$\mathbf{2.410 \times 10^{-3}}$	8.703×10^{-3}	3.61
2	$\mathbf{2.434 \times 10^{-4}}$	7.850×10^{-3}	32.25	$\mathbf{2.213 \times 10^{-3}}$	4.872×10^{-2}	22.01
3	$\mathbf{2.864 \times 10^{-4}}$	5.597×10^{-4}	1.95	$\mathbf{2.517 \times 10^{-3}}$	7.518×10^{-3}	2.99
4	$\mathbf{5.100 \times 10^{-4}}$	3.355×10^{-3}	6.58	$\mathbf{3.012 \times 10^{-3}}$	1.566×10^{-2}	5.20
5	$\mathbf{2.846 \times 10^{-4}}$	4.884×10^{-4}	1.72	$\mathbf{2.541 \times 10^{-3}}$	7.343×10^{-3}	2.89

(b) Outer convergence

Seed	Direct δ	Reduced δ	Direct tol.	Reduced tol.
1	4.764×10^{-4}	7.724×10^{-3}	Yes	No
2	4.761×10^{-4}	2.415×10^{-3}	Yes	No
3	9.496×10^{-4}	7.484×10^{-3}	Yes	No
4	4.828×10^{-3}	9.845×10^{-3}	No	No
5	1.547×10^{-4}	5.085×10^{-3}	Yes	No

Table 12: Direct LySep versus the MIM-style nonlinear mixed-residual PINN across five paired KdV seeds. Ratios are MIM-style/direct; bold marks lower errors. Outer tolerance is inapplicable to the MIM-style route.

	e_u			e_{PDE}		
Seed	Direct	MIM-style	Ratio	Direct	MIM-style	Ratio
1	$\mathbf{3.112 \times 10^{-4}}$	6.809×10^{-4}	2.19	$\mathbf{2.410 \times 10^{-3}}$	4.510×10^{-3}	1.87
2	$\mathbf{2.434 \times 10^{-4}}$	3.783×10^{-4}	1.55	$\mathbf{2.213 \times 10^{-3}}$	5.460×10^{-3}	2.47
3	$\mathbf{2.864 \times 10^{-4}}$	8.503×10^{-4}	2.97	$\mathbf{2.517 \times 10^{-3}}$	6.024×10^{-3}	2.39
4	$\mathbf{5.100 \times 10^{-4}}$	2.780×10^{-3}	5.45	$\mathbf{3.012 \times 10^{-3}}$	3.065×10^{-2}	10.17
5	$\mathbf{2.846 \times 10^{-4}}$	9.432×10^{-4}	3.31	$\mathbf{2.541 \times 10^{-3}}$	8.737×10^{-3}	3.44

Table 13: Five-seed aggregate accuracy statistics. Std. is the sample standard deviation (denominator $5 - 1$). Paired ratios in Table 1 are medians of per-seed ratios, not ratios of these medians.

Method	Mean	Std.	Median	Min.	Max.
e_u					
Direct LySep	3.271×10^{-4}	1.051×10^{-4}	2.864×10^{-4}	2.434×10^{-4}	5.100×10^{-4}
Reduced LySep	2.562×10^{-3}	3.198×10^{-3}	5.597×10^{-4}	4.884×10^{-4}	7.850×10^{-3}
MIM-style PINN	1.127×10^{-3}	9.489×10^{-4}	8.503×10^{-4}	3.783×10^{-4}	2.780×10^{-3}
e_{PDE}					
Direct LySep	2.539×10^{-3}	2.946×10^{-4}	2.517×10^{-3}	2.213×10^{-3}	3.012×10^{-3}
Reduced LySep	1.759×10^{-2}	1.773×10^{-2}	8.703×10^{-3}	7.343×10^{-3}	4.872×10^{-2}
MIM-style PINN	1.108×10^{-2}	1.105×10^{-2}	6.024×10^{-3}	4.510×10^{-3}	3.065×10^{-2}

J DEVELOPMENT-STAGE QUALIFICATIONS AND SCOPE OF INTERPRETATION

J.1. WHY THE FINAL NS FORMULATION WAS SELECTED

The development sequence isolates distinct obstacles; it is not presented as a controlled ablation in which only derivative order changes.

Stage / formulation	Main observation	Consequence
N1, lagged Burgers	Solution error 6.8119×10^{-7}, nonlinear residual 4.0849×10^{-6}	Supports the nonlinear lagging bridge
N2, primitive soft-divergence NS	Velocity 0.0013573, momentum 0.0024922, pressure gradient 0.0718314; relative divergence 0.0021294 exceeds 0.001	Seek an exact solenoidal representation
N3, mixed vorticity/streamfunction	Hard divergence 1.73×10^{-18}, but velocity 0.364425, vorticity 0.683374, nonlinear vorticity residual 0.700538, kinematic residual 0.676062	Observed inner stagnation motivates a different strong formulation
N3R1, ψ, χ, p	Inner stagnation removed and outer defect 0.000130284; relation error 0.108103 and pressure-gradient error 0.522502 remain	Diagnose the harmonic compensation mechanism of Appendix E
N3R2, third-order contract	Eliminate independent χ, retaining primitive pressure and direct third derivatives	Mathematical design stage; no numerical run
N3R3, final ψ, p	Velocity 3.7893×10^{-5}, momentum 8.0630×10^{-4}, pressure gradient 0.019159, hard divergence 0	Final NS evidence

