PIVOT: PHYSICS-INFORMED VIRTUAL PROBING FOR OUTPUT-SPACE TUNING OF NEURAL OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Partial differential equations (PDEs) are fundamental to modeling complex physical systems, and neural operators provide efficient surrogates for traditional numerical solvers. However, when test conditions depart from the training distribution, pretrained neural operators can suffer substantial accuracy degradation and accumulate physical errors over long rollouts. Existing adaptation methods update model parameters or introduce auxiliary modules, but these adjustments often lack direct constraints from the target physics. To bridge this gap, we introduce PIVOT, a general framework for output-space adaptation through the local responses of frozen neural operators. PIVOT exploits the observation that frozen operators can retain reliable local responses even when their absolute predictions are inaccurate. It learns physically admissible perturbation directions and probes inputs with small, opposite perturbations. Differencing the paired predictions attenuates shared prediction bias, extracting responses that reflect the true system's local behavior. PIVOT then assesses response reliability through physical consistency checks and learns spatially varying combinations of reliable responses to correct the current prediction. Finally, it bounds correction magnitudes and progressively scales down updates that increase physical residuals, keeping the pretrained neural operator fully frozen throughout adaptation. Experiments on APEBench and RealPDEBench across diverse physical systems and distribution shifts demonstrate that PIVOT jointly improves predictive accuracy, physical consistency, and long-horizon rollout stability. The complete model implementation is provided in the supplementary material.

1 INTRODUCTION

Partial differential equations (PDEs) underpin the modeling of complex physical systems, but the cost of repeatedly solving them with conventional numerical methods limits their use in applications requiring real-time computation. Neural operators offer efficient surrogates by learning mappings between function spaces (Lu et al., 2021; Li et al., 2021; Kovachki et al., 2023), with recent advances strengthening their ability to predict complex physical processes (Tran et al., 2023; Hao et al., 2024; Herde et al., 2024; Buitrago Ruiz et al., 2025; Zhou et al., 2026). However, when test conditions depart from the training distribution, predictive accuracy can deteriorate and predictions may become inconsistent with the governing physics. These errors can accumulate over successive predictions, driving trajectories away from the true dynamics and destabilizing long rollouts (Takamoto et al., 2022; Lippe et al., 2023). Adapting neural operators to new conditions therefore requires improving predictive accuracy while maintaining consistency with physical laws.

Existing adaptation methods typically update model parameters, introduce lightweight trainable modules, or learn output corrections (Goswami et al., 2022; Hu et al., 2021; Liu et al., 2024; Zhao et al., 2024; Albert et al., 2025). These mechanisms provide flexibility to fit target data, but prediction error alone does not determine whether a correction respects the target dynamics. Physically consistent predictions must also satisfy boundary conditions and applicable conservation or balance laws. Corrections that overlook these requirements may improve short-term accuracy while introducing inconsistencies that compromise subsequent evolution. Adaptation therefore requires physical information to guide the construction and assessment of corrections alongside the reduction of prediction error.

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

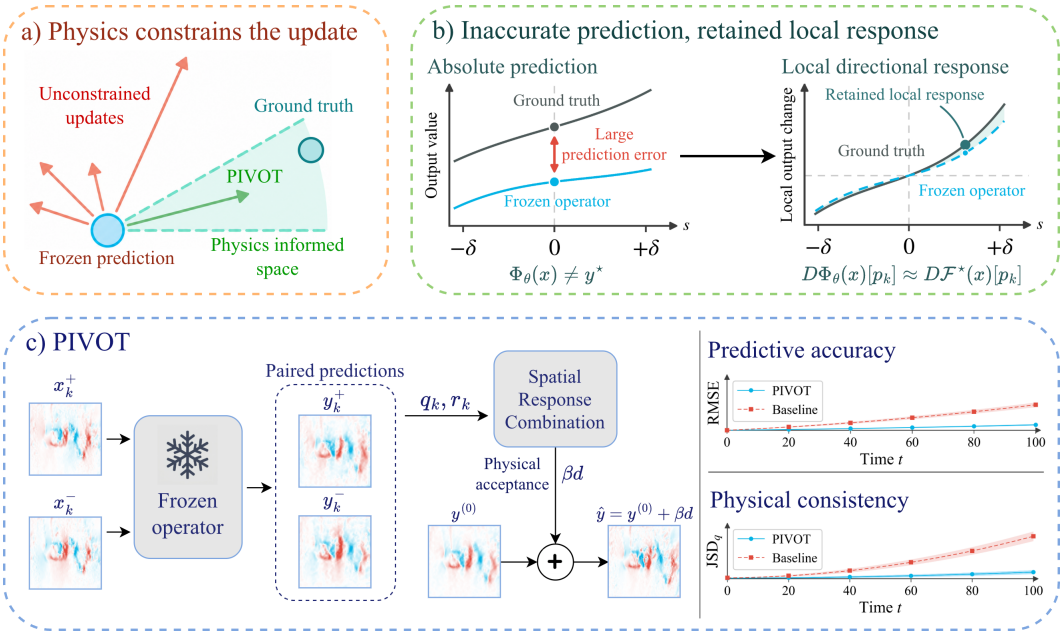


Figure 1: Frozen neural operators can retain reliable local responses under distribution shift. PIVOT extracts these responses through physically admissible input perturbations and uses them to construct prediction corrections guided by physical consistency checks.

We seek this physical information in the predictive behavior of the frozen operator itself. Our empirical analysis shows that even when its absolute predictions exhibit substantial errors, its responses to small, physically admissible input perturbations can closely match those of the reference dynamics in both direction and magnitude. This observation indicates that inaccurate absolute predictions can coexist with a reliable representation of local dynamical changes. By applying controlled input perturbations and comparing the corresponding outputs, we can extract these retained responses and use them as candidate correction directions grounded in the operator's retained dynamical knowledge.

Building on this observation, we introduce **PIVOT**, a general framework for output-space adaptation that exploits the local responses of a frozen neural operator. Given a target sample, we first obtain an initial prediction from the frozen operator and construct a set of small, physically admissible perturbations around the current input. For each perturbation direction, we query the operator with positive and negative perturbations and extract the local response from the difference between the paired predictions, reducing the influence of absolute prediction bias on response estimation. We then retain physically consistent responses and combine them to correct the current prediction. Finally, PIVOT checks the magnitude and physical residual of the resulting correction, and attenuates the correction whenever it increases the physical residual, yielding the final prediction. Throughout the entire adaptation process, the pretrained neural operator remains fully frozen.

We evaluate PIVOT on six datasets from APEBench (Koehler et al., 2024) and RealPDEBench (Hu et al., 2026), covering diverse physical systems and distribution shifts. Experiments demonstrate joint improvements in predictive accuracy and physical consistency, with long-horizon evaluations further showing improved rollout stability. These results support the use of retained local responses for effective adaptation of frozen neural operators.

Our contributions are threefold:

1. We empirically show that, under distribution shift, frozen neural operators can retain physically consistent local responses to small, physically admissible input perturbations despite substantial absolute prediction errors, revealing a source of dynamical information for model adaptation.
2. We introduce PIVOT, a framework for physically admissible prediction correction that extracts local responses through paired perturbations and learns spatially varying combina-

tions of these responses to refine predictions, incorporating physical information throughout adaptation while keeping the pretrained neural operator fully frozen.

3. We evaluate PIVOT on six datasets from APEBench and RealPDEBench, demonstrating improvements in predictive accuracy, physical consistency, and long-horizon rollout stability across diverse physical systems and distribution shifts.

2 Related Work

Neural operators and PDE surrogate modeling. Neural PDE solvers employ graph-based, spectral, wavelet, transformer, geometry-aware, latent, and reduced-order formulations to capture complex physical dynamics across resolutions, geometries, and temporal horizons (Li et al., 2020; Sanchez-Gonzalez et al., 2020; Cao, 2021; Gupta et al., 2021; Pfaff et al., 2021; Brandstetter et al., 2022; Lienen & Günnemann, 2022; Seidman et al., 2022; Huang et al., 2022; Rahman et al., 2023; Li et al., 2023b; Hao et al., 2023; Li et al., 2023a; Wang & Wang, 2024). Recent advances in large-scale pretraining, explicit temporal memory, and flexible spatial representations further support transferable physical modeling (Tran et al., 2023; Hao et al., 2024; Herde et al., 2024; Buitrago Ruiz et al., 2025; Zhou et al., 2026). Despite these expanded capabilities, accuracy within the training distribution does not establish physical consistency under new conditions. Shifts in boundary conditions, control parameters, geometry, or physical states can degrade predictions and introduce physical inconsistencies that accumulate over long rollouts. Adapting pretrained operators to shifted dynamics while assessing preservation of target-system physical properties therefore remains challenging.

Adaptation and physics-guided correction. Parameter-efficient adaptation uses lightweight trainable components, sparse updates, prompts, or low-rank parameterizations to adapt pretrained models at limited additional cost (Houlsby et al., 2019; Li & Liang, 2021; Lester et al., 2021; Mahabadi et al., 2021; Liu et al., 2022; Ben Zaken et al., 2022; Zhang et al., 2023; Dettmers et al., 2023; Kopiczko et al., 2024; Hayou et al., 2024; Jung et al., 2025). These mechanisms reduce trainable parameters without inherently constraining physical behavior. Scientific extensions employ contextual conditioning, compact parameter adaptation, spectral refinement, and physics-preserving transfer across systems and operating regimes (Kirchmeyer et al., 2022; Berman & Peherstorfer, 2024; Koupaï et al., 2024; Cao et al., 2025; Zhang et al., 2025; Yang & Ren, 2025). However, architectural dependencies restrict some mechanisms' direct reuse across heterogeneous backbones, while validation on selected equation families and transfer settings leaves effectiveness on substantially different systems unestablished. Broad applicability with consistent physical fidelity thus remains insufficiently validated. Other approaches incorporate physical laws into training and prediction refinement to promote physically consistent outputs (Wang et al., 2021; Li et al., 2024). Their success motivates explicit physical assessment, while exploiting frozen operators' retained local responses for adaptation across architectures and physical regimes remains open. PIVOT extracts informative responses to physically admissible perturbations from a frozen operator, learns sample-specific response combinations, and checks the assembled correction for physical consistency.

3 Methodology

3.1 Problem Formulation and Overview

We consider adapting a pretrained neural operator Φ_θ to a target distribution that differs from its training distribution. Given an input x, the operator produces the baseline prediction $y^{(0)} = \Phi_\theta(x)$. Our goal is to improve target-domain prediction accuracy and physical consistency while keeping the pretrained parameters θ fixed. Adaptation uses paired target-domain observations $(x, y^\star)$, where $y^\star$ is the observed output. The governing physical constraints are available during both training and inference, whereas $y^\star$ is used only during training.

PIVOT addresses this task by exploiting local responses retained by the frozen operator. A probe controller g_ψ learns physically admissible input perturbations, and paired operator queries reveal the resulting prediction changes. Responses that increase physical residuals receive lower weights, and a local readout h_φ learns how to combine the responses across space. The assembled correction is bounded in magnitude and checked against the baseline's physical residuals before acceptance. Only ψ and φ are optimized.

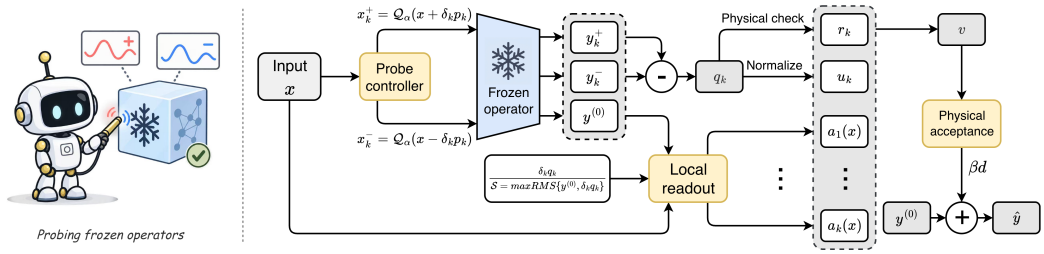


Figure 2: PIVOT extracts retained local responses from frozen operators for physically checked output correction.

3.2 Physically Admissible Probing

We learn input perturbations to reveal local responses useful for correcting the baseline prediction. For each input, the controller learns K physically admissible probe directions and their perturbation amplitudes. For probe k, g_ψ reads channelwise and low-frequency statistics of x and $y^{(0)}$ and predicts weights over a fixed collection of smooth fields. Adjusting these weights changes the perturbation's spatial pattern. The smooth fields limit rapid spatial oscillations, allowing the controller to search for structured input changes through a small set of coefficients.

We impose the perturbation constraints on each combined field and normalize valid fields to unit root-mean-square (RMS) magnitude, obtaining the probe direction p_k. The controller separately chooses a bounded perturbation size relative to the current input magnitude. We convert this relative size into the step δ_k, accounting for data normalization so that the perturbation magnitude follows the input scale in physical units. The paired predictions are

$$y_k^\pm = \Phi_\theta(\mathcal{Q}_x(x \pm \delta_k p_k)), \tag{1}$$

where $\mathcal{Q}_x$ enforces input feasibility. The resulting feasible inputs are denoted by $x_k^\pm$. The opposite perturbations reveal how the prediction changes in both directions around the same physical state.

Feasibility enforcement can alter the perturbation lengths. Let $h_k^\pm$ denote the signed projections of $x_k^\pm - x$ onto p_k. We account for these realized displacements when measuring the response:

$$\widetilde{q}_k = \frac{(h_k^-)^2(y_k^+ - y^{(0)}) - (h_k^+)^2(y_k^- - y^{(0)})}{h_k^- h_k^+ (h_k^- - h_k^+)}. \tag{2}$$

For unchanged symmetric perturbations, this reduces to $(y_k^+ - y_k^-)/(2\delta_k)$. The difference cancels an additive offset shared by the queried predictions, isolating their local variation. Using the realized displacement lengths accounts for changes introduced by feasibility enforcement.

We also check whether the feasible perturbations retain sufficient length and remain aligned with the intended direction. A perturbation that nearly disappears or turns away from that direction receives less weight. We apply this weight to the measured response and impose the prescribed response constraints, obtaining q_k.

3.3 Spatial Response Combination

We use physical consistency to determine each response's influence and learn spatial coefficients to match local correction needs. First, we compare the physical residuals of the two perturbed transitions with that of the baseline transition. A response is downweighted when its perturbed predictions increase the residual. Its weight decreases exponentially with the relative increase, with stronger attenuation for the same relative increase when the baseline residual is larger.

This physical assessment is combined with the checks on perturbation direction and length. Responses below the reliability threshold are rejected at inference. Training uses a smooth threshold so that poorly chosen probes can still receive gradients and improve. We then remove the parts of each response already represented by preceding responses and normalize the remaining field. This two-pass orthogonalization gives response fields u_k with comparable scales. Their final reliabilities $r_k \in [0, 1]$ include rejection and validity checks, with zero weight assigned to degenerate fields.

Prediction errors vary across space, so a response that helps in one region may require a different strength or sign elsewhere. The readout h_φ therefore predicts a spatial coefficient field a_k for each response. It receives x, $y^{(0)}$, and the response displacements $\delta_k q_k / S$, concatenated across channels. Here S is the largest RMS magnitude among the baseline and the response displacements, bounded below by a positive numerical floor. This scaling lets the readout assess response magnitudes relative to the current output scale.

The readout is a lightweight convolutional network. A hyperbolic-tangent output bounds its coefficients in $[-1, 1]$, and fixed averaging between neighboring locations smooths abrupt coefficient changes. The assembled correction is

$$v = \sum_k \mathcal{P}_x \left(\frac{A}{\sqrt{K}} \, r_k \, a_k \odot u_k \right), \tag{3}$$

where the sum covers all K probes and A is a fixed fraction of S. The symbol $\odot$ denotes pointwise multiplication, with spatial coefficients broadcast across output channels. The map $\mathcal{P}_x$ reapplies the response constraints after spatial weighting. The frozen operator supplies the response patterns, while the readout learns where to strengthen, weaken, or reverse their contributions.

3.4 Physical Acceptance

We check the complete correction because its physical effect depends on how the responses interact after spatial weighting. First, we uniformly rescale v if its RMS magnitude exceeds A. This bounds the total update while preserving its spatial pattern.

We then refine the correction using the physical residual. A damped least-squares step adjusts coordinates within the span of the assembled response components. An adjustment is accepted only when the overall residual decreases and no checked residual component increases beyond numerical tolerance. The refined field is bounded again by A, yielding the proposal d. This step improves the response combination without additional operator queries.

At inference, we evaluate the proposal under the original input. Let $\rho_c(x, y)$ denote the spatial RMS of physical residual component c for transition (x, y). Starting with the full correction, we successively halve its amplitude until every component satisfies

$$\rho_c(x, y^{(0)} + \beta d) \le (1 + \eta)\rho_c(x, y^{(0)}) + \zeta, \qquad \forall c, \tag{4}$$

where β is the tested correction fraction and η, ζ are nonnegative relative and absolute tolerances. Checking components separately prevents improvement in one physical condition from hiding deterioration in another. Candidates with nonfinite predictions or residuals are rejected. We retain the first passing fraction; if none passes within the prescribed number of halvings, β is set to zero. The final prediction is

$$\widehat{y} = y^{(0)} + \beta d. \tag{5}$$

When the baseline prediction and its residuals are finite, no checked residual component increases beyond the specified tolerance relative to the baseline. This final check evaluates the complete correction after spatial weighting and refinement.

3.5 Training

Training learns probes that expose useful responses and coefficients that turn them into accurate corrections:

$$\mathcal{L} = \mathcal{L}_y + \lambda_\rho \mathcal{L}_\rho + \lambda_d \mathcal{L}_d + \mathcal{L}_q. \tag{6}$$

$\mathcal{L}_y$ supervises $y^{(0)} + d$ against the observed target $y^\star$. $\mathcal{L}_\rho$ penalizes residual excess over the observed transition and componentwise deterioration from the baseline. $\mathcal{L}_d$ penalizes correction RMS magnitudes exceeding a prescribed multiple of the baseline error RMS. The nonnegative weights λ_ρ and λ_d control these penalties.

$\mathcal{L}_q$ combines error-reconstruction alignment and coefficient distillation with their respective weights. It encourages response combinations to reconstruct the baseline error and trains the readout toward clipped, smoothed local ridge-fit coefficients using a weighted smooth-ℓ_1 loss. For multiple probes, it also discourages correlated probes and responses. All terms reuse the same training targets.