The soft divergence column above is a relative discrete L^2 quantity; hard divergence is maximum absolute divergence. Their numerical values are not the same metric. No weight sweep was performed after the soft-divergence failure. For the independent-χ formulation, increasing a finite L^2 relation weight cannot create a uniform H^1-by-L^2 estimate over the admissible smooth class. The harmonic example $g_k = \cosh(kx)\cos(ky)$ has derivative-to-value norm ratio of order k; after normalization and scaling by $1/k$, its value norm tends to zero while its gradient need not. A pressure correction compensates that gradient in momentum. This is an analytical limitation of the adopted formulation, not a theorem that reweighting cannot improve any particular finite run.

J.2. GEOMETRIC NONDEGENERACY VERSUS DYNAMIC COUPLING IN MHD

Write $Q(s) = 1 - s^2$, $H = Q(x)^2 Q(y)^2$, $K = Q(x)Q(y)$ and $a(t) = te^{-t}$. H1 used $\psi^* = 0.05aH$, $A^* = 0.04aK$, and $T = 0.25$. Because $H = K^2$, their gradients align and $u^* \cdot \nabla A^* = 0$. H1 therefore remains a structural MHD qualification, not evidence of dynamically active induction transport. It passed with velocity error 3.79795×10^{-5}, magnetic error 1.65474×10^{-5}, and scalar-induction residual 1.70688×10^{-5}.

H1R1 changed the magnetic shape to $G = K(1 + \gamma x)$ with $\gamma = 0.5$, retaining the small amplitudes and short time. Then

$$H_y G_x - H_x G_y = -4\gamma y Q(x)^3 Q(y)^2,$$

which is not identically zero. Its reported geometric alignment ratio 0.233467 measures nondegeneracy, not transport relative to the linear induction dynamics. The exact maximum transport was only $3.97984508 \times 10^{-5}$. H1R1 passed with velocity 3.79566×10^{-5}, magnetic 1.47198×10^{-5}, scalar-induction residual 3.76036×10^{-5}, and pressure gradient 0.0191605. It supplies a geometric qualification; it is not substituted for the final moderate-dynamic-coupling result.

The final moderate-coupling MHD case retained $\gamma = 0.5$, set $T = 1$, velocity amplitude 0.75 and magnetic amplitude 0.40, with $\nu = \eta = 0.1$. Define $L_u = u_t - \nu\Delta u + \nabla p$ and

$$R_{ind} = \frac{\|u \cdot \nabla A\|}{\max(\|A_t - \eta \Delta A\|, 10^{-14})},$$
$$R_{adv} = \frac{\|(u \cdot \nabla)u\|}{\max(\|L_u\|, 10^{-14})},$$
$$R_L = \frac{\|(B \cdot \nabla)B\|}{\max(\|L_u\|, 10^{-14})}.$$

Exact training ratios were 0.05148298771, 0.21528963508, 0.06877681110, each above the prescribed 0.03 dynamic threshold. Learned ratios were 0.05177277634, 0.21619830654, 0.06850474273. These are term-size diagnostics, not relative prediction errors. This case is the final MHD evidence because it combines the hard representation, nonaligned geometry and explicitly qualified moderate dynamic coupling. It is not a turbulent or strong-coupling result.

J.3. Reproducibility and interpretation of the matched comparator

The five-pair comparison is a descriptive assessment under the specified optimizer budgets. Conditional quadraticity concerns the output block; full-objective convexity and global Picard convergence are not established by the present results.

The LySep seed 1 state and runtime in the paired comparison are the existing seed-1 final MHD result, reused rather than rerun. Seeds 2–5 are separate prescribed paired initializations. Training-free identity and gradient checks are excluded from optimization runtimes. Later frozen coefficients are method specific, and the end-to-end method uses no auxiliary loss. Consequently, comparing the numerical values of each method's separated loss is invalid, and even the last frozen physical loss is not evaluated at a common frozen state. The main comparison therefore uses current-field nonlinear residuals and solution/pressure-gradient errors.

Hard divergence is a shared representation property of both methods. The scalar-to-vector identity differentiates the scalar induction residual, so controlling its scalar L^2 norm alone supplies no vector-residual bound without derivative control. The comparison uses the field errors and current-field nonlinear residuals specified above; its sampling and computational interpretation is given in Section 5.4.

J.4. Numerical theorem checks already performed

Exact auxiliary recovery and conditional output stationarity were checked on the original qualification states. Burgers gave relative loss difference 5.55×10^{-17} and output stationarity 1.62×10^{-18}; primitive NS gave zero loss difference and 9.29×10^{-18} stationarity. Final third-order NS reported zero relative errors for the four spatial third derivatives and mixed-permutation checks. MHD passed analytic-source, hard-representation, frozen-operator, exact-recovery and output-stationarity checks. The forced KdV-type case's distinct checks are in I.3. These are implementation qualifications supporting the mathematical formulas, not experimental proofs of the general theorems.