Only ψ and φ are optimized, retaining input gradients through the frozen operator. Target-dependent coefficient fitting is used only during training. Each prediction requires $2K+1$ operator evaluations; correction refinement and acceptance checks require only physical-residual evaluations. Appendix C details the training protocol and computational cost.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Data and protocols. We evaluate PIVOT on six benchmarks from APEBench (Koehler et al., 2024) and RealPDEBench (Hu et al., 2026), spanning diverse dynamics, geometries, control conditions, and coupled systems. APEBench includes Burgers, Kuramoto Sivashinsky (KS), and Kolmogorov flow, while RealPDEBench includes Cylinder Flow, Controlled Cylinder Flow, and Fluid Structure Interaction (FSI). To assess generality across architectures, we apply PIVOT to three representative neural operators: FNO (Li et al., 2021), CNO (Raonić et al., 2023), and Transolver (Wu et al., 2024).

In the main experiments, we conduct two complementary comparisons. First, we evaluate adaptation effectiveness by comparing PIVOT with Bottleneck (Houlsby et al., 2019), LoRA (Hu et al., 2021), DoRA (Liu et al., 2024), GraLoRA (Jung et al., 2025), and RandLoRA (Albert et al., 2025) on the same pretrained backbones under shared training and evaluation settings, with trainable parameter budgets matched as closely as possible. Second, we assess overall predictive performance against strong neural operators: F-FNO (Tran et al., 2023), DPOT (Hao et al., 2024), Poseidon (Herde et al., 2024), MemNO (Buitrago Ruiz et al., 2025), and Transolver-3 (Zhou et al., 2026). In the cross-domain matching experiments, we additionally include POTT (Yang & Ren, 2025) and F-Adapter (Zhang et al., 2025) as adaptation baselines for physics-preserving transfer and frequency-aware operator adaptation, respectively. All methods share the same data splits and evaluation protocols. APEBench is evaluated with autoregressive rollouts, while RealPDEBench uses fixed forecast windows.

Metrics. Let $\hat{\mathbf{u}}_t$ and $\mathbf{u}_t$ denote predicted and reference states at forecast step t, with C channels and horizon T. We measure predictive accuracy on APEBench (Koehler et al., 2024) and RealPDEBench (Hu et al., 2026) using nRMSE and Rel. L2, respectively:

$$\mathrm{nRMSE}_t = \sum_{c=1}^{C} \frac{\|\hat{\mathbf{u}}_{t,c} - \mathbf{u}_{t,c}\|_2}{\|\mathbf{u}_{t,c}\|_2 + \epsilon}, \qquad \text{Rel. L2} = \frac{\|\hat{\mathbf{u}} - \mathbf{u}\|_2}{\|\mathbf{u}\|_2 + \epsilon}. \tag{7}$$

Here, $\hat{\mathbf{u}}$ and $\mathbf{u}$ contain all forecast steps, and $\epsilon > 0$ ensures numerical stability. Mean nRMSE takes the arithmetic mean of stepwise errors, whereas Rel. L2 jointly evaluates the complete forecast window; terminal variants use only the final step. Both H1 variants additionally account for first-order spatial derivatives: H1 nRMSE adds separately normalized derivative errors, while H1 Rel. L2 jointly normalizes field and gradient errors through the spatial H^1 norm.

To assess physical fidelity on APEBench(Koehler et al., 2024), we report PDE Defect and Balance Error, which average normalized discrete PDE residuals and global balance violations over the rollout, respectively. On RealPDEBench, Frequency Error (FE) measures temporal Fourier discrepancies after spatial aggregation, while Kinetic Energy Error (KE) measures the absolute discrepancy in velocity fluctuation energy (Hu et al., 2026).

We further evaluate distributional fidelity using the Jensen-Shannon distance:

$$\mathrm{JSD}_q = \sqrt{\tfrac{1}{2} D_{\mathrm{KL}}(p^q_{\mathrm{pred}} \| m^q) + \tfrac{1}{2} D_{\mathrm{KL}}(p^q_{\mathrm{gt}} \| m^q)}. \tag{8}$$

Here, p^q_{pred} and p^q_{gt} are predicted and reference distributions of physical quantity q, estimated over evaluated spatial locations and forecast steps using 200-bin histograms with shared bin edges; m^q is their equal mixture. For flow benchmarks with two velocity components, we report JSD for u, v, and vorticity magnitude $|\omega|$, together with their arithmetic mean.

Table 1: CNO-based adaptation on KS and Cylinder Flow.

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
Kuramoto–Sivashinsky Equation from APEBench Koehler et al. (2024)					
Bottleneck Houlsby et al. (2019)	0.5865	<u>1.7852</u>	1.6461	<u>0.0427</u>	9.089×10^{-4}
LoRA Hu et al. (2021)	0.5539	1.8089	1.5518	0.0432	9.596×10^{-4}
DoRA Liu et al. (2024)	0.5656	1.8294	1.6178	0.0438	9.912×10^{-4}
GraLoRA Jung et al. (2025)	0.6404	1.8588	1.8414	0.0444	1.241×10^{-3}
RandLoRA Albert et al. (2025)	<u>0.4297</u>	1.8600	<u>1.2056</u>	0.0457	1.047×10^{-3}
PIVOT	**0.1062**	**1.2409**	**0.2601**	**0.0127**	$\mathbf{1.150 \times 10^{-8}}$
Method	**Rel. ℓ_2 ↓**	**Terminal Rel. ℓ_2 ↓**	**H1 Rel. ℓ_2 ↓**	**FE ↓**	**KE ↓**
Cylinder Flow from RealPDEBench Hu et al. (2026)					
Bottleneck Houlsby et al. (2019)	0.0645	0.0891	0.4766	47.1043	6.066×10^{-4}
LoRA Hu et al. (2021)	0.0644	0.0885	0.4762	44.8914	6.018×10^{-4}
DoRA Liu et al. (2024)	0.0636	0.0883	0.4763	<u>41.3211</u>	5.870×10^{-4}
GraLoRA Jung et al. (2025)	<u>0.0634</u>	0.0881	<u>0.4758</u>	42.7354	5.976×10^{-4}
RandLoRA Albert et al. (2025)	0.0641	<u>0.0879</u>	0.4766	41.9609	<u>5.697×10^{-4}</u>
PIVOT	**0.0631**	**0.0875**	**0.4752**	**34.7501**	$\mathbf{4.209 \times 10^{-4}}$

4.2 Predictive Performance and Rollout Fidelity

PIVOT consistently improves both predictive accuracy and physical fidelity in the CNO based comparisons on KS and Cylinder Flow (Table 1). On KS, these gains remain visible throughout long horizon rollout: PIVOT preserves the characteristic spatial structures of the reference dynamics, while several competing adapters progressively collapse toward large, nearly uniform regions (Figure 3). This qualitative separation mirrors the quantitative results, indicating that PIVOT limits error accumulation while retaining the evolving structure of the solution. On Cylinder Flow, PIVOT achieves stronger physical fidelity even when predictive errors are closely matched, highlighting the importance of evaluating whether adapted predictions remain dynamically consistent in addition to measuring field error. The comparison against strong neural operators on KS further confirms that this advantage extends beyond adapter based baselines, with PIVOT maintaining a favorable combination of predictive accuracy and physical consistency (Table 2).

Taken together, these results support PIVOT's central premise that physical information retained in frozen operators can provide effective guidance for adaptation, improving predictive accuracy while preserving the dynamics over extended rollouts. Additional results across datasets and backbones are reported in Appendix B.1.

4.3 Local Response Fidelity and Probe Quality

Frozen operators retain useful local physical responses even when their absolute predictions require adaptation. Table 3 compares frozen FNO and numerical-solver responses across Burgers, KS, and Kolmogorov under identical admissible perturbations generated independently of model outputs and targets. Agreement in direction and magnitude before output projection or physics filtering supports retained physical sensitivity as a source of candidate correction directions.

Figure 5 examines how probe construction affects response quality. Random admissible probes are obtained by projecting random combinations of smooth basis functions; unconstrained probes are optimized without feasibility restrictions; learned admissible probes are optimized within PIVOT's feasible space. The first two perform similarly on Burgers, potentially reflecting the relatively mild kinematic restrictions of this periodic setting. Learned admissible probes yield more faithful responses and better recover the reference spatial structure, highlighting the value of learning informative feasible directions for physical correction. Appendix B.2 further shows that higher-order responses agree poorly with numerical-solver references, supporting PIVOT's reliance on the validated directional responses. Additional results and analyses are provided there.

4.4 Cross-Domain Matching under Distribution Shift

We evaluate cross-domain matching on Cylinder Flow by adapting a frozen FNO pretrained on CFD at $Re \in [1800, 6000]$ to experimental trajectories at $Re \in [9000, 12000]$, combining simulation-to-

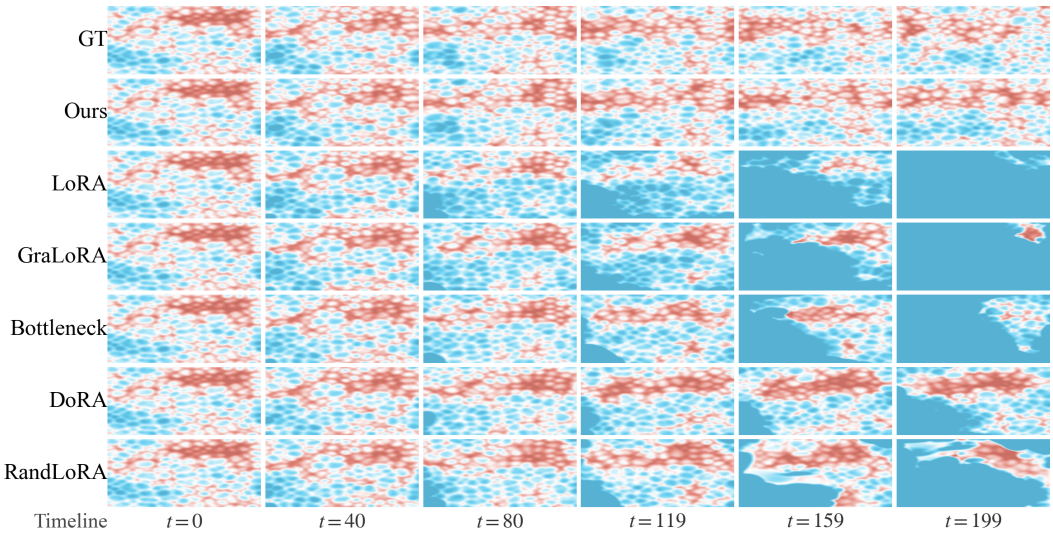


Figure 3: Qualitative comparison of different adaptation methods for long-horizon spatiotemporal prediction on the same ground-truth trajectory.

Table 2: Comparison with strong neural operators on KS.

Method	Mean nRMSE ↓	PDE Defect ↓	Balance Error ↓
DPOT	0.3949	0.0134	9.569×10^{-4}
FFNO	0.4195	**0.0105**	1.095×10^{-3}
MemNO	<u>0.3096</u>	0.0271	<u>9.365×10^{-4}</u>
Poseidon	0.4718	0.1150	5.254×10^{-3}
Transolver-3	0.8543	0.1102	1.407×10^{-3}
CNO+PIVOT	**0.1062**	<u>0.0127</u>	$\mathbf{1.150 \times 10^{-8}}$

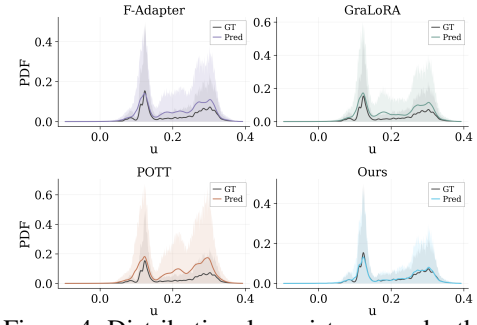


Figure 4: Distributional consistency under the transfer setting on Cylinder.

real transfer with parameter extrapolation. Target splits contain disjoint trajectories and Reynolds settings, with higher-Reynolds conditions reserved for testing. All adapters share the checkpoint, adaptation data, source normalization, and evaluation protocol.

Figure 6 shows that PIVOT leads all reported predictive and physical metrics. Notably, POTT Yang & Ren (2025) improves physical diagnostics but increases prediction errors relative to FNO, highlighting challenges in transferring physics-guided corrections across physical systems. PIVOT improves both, supporting response-based correction under the combined shift. Appendix B.3 reports protocols and Burgers results with diffusion halved from 1.5 to 0.75.

4.5 Physical Consistency under Cross-Domain Matching

We further examine whether this physical advantage extends to the velocity statistics of the target flow. Figure 4 visualizes the probability density functions (PDFs) of velocity component u on Cylinder Flow. These PDFs depict the empirical distributions whose discrepancy is quantified by JSD, using samples aggregated across evaluated spatial locations and forecast steps. PIVOT reproduces the reference distribution more faithfully across the dominant peak and the broader velocity range, capturing the relative prevalence of target-domain flow states.

Together with FE and KE, this distributional agreement establishes a consistent advantage across temporal spectral, energetic, and statistical properties. These complementary diagnostics show that PIVOT's improvements extend beyond field reconstruction to the physical characteristics of the target dynamics. The results support the central premise of PIVOT: local responses retained by a

Table 3: Frozen response accuracy of FNO against the numerical reference solver.

Dataset	Rel. error (%) ↓	Response cosine ↑	Amplitude ratio
Burgers	2.039	1.000	0.995
KS	1.764	1.000	0.998
Kolmogorov	7.620	0.997	0.980

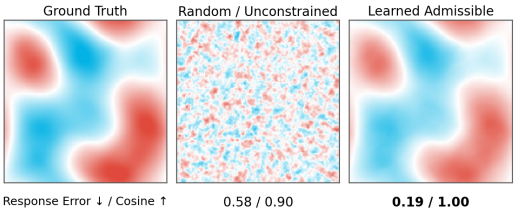


Figure 5: Learned admissible probing closely recovers the ground-truth response on Burgers.

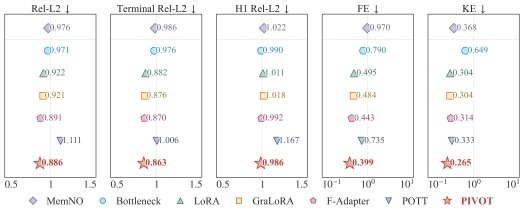


Figure 6: Cross-domain Cylinder performance: each metric normalized by frozen FNO.

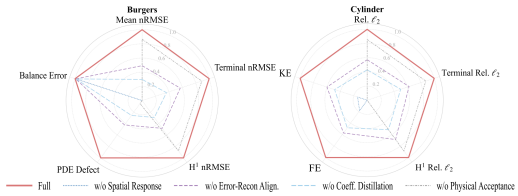


Figure 7: Component ablation on Burgers and Cylinder.

frozen source operator remain useful under distribution shift, and combining them under physical constraints enables accurate and physically consistent target-domain correction. Further analyses appear in Appendix B.4.

4.6 Ablation Study

We ablate four components on Burgers and Cylinder while fixing the backbone and training protocol. Each variant removes one component: spatially varying response combination (*w/o Spatial Response*), the error-reconstruction alignment objective (*w/o Error-Recon Align.*), supervision from fitted coefficients (*w/o Coeff. Distillation*), or the final physical check on the assembled correction (*w/o Physical Acceptance*).

Figure 7 shows that full PIVOT achieves the strongest overall predictive and physical performance, supporting the complementary contributions of response learning, spatial combination, and physical acceptance. The central finding is the separation between predictive accuracy and physical consistency: removing physical acceptance retains comparatively strong accuracy yet markedly weakens physical consistency on both systems. These results show that accurate field reconstruction alone does not ensure physical fidelity and support checking the assembled correction after spatial weighting. Additional ablations appear in Appendix B.5.

5 Conclusion

PIVOT builds on the finding that frozen neural operators can retain reliable local responses to physically admissible perturbations despite predictive degradation under distribution shift. Learned admissible probing and spatial response combination convert this retained physical information into target-domain corrections, while physical checks assess response reliability and constrain the assembled update. Experiments spanning six benchmarks and three backbones demonstrate improvements in predictive accuracy, physical consistency, and long-horizon fidelity, with further support from cross-domain matching. Physical diagnostics reveal substantial differences in physical behavior at comparable prediction errors, while ablations highlight the importance of final physical acceptance after spatial combination. These results support evaluating adaptation through both field accuracy and the physical properties of predicted dynamics. Together, these findings identify retained local responses as a useful resource for adaptation and show how learning to extract and combine them, coupled with explicit physical assessment, can extend the usefulness of a pretrained operator under new physical conditions while keeping its parameters frozen.

AI USE STATEMENT

AI tools assisted with research ideation and experimental design, drafting portions of the manuscript, and refining the writing. The authors reviewed and verified all AI-assisted content, including technical claims and experimental procedures, and take full responsibility for the final manuscript and reported results.

REFERENCES

Paul Albert, Frederic Zhang, Hemanth Saratchandran, Cristian Rodriguez-Opazo, Anton van den Hengel, and Ehsan Abbasnejad. Randlora: Full rank parameter-efficient fine-tuning of large models. In *International Conference on Learning Representations*, 2025.

Elad Ben Zaken, Yoav Goldberg, and Shauli Ravfogel. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pp. 1–9, 2022.

Jules Berman and Benjamin Peherstorfer. Colora: Continuous low-rank adaptation for reduced implicit neural modeling of parameterized partial differential equations. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 3565–3583, 2024.

Johannes Brandstetter, Daniel E. Worrall, and Max Welling. Message passing neural pde solvers. In *International Conference on Learning Representations*, 2022.

Ricardo Buitrago Ruiz, Tanya Marwah, Albert Gu, and Andrej Risteski. On the benefits of memory for modeling time-dependent pdes. In *International Conference on Learning Representations*, 2025.

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

Shuhao Cao, Francesco Brarda, Ruipeng Li, and Yuanzhe Xi. Spectral-refiner: Accurate fine-tuning of spatiotemporal fourier neural operator for turbulent flows. In *International Conference on Learning Representations*, 2025.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Somdatta Goswami, Katiana Kontolati, Michael D. Shields, and George Em Karniadakis. Deep transfer operator learning for partial differential equations under conditional shift. *Nature Machine Intelligence*, 4:1155–1164, 2022.

Gaurav Gupta, Xiongye Xiao, and Paul Bogdan. Multiwavelet-based operator learning for differential equations. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. Gnot: A general neural operator transformer for operator learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 12556–12569, 2023.

Zhongkai Hao, Chang Su, Songming Liu, Julius Berner, Chengyang Ying, Hang Su, Anima Anandkumar, Jian Song, and Jun Zhu. Dpot: Auto-regressive denoising operator transformer for large-scale pde pre-training. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 17616–17635, 2024.

Soufiane Hayou, Nikhil Ghosh, and Bin Yu. Lora+: Efficient low rank adaptation of large models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 17783–17806, 2024.

Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for pdes. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 2790–2799, 2019.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021. URL https://arxiv.org/abs/2106.09685.

Peiyan Hu, Haodong Feng, Hongyuan Liu, Tongtong Yan, Wenhao Deng, Tianrun Gao, Rong Zheng, Haoren Zheng, Chenglei Yu, Chuanrui Wang, Kaiwen Li, Zhi-Ming Ma, Dezhi Zhou, Xingcai Lu, Dixia Fan, and Tailin Wu. Realpdebench: A benchmark for complex physical systems with real-world data. In *International Conference on Learning Representations*, 2026.

Xiang Huang, Zhanhong Ye, Hongsheng Liu, Shi Ji, Zidong Wang, Kang Yang, Yang Li, Min Wang, Haotian Chu, Fan Yu, Bei Hua, Lei Chen, and Bin Dong. Meta-auto-decoder for solving parametric partial differential equations. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Yeonjoon Jung, Daehyun Ahn, Hyungjun Kim, Taesu Kim, and Eunhyeok Park. Gralora: Granular low-rank adaptation for parameter-efficient fine-tuning. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Matthieu Kirchmeyer, Yuan Yin, Jeremie Dona, Nicolas Baskiotis, Alain Rakotomamonjy, and Patrick Gallinari. Generalizing to new physical systems via context-informed dynamics model. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 11283–11301, 2022.

Felix Koehler, Simon Niedermayr, Rüdiger Westermann, and Nils Thuerey. APEBench: A benchmark for autoregressive neural emulators of PDEs. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Dawid Jan Kopiczko, Tijmen Blankevoort, and Yuki M. Asano. Vera: Vector-based random matrix adaptation. In *International Conference on Learning Representations*, 2024.

Armand Kassaï Koupaï, Jorge Mifsut Benet, Yuan Yin, Jean-Noël Vittaut, and Patrick Gallinari. Geps: Boosting generalization in parametric pde neural solvers through adaptive conditioning. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL http://jmlr.org/papers/v24/21-1524.html.

Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 3045–3059, 2021.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pp. 4582–4597, 2021.

Zijie Li, Dule Shu, and Amir Barati Farimani. Scalable transformer for pde surrogate modeling. In *Advances in Neural Information Processing Systems*, volume 36, 2023a.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Andrew Stuart, Kaushik Bhattacharya, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023b.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27, 2024. doi: 10.1145/3648506.

Marten Lienen and Stephan Günnemann. Learning the dynamics of physical systems from sparse observations with finite element networks. In *International Conference on Learning Representations*, 2022.

Phillip Lippe, Bas Veeling, Paris Perdikaris, Richard Turner, and Johannes Brandstetter. Pde-refiner: Achieving accurate long rollouts with neural pde solvers. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin Raffel. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Shih-Yang Liu, Chien-Yi Wang, Hongxu Yin, Pavlo Molchanov, Yu-Chiang Frank Wang, Kwang-Ting Cheng, and Min-Hung Chen. Dora: Weight-decomposed low-rank adaptation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 32100–32121, 2024.

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

Rabeeh Karimi Mahabadi, James Henderson, and Sebastian Ruder. Compacter: Efficient low-rank hypercomplex adapter layers. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter Battaglia. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations*, 2021.

Md Ashiqur Rahman, Zachary E. Ross, and Kamyar Azizzadenesheli. U-no: U-shaped neural operators. *Transactions on Machine Learning Research*, 2023.

Bogdan Raonić, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel de Bézenac. Convolutional neural operators for robust and accurate learning of pdes. *Advances in Neural Information Processing Systems*, 36:77187–77200, 2023.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 8459–8468, 2020.

Jacob H. Seidman, Georgios Kissas, Paris Perdikaris, and George J. Pappas. Nomad: Nonlinear manifold decoders for operator learning. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Daniel MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. Pdebench: An extensive benchmark for scientific machine learning. In *Advances in Neural Information Processing Systems*, volume 35, pp. 1596–1611, 2022.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators, 2023. URL `https://arxiv.org/abs/2111.13802`.

Sifan Wang, Hanwen Wang, and Paris Perdikaris. Learning the solution operator of parametric partial differential equations with physics-informed deeponets. *Science Advances*, 7(40):eabi8605, 2021. doi: 10.1126/sciadv.abi8605.

Tian Wang and Chuang Wang. Latent neural operator for solving forward and inverse pde problems. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53681–53705. PMLR, 2024.

Hao-Ran Yang and Chuan-Xian Ren. A physics-preserved transfer learning method for differential equations. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Hangwei Zhang, Chun Kang, Yan Wang, and Difan Zou. F-adapter: Frequency-adaptive parameter-efficient fine-tuning in scientific machine learning. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. Adaptive budget allocation for parameter-efficient fine-tuning. In *International Conference on Learning Representations*, 2023.

Jiawei Zhao, Zhenyu Zhang, Beidi Chen, Zhangyang Wang, Anima Anandkumar, and Yuandong Tian. Galore: Memory-efficient llm training by gradient low-rank projection. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 61121–61143, 2024.

Hang Zhou, Haixu Wu, Haonan Shangguan, Yuezhou Ma, Huikun Weng, Jianmin Wang, and Mingsheng Long. Transolver-3: Scaling up transformer solvers to industrial-scale geometries. *arXiv preprint arXiv:2602.04940*, 2026.

A Theoretical Analysis

PIVOT uses changes in a frozen operator's predictions to construct corrections. We show how these finite responses depend on nonlinear equation residuals, how response errors affect the assembled proposal, and when physical acceptance yields a more accurate prediction.

Let $F^\star$ be the target discrete evolution map, $y^\star = F^\star(x)$, $y_0 = y^{(0)} = \Phi_\theta(x)$, and $e = y^\star - y_0$. As in Section 3.4, the clipped proposal satisfies $\|d\| \le A < \infty$ and the returned prediction is $\widehat{y} = y_0 + \beta d$, where the search tests $\beta \in \{1, 2^{-1}, \dots, 2^{-J}\}$ and uses $\beta = 0$ for fallback. We analyze these discrete operations in exact arithmetic. All spaces are finite-dimensional; vector norms are RMS norms with their associated inner products and induced operator norms. Residuals retain their prescribed physical scaling.

A.1 Finite Responses under Nonlinear Dynamics

Fix a retained probe p with $\|p\| = 1$ and its realized feasible queries x_-, x, and x_+. Write $a = -\langle p, x_- - x\rangle > 0$, $b = \langle p, x_+ - x\rangle > 0$, $w_- = b/(a+b)$, and $w_+ = a/(a+b)$. The response stencil in Eq. (2) is

$$\mathcal{S}f = w_- \frac{f(x) - f(x_-)}{a} + w_+ \frac{f(x_+) - f(x)}{b}. \tag{9}$$

For probe k, let $g_k = \mathcal{S}_k F^\star$ and $\widetilde{q}_k = \mathcal{S}_k \Phi_\theta$ denote the target and frozen responses before weighting and constraints. With $E(z) = \Phi_\theta(z) - F^\star(z)$, linearity gives

$$\widetilde{q}_k - g_k = \mathcal{S}_k E. \tag{10}$$

Thus an additive prediction bias shared across the queries cancels. For nonlinear dynamics, the remaining response error can be related to how the governing equations change across those same queries.

Proposition 1. Suppose the full target discretization satisfies $\mathcal{H}(z, F^\star(z)) = 0$ at the three queries, and $\mathcal{H}$ is continuously differentiable in its output argument on the segments joining $F^\star(z)$ and $\Phi_\theta(z)$. Define the equation residual $R(z) = \mathcal{H}(z, \Phi_\theta(z))$ and the averaged Jacobian

$$M(z) = \int_0^1 D_y\mathcal{H}\big(z, F^\star(z) + tE(z)\big)\, dt. \tag{11}$$

For finite constants κ, L_M, assume $\|v\| \le \kappa\|M(x)v\|$ for every output increment v, and $\|M(x_\pm) - M(x)\| \le L_M\|x_\pm - x\|$. Set $B_E = \max_{z\in\{x_-,x,x_+\}} \|E(z)\|$ and $\chi = w_-\|x_- - x\|/a + w_+\|x_+ - x\|/b$. Then

$$\|\widetilde{q}_k - g_k\| \le \kappa\big(\|\mathcal{S}_k R\| + L_M B_E \chi\big). \tag{12}$$

Proof. The fundamental theorem of calculus gives $R(z) = M(z)E(z)$ exactly. Hence $M(x)\mathcal{S}_k E = \mathcal{S}_k R - \mathcal{S}_k((M - M(x))E)$. The second term has zero central value, so its norm is at most $L_M B_E \chi$. Apply the stability inequality for $M(x)$ and Eq. (10). □

The two terms have distinct meanings: $\mathcal{S}_k R$ measures variation in the equation residual, while $L_M B_E \chi$ measures how prediction error interacts with state-dependent dynamics. The factor χ records distortion of the probe by feasibility processing; straight probes have $\chi = 1$. Equation curvature and local variation of the two maps control L_M. If $D_y\mathcal{H}$ has Lipschitz constants K_x, K_y in its two arguments on the relevant region, and $F^\star, \Phi_\theta$ have local Lipschitz constants $L_\star, L_\Phi$, then integration in Eq. (11) gives $L_M \le K_x + K_y(L_\star + L_\Phi)/2$. Consequently, the nonlinear term stays bounded as the probes shrink when these constants and χ remain bounded. For a fixed linear residual $\mathcal{H}(z, y) = My - Cz - f$, $L_M = 0$ and $\kappa = \|M^{-1}\|$ recover the linear bound.

Connection to nonlinear PDEs and parameter shifts. At a fixed spatial discretization, Burgers, KS, and incompressible Navier–Stokes/Kolmogorov dynamics (Koehler et al., 2024) have the form $\dot{u} = Lu + \mathcal{B}(u, u) + f$, where $\mathcal{B}$ is bilinear; for incompressible flow, pressure is eliminated or velocity is restricted to the divergence-free space. Diffusion and other linear terms enter L, while quadratic nonlinear terms enter $\mathcal{B}$. An implicit step of size $\tau > 0$ provides a concrete instance of Proposition 1:

$$\mathcal{H}(z, y) = y - z - \tau\big[Ly + \mathcal{B}(y, y) + f\big]. \tag{13}$$

If $\|\mathcal{B}(u, v)\| \le C_\mathcal{B}\|u\|\|v\|$, then $K_x = 0$, $K_y = 2\tau C_\mathcal{B}$, and $L_M \le \tau C_\mathcal{B}(L_\star + L_\Phi)$. The additional response error is therefore bounded by $\kappa\tau C_\mathcal{B} B_E \chi(L_\star + L_\Phi)$: nonlinear advection couples value error to response error through the step size, local variation, and equation stability.

For source and target instances of this step with a common τ and $\mathcal{B}$ but coefficients (L_s, f_s) and (L, f), let R_s be the source-equation residual of the same frozen predictor. Subtracting the two equations yields

$$\mathcal{S}_k R = \mathcal{S}_k R_s + \tau(L_s - L)\mathcal{S}_k\Phi_\theta. \tag{14}$$

State-independent forcing shifts cancel from this contrast. For a diffusion shift, $L_s - L = (\nu_s - \nu)\Delta_h$, so the shift acts through spatial derivatives of the extracted response. These identities explain how accurate responses can coexist with substantial value error: source residual contrasts and the shifted linear response can stay small, while the nonlinear contribution is controlled explicitly by Eq. (12).

A.2 From Responses to Correction Quality

Fix the realized input, baseline, query geometry, paired-query physical scores, and budget A. Pairs rejected before stencil construction remain inactive. Let $\mathcal{A}$ map a raw response bank to the final clipped proposal, including response constraints, normalization, spatial weighting, and residual refinement. For $\widetilde{q} = (\widetilde{q}_k)_k$ and $g = (g_k)_k$, set $d = \mathcal{A}(\widetilde{q})$ and $\bar{d} = \mathcal{A}(g)$. Both evaluations use the fixed auxiliary quantities above and recompute all subsequent response-dependent operations. Assume both are defined; final clipping gives $\|d\|, \|\bar{d}\| \le A$.

Proposition 2. For every $\beta \in [0, 1]$,

$$\|y_0 + \beta d - y^\star\| \le (1 - \beta)\|e\| + \beta\big(\|\bar{d} - e\| + \|d - \bar{d}\|\big). \tag{15}$$

Proof. Write $y_0 + \beta d - y^\star = -(1 - \beta)e + \beta(\bar{d} - e) + \beta(d - \bar{d})$ and apply the triangle inequality. □

The first correction term measures the error when the same learned pipeline receives target responses. The second measures the effect of replacing them with frozen responses. This decomposition matches the roles of response alignment, coefficient distillation, and spatial combination: they train the pipeline to turn the available responses into a correction of the baseline error. If $\mathcal{A}$ is $L_{\mathcal{A}}$-Lipschitz between the two banks in the product norm $\|(v_k)_k\| = (\sum_k \|v_k\|^2)^{1/2}$, then $\|d - \bar{d}\| \le L_{\mathcal{A}}(\sum_k \|\widetilde{q}_k - g_k\|^2)^{1/2}$. Proposition 1 thus propagates nonlinear response error through the complete correction procedure.

Every $\beta \in (0, 1]$ improves accuracy when $\|\bar{d} - e\| + \|d - \bar{d}\| < \|e\|$. For a particular $\beta > 0$, the exact improvement condition is $2\langle d, e\rangle > \beta\|d\|^2$, obtained by expanding $\|\beta d - e\|^2$. The acceptance test determines which fractions satisfy the physical residual thresholds.

A.3 Physical Acceptance and Accuracy Improvement

Let $\rho_c(x, y) \ge 0$ be the finite collection of monitored residuals in Eq. (4), with $\eta, \zeta \ge 0$. The search accepts the first finite candidate satisfying every componentwise threshold and otherwise returns y_0. Assume the baseline and its residuals are finite, and the checked candidate is returned unchanged.

Theorem 3. Every returned field satisfies

$$\rho_c(x, \widehat{y}) \le (1 + \eta)\rho_c(x, y_0) + \zeta \qquad \text{for every } c. \tag{16}$$

Suppose additionally that each residual is finite and L_c-Lipschitz on the closed ball of radius A around y_0. For $d \neq 0$, let $s_c = \eta\rho_c(x, y_0) + \zeta$ and assume $s_c > 0$ whenever $L_c > 0$. Define

$$\beta_{\text{safe}} = \min\left\{1,\ \min_{c:L_c>0} \frac{s_c}{L_c\|d\|}\right\}, \tag{17}$$

with the inner minimum equal to $+\infty$ when all $L_c = 0$. If $2^{-J} \le \beta_{\text{safe}}$, the search accepts a positive fraction $\beta \ge \beta_{\text{safe}}/2$. The returned prediction strictly improves accuracy whenever $\|\bar{d} - e\| + \|d - \bar{d}\| < \|e\|$.

Proof. An accepted candidate satisfies Eq. (16) by construction; fallback satisfies it because η, ζ and the residuals are nonnegative. Every trial lies in the stated ball, and Lipschitz continuity gives $\rho_c(x, y_0 + \beta d) - \rho_c(x, y_0) \le L_c\beta\|d\|$. Thus all fractions $0 < \beta \le \beta_{\text{safe}}$ pass. The first grid fraction in this interval lies in $(\beta_{\text{safe}}/2, \beta_{\text{safe}}]$; descending search accepts it or a larger one. Proposition 2 then gives a strict decrease in prediction error under the stated correction condition. □

Residual slack determines how much of a proposal the test can accept. Correction quality determines the accuracy gained from that accepted fraction. The theorem connects these two parts of PIVOT's update rule.

A.4 Balance Preservation and Rollout Error

Why constraints are reapplied after spatial weighting. Suppose a fixed linear operator C satisfies $CF^\star(z) = b(z)$ at the queried inputs. Then $Cg_k = \mathcal{S}_k b$. If $b(x_-) = b(x) = b(x_+)$, the target response lies in $\ker C$, and scalar combinations preserve the baseline balance residual. Spatial weights can change that balance: $(1, -1)$ lies in the kernel of $C = (1, 1)$, but its pointwise product with $(1, 0)$ does not. This motivates applying $\mathcal{P}_x$ to the weighted responses in Eq. (3) and checking the complete candidate after refinement.

For a monitored nonlinear balance B_x, a tangent correction satisfying $D_y B_x(y_0)d = 0$ obeys $\|B_x(y_0 + \beta d) - B_x(y_0)\| \le H\beta^2\|d\|^2/2$ when $D_y B_x$ is H-Lipschitz along the update segment. This follows by integrating $[D_y B_x(y_0 + td) - D_y B_x(y_0)]d$ over $t \in [0, \beta]$. The final acceptance check evaluates the monitored balances at the complete candidate.

Propagation along an autoregressive trajectory. Let $x^\star_{t+1} = F^\star(x^\star_t)$ and define $E_t = \|\widehat{x}_t - x^\star_t\|$ and $\varepsilon_t = \|\widehat{x}_{t+1} - F^\star(\widehat{x}_t)\|$. Proposition 2 bounds ε_t at the input actually visited. For history-based predictors, include the history shift in both state maps and use the induced norm on the newly predicted block; the shifts coincide at a common input. If $\|F^\star(\widehat{x}_t) - F^\star(x^\star_t)\| \le L_t E_t$, with finite

$L_t \geq 0$, the triangle inequality gives $E_{t+1} \leq L_t E_t + \varepsilon_t$ and, by induction,

$$E_T \leq \left(\prod_{j=0}^{T-1} L_j \right) E_0 + \sum_{s=0}^{T-1} \left(\prod_{j=s+1}^{T-1} L_j \right) \varepsilon_s. \tag{18}$$

Empty products equal one. Each accepted local error is weighted by the subsequent sensitivity of the target dynamics, connecting response quality and correction accuracy to the resulting rollout.

Table 4: FNO-based adapter comparison on Burgers from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
Bottleneck Houlsby et al. (2019)	0.4910	1.3553	1.3104	0.0042	8.769×10^{-5}
LoRA Hu et al. (2021)	0.4907	1.3497	1.3064	0.0042	8.767×10^{-5}
DoRA Liu et al. (2024)	0.4909	1.3290	1.3071	0.0041	8.759×10^{-5}
GraLoRA Jung et al. (2025)	0.4802	1.3480	1.3058	0.0042	8.391×10^{-5}
RandLoRA Albert et al. (2025)	**0.4790**	**1.3208**	1.2983	0.0041	8.614×10^{-5}
PIVOT	0.4798	1.3352	**1.2818**	**0.0039**	$\mathbf{5.510 \times 10^{-9}}$

Table 5: CNO-based adapter comparison on Burgers from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
Bottleneck Houlsby et al. (2019)	0.3892	1.4751	1.8906	0.0227	2.975×10^{-4}
LoRA Hu et al. (2021)	0.3920	2.0132	1.9570	0.0169	1.115×10^{-3}
DoRA Liu et al. (2024)	0.3859	1.4957	1.8928	0.0234	3.618×10^{-4}
GraLoRA Jung et al. (2025)	0.3776	1.6218	1.8861	0.0242	3.290×10^{-4}
RandLoRA Albert et al. (2025)	0.3789	1.5215	1.9026	0.0221	3.598×10^{-4}
PIVOT	**0.2990**	**1.1647**	**0.9150**	**0.0056**	$\mathbf{5.123 \times 10^{-9}}$

Table 6: Adapter comparison using Transolver Wu et al. (2024) as the backbone on Burgers from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
Bottleneck Houlsby et al. (2019)	0.8864	12.6576	11.0639	0.9640	1.875×10^{-4}
LoRA Hu et al. (2021)	0.7979	5.4011	4.9190	0.2442	1.627×10^{-4}
DoRA Liu et al. (2024)	0.9173	14.5891	11.5001	1.1075	2.895×10^{-4}
GraLoRA Jung et al. (2025)	1.0660	15.5771	14.7781	1.1820	9.314×10^{-4}
RandLoRA Albert et al. (2025)	1.1711	15.9498	17.4025	1.3693	3.881×10^{-4}
PIVOT	**0.5266**	**2.4747**	**2.6891**	**0.0632**	$\mathbf{5.829 \times 10^{-9}}$

Table 7: Strong neural-operator comparison on Burgers from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	PDE Defect ↓	Balance Error ↓
DPOT Hao et al. (2024)	0.0748	0.0019	8.678×10^{-5}
F-FNO Tran et al. (2023)	0.0795	0.0015	9.932×10^{-5}
MemNO Buitrago Ruiz et al. (2025)	**0.0159**	**0.0004**	1.567×10^{-5}
Poseidon Herde et al. (2024)	0.5551	0.0102	2.750×10^{-4}
Transolver-3 Zhou et al. (2026)	1.1762	0.0276	5.670×10^{-4}
FNO+PIVOT	0.4798	0.0039	$\mathbf{5.510 \times 10^{-9}}$
CNO+PIVOT	0.2990	0.0056	$\mathbf{5.123 \times 10^{-9}}$
Transolver+PIVOT	0.5266	0.0632	5.829×10^{-9}

Table 8: Backbone comparison with and without PIVOT on Burgers from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
FNO	0.4911	1.3569	1.3086	0.0042	8.781×10^{-5}
FNO+PIVOT	**0.4798**	**1.3352**	**1.2818**	**0.0039**	$\mathbf{5.510 \times 10^{-9}}$
CNO	0.6825	1.5122	2.2923	0.0227	3.567×10^{-4}
CNO+PIVOT	**0.2990**	**1.1647**	**0.9150**	**0.0056**	$\mathbf{5.123 \times 10^{-9}}$
Transolver Wu et al. (2024)	0.8818	2.9641	3.6356	0.0995	2.174×10^{-4}
Transolver+PIVOT	**0.5266**	**2.4747**	**2.6891**	**0.0632**	$\mathbf{5.829 \times 10^{-9}}$

Table 9: FNO-based adapter comparison on Kuramoto–Sivashinsky (KS) from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
Bottleneck Houlsby et al. (2019)	0.4769	1.3332	1.1471	0.0270	1.204×10^{-4}
LoRA Hu et al. (2021)	0.4769	1.3318	1.1462	0.0270	1.204×10^{-4}
DoRA Liu et al. (2024)	0.4752	1.3356	1.1424	0.0269	1.210×10^{-4}
GraLoRA Jung et al. (2025)	0.4768	1.3316	1.1473	0.0271	1.198×10^{-4}
RandLoRA Albert et al. (2025)	0.4764	1.3247	1.1444	0.0269	1.198×10^{-4}
PIVOT	**0.4665**	**1.3177**	**1.1044**	**0.0202**	$\mathbf{1.076 \times 10^{-8}}$

Table 10: Adapter comparison using Transolver Wu et al. (2024) as the backbone on Kuramoto–Sivashinsky (KS) from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
Bottleneck Houlsby et al. (2019)	0.7447	1.4376	2.1201	0.0341	7.850×10^{-4}
LoRA Hu et al. (2021)	0.7487	1.4870	1.6557	0.2395	9.365×10^{-4}
DoRA Liu et al. (2024)	0.7511	1.4896	2.1407	0.0353	8.750×10^{-4}
GraLoRA Jung et al. (2025)	0.7748	1.3139	2.7157	0.4166	1.626×10^{-3}
RandLoRA Albert et al. (2025)	0.7984	1.5232	2.3743	0.0372	9.416×10^{-4}
PIVOT	**0.3106**	**1.0002**	**0.9127**	**0.0109**	$\mathbf{6.449 \times 10^{-9}}$

Table 11: Backbone comparison with and without PIVOT on Kuramoto–Sivashinsky (KS) from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
FNO	0.4769	1.3336	1.1467	0.0270	1.204×10^{-4}
FNO+PIVOT	**0.4665**	**1.3177**	**1.1044**	**0.0202**	$\mathbf{1.076 \times 10^{-8}}$
CNO	0.4659	1.3356	1.1844	0.0271	9.365×10^{-4}
CNO+PIVOT	**0.1062**	**1.2409**	**0.2601**	**0.0127**	$\mathbf{1.150 \times 10^{-8}}$
Transolver Wu et al. (2024)	0.7437	**1.0001**	1.6052	0.0193	8.146×10^{-4}
Transolver+PIVOT	**0.3106**	1.0002	**0.9127**	**0.0109**	$\mathbf{6.449 \times 10^{-9}}$

Table 12: FNO-based adapter comparison on Kolmogorov Flow from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
Bottleneck Houlsby et al. (2019)	0.4123	1.1577	0.9987	0.0193	1.345×10^{-4}
LoRA Hu et al. (2021)	0.4117	1.1565	0.9979	0.0194	1.345×10^{-4}
DoRA Liu et al. (2024)	0.4110	1.1521	0.9974	0.0188	1.345×10^{-4}
GraLoRA Jung et al. (2025)	0.4112	1.1561	0.9980	0.0189	1.331×10^{-4}
RandLoRA Albert et al. (2025)	0.4111	1.1503	0.9964	0.0185	1.339×10^{-4}
PIVOT	**0.4054**	**0.6537**	**0.9950**	**0.0182**	$\mathbf{1.129 \times 10^{-8}}$

Table 13: CNO-based adapter comparison on Kolmogorov Flow from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
Bottleneck Houlsby et al. (2019)	0.6640	1.4621	1.7858	0.1720	8.276×10^{-4}
LoRA Hu et al. (2021)	0.6565	1.3339	1.7616	0.3504	9.071×10^{-4}
DoRA Liu et al. (2024)	0.6581	1.4984	1.7800	0.1767	9.025×10^{-4}
GraLoRA Jung et al. (2025)	0.6752	1.2539	1.8254	0.3766	1.205×10^{-3}
RandLoRA Albert et al. (2025)	**0.6298**	1.5234	1.6801	0.1840	9.538×10^{-4}
PIVOT	0.6403	**0.9039**	**1.5727**	**0.0654**	$\mathbf{1.095 \times 10^{-8}}$

Table 14: Adapter comparison using Transolver Wu et al. (2024) as the backbone on Kolmogorov Flow from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
Bottleneck Houlsby et al. (2019)	**0.7819**	1.4218	2.2438	0.1334	4.756×10^{-4}
LoRA Hu et al. (2021)	0.7031	1.4517	2.1232	0.1993	5.202×10^{-4}
DoRA Liu et al. (2024)	0.7388	1.4731	2.2656	0.1381	5.301×10^{-4}
GraLoRA Jung et al. (2025)	0.7387	1.4556	2.4982	0.2292	7.580×10^{-4}
RandLoRA Albert et al. (2025)	0.7597	1.5063	2.5128	0.1456	5.704×10^{-4}
PIVOT	0.7093	**0.9826**	**1.2367**	**0.0220**	$\mathbf{5.814 \times 10^{-9}}$

Table 15: Strong neural-operator comparison on Kolmogorov Flow from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	PDE Defect ↓	Balance Error ↓
DPOT Hao et al. (2024)	0.4230	0.0600	1.956×10^{-4}
F-FNO Tran et al. (2023)	0.4494	0.0471	2.238×10^{-4}
MemNO Buitrago Ruiz et al. (2025)	3.4290	0.0754	5.290×10^{-4}
Poseidon Herde et al. (2024)	0.7256	0.0968	1.059×10^{-3}
Transolver-3 Zhou et al. (2026)	0.7922	0.0752	5.644×10^{-4}
FNO+PIVOT	**0.4054**	**0.0182**	1.129×10^{-8}
CNO+PIVOT	0.6403	0.0654	$\mathbf{1.095 \times 10^{-8}}$
Transolver+PIVOT	0.7093	0.0220	$\mathbf{5.814 \times 10^{-9}}$

Table 16: Backbone comparison with and without PIVOT on Kolmogorov Flow from APEBench Koehler et al. (2024).

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
FNO	0.4113	1.1580	0.9983	0.0191	1.346×10^{-4}
FNO+PIVOT	**0.4054**	**0.6537**	**0.9950**	**0.0182**	$\mathbf{1.129 \times 10^{-8}}$
CNO	0.7095	1.0939	1.7377	0.1092	8.527×10^{-4}
CNO+PIVOT	**0.6403**	**0.9039**	**1.5727**	**0.0654**	$\mathbf{1.095 \times 10^{-8}}$
Transolver Wu et al. (2024)	0.8009	0.9890	1.6988	0.0756	4.935×10^{-4}
Transolver+PIVOT	**0.7093**	**0.9826**	**1.2367**	**0.0220**	$\mathbf{5.814 \times 10^{-9}}$

Table 17: FNO-based adapter comparison on Controlled Cylinder Flow from RealPDEBench Hu et al. (2026).

Method	Rel. L_2 ↓	Terminal Rel. L_2 ↓	H1 Rel. L_2 ↓	FE ↓	KE ↓
Bottleneck Houlsby et al. (2019)	0.0593	0.0839	0.4116	15.8656	3.284×10^{-5}
LoRA Hu et al. (2021)	0.0590	0.0836	0.4134	15.5003	3.261×10^{-5}
DoRA Liu et al. (2024)	0.0576	0.0833	0.4110	16.0729	3.247×10^{-5}
GraLoRA Jung et al. (2025)	**0.0574**	**0.0831**	**0.4099**	15.7838	3.246×10^{-5}
RandLoRA Albert et al. (2025)	0.0591	0.0837	0.4128	16.0866	3.275×10^{-5}
PIVOT	0.0579	0.0835	0.4103	**12.7983**	$\mathbf{3.068 \times 10^{-5}}$

Table 18: CNO-based adapter comparison on Controlled Cylinder Flow from RealPDEBench Hu et al. (2026).

Method	Rel. L_2 ↓	Terminal Rel. L_2 ↓	H1 Rel. L_2 ↓	FE ↓	KE ↓
Bottleneck Houlsby et al. (2019)	0.0563	0.0812	0.3901	15.4885	3.102×10^{-5}
LoRA Hu et al. (2021)	0.0560	0.0807	0.3895	15.0032	3.096×10^{-5}
DoRA Liu et al. (2024)	0.0558	0.0809	0.3892	15.2129	3.103×10^{-5}
GraLoRA Jung et al. (2025)	**0.0556**	**0.0801**	**0.3891**	14.9456	3.102×10^{-5}
RandLoRA Albert et al. (2025)	0.0562	0.0808	0.3897	15.2236	3.112×10^{-5}
PIVOT	0.0559	0.0803	0.3892	**13.2137**	$\mathbf{2.958 \times 10^{-5}}$

Table 19: Adapter comparison using Transolver Wu et al. (2024) as the backbone on Controlled Cylinder Flow from RealPDEBench Hu et al. (2026).

Method	Rel. L_2 ↓	Terminal Rel. L_2 ↓	H1 Rel. L_2 ↓	FE ↓	KE ↓
Bottleneck Houlsby et al. (2019)	0.0526	0.0765	0.3666	14.8284	2.900×10^{-5}
LoRA Hu et al. (2021)	0.0526	0.0768	0.3666	14.8295	2.897×10^{-5}
DoRA Liu et al. (2024)	0.0527	0.0762	0.3663	15.2688	2.896×10^{-5}
GraLoRA Jung et al. (2025)	0.0526	0.0765	0.3662	15.0606	2.899×10^{-5}
RandLoRA Albert et al. (2025)	0.0526	0.0757	0.3661	15.1467	2.898×10^{-5}
PIVOT	**0.0514**	**0.0753**	**0.3623**	**13.0498**	$\mathbf{2.714 \times 10^{-5}}$

Table 20: Strong neural-operator comparison on Controlled Cylinder Flow from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
DPOT Hao et al. (2024)	0.0633	18.4904	3.315×10^{-5}
F-FNO Tran et al. (2023)	<u>0.0535</u>	16.2784	<u>2.908×10^{-5}</u>
MemNO Buitrago Ruiz et al. (2025)	0.0780	23.0411	3.846×10^{-5}
Poseidon Herde et al. (2024)	0.0572	16.7222	3.132×10^{-5}
Transolver-3 Zhou et al. (2026)	0.0586	16.0429	3.208×10^{-5}
FNO+PIVOT	0.0579	**12.7983**	3.068×10^{-5}
CNO+PIVOT	0.0559	13.2137	2.958×10^{-5}
Transolver+PIVOT	**0.0514**	<u>13.0498</u>	$\mathbf{2.714 \times 10^{-5}}$

Table 21: Backbone comparison with and without PIVOT on Controlled Cylinder Flow from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	Terminal Rel. $L_2 \downarrow$	H1 Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
FNO	0.0606	0.0855	0.4113	23.5055	3.290×10^{-5}
FNO+PIVOT	**0.0579**	**0.0835**	**0.4103**	**12.7983**	$\mathbf{3.068 \times 10^{-5}}$
CNO	0.0572	0.0821	0.3916	19.8033	3.189×10^{-5}
CNO+PIVOT	**0.0559**	**0.0803**	**0.3892**	**13.2137**	$\mathbf{2.958 \times 10^{-5}}$
Transolver Wu et al. (2024)	0.0528	0.0766	0.3661	16.2371	2.919×10^{-5}
Transolver+PIVOT	**0.0514**	**0.0753**	**0.3623**	**13.0498**	$\mathbf{2.714 \times 10^{-5}}$

Table 22: FNO-based adapter comparison on Cylinder Flow from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	Terminal Rel. $L_2 \downarrow$	H1 Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
Bottleneck Houlsby et al. (2019)	0.0666	0.0896	0.5023	<u>45.5330</u>	5.406×10^{-4}
LoRA Hu et al. (2021)	0.0665	0.0895	0.5022	45.6907	<u>5.403×10^{-4}</u>
DoRA Liu et al. (2024)	0.0660	0.0893	0.5015	45.5889	5.506×10^{-4}
GraLoRA Jung et al. (2025)	**0.0651**	**0.0891**	**0.4998**	46.3588	5.554×10^{-4}
RandLoRA Albert et al. (2025)	0.0658	0.0884	0.5026	45.9936	5.529×10^{-4}
PIVOT	<u>0.0652</u>	<u>0.0882</u>	<u>0.5013</u>	**38.4092**	$\mathbf{5.138 \times 10^{-4}}$

Table 23: Adapter comparison using Transolver Wu et al. (2024) as the backbone on Cylinder Flow from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	Terminal Rel. $L_2 \downarrow$	H1 Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
Bottleneck Houlsby et al. (2019)	0.0711	0.0935	0.4612	152.6867	2.650×10^{-4}
LoRA Hu et al. (2021)	0.0708	0.0936	0.4609	<u>151.2803</u>	2.652×10^{-4}
DoRA Liu et al. (2024)	0.0705	0.0938	0.4606	151.3007	2.648×10^{-4}
GraLoRA Jung et al. (2025)	0.0704	0.0936	<u>0.4604</u>	153.5099	<u>2.647×10^{-4}</u>
RandLoRA Albert et al. (2025)	<u>0.0702</u>	<u>0.0934</u>	0.4604	151.5093	2.647×10^{-4}
PIVOT	**0.0698**	**0.0931**	**0.4600**	**125.7553**	$\mathbf{2.468 \times 10^{-4}}$

Table 24: Strong neural-operator comparison on Cylinder Flow from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
DPOT Hao et al. (2024)	0.0725	104.2704	2.998×10^{-4}
F-FNO Tran et al. (2023)	**0.0629**	64.8513	3.353×10^{-4}
MemNO Buitrago Ruiz et al. (2025)	0.0889	105.8922	3.206×10^{-4}
Poseidon Herde et al. (2024)	0.0768	170.4426	2.796×10^{-4}
Transolver-3 Zhou et al. (2026)	0.0768	162.6757	$\mathbf{2.741 \times 10^{-4}}$
FNO+PIVOT	0.0652	38.4092	5.138×10^{-4}
CNO+PIVOT	0.0631	**34.7501**	4.209×10^{-4}
Transolver+PIVOT	0.0698	125.7553	$\mathbf{2.468 \times 10^{-4}}$

Table 25: Backbone comparison with and without PIVOT on Cylinder Flow from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	Terminal Rel. $L_2 \downarrow$	H1 Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
FNO	0.0674	0.0905	**0.5010**	51.3662	5.420×10^{-4}
FNO+PIVOT	**0.0652**	**0.0882**	0.5013	**38.4092**	$\mathbf{5.138 \times 10^{-4}}$
CNO	0.0652	0.0891	0.4781	52.4399	4.903×10^{-4}
CNO+PIVOT	**0.0631**	**0.0875**	**0.4752**	**34.7501**	$\mathbf{4.209 \times 10^{-4}}$
Transolver Wu et al. (2024)	0.0709	0.0944	0.4609	153.8597	2.661×10^{-4}
Transolver+PIVOT	**0.0698**	**0.0931**	**0.4600**	**125.7553**	$\mathbf{2.468 \times 10^{-4}}$

Table 26: FNO-based adapter comparison on Fluid–Structure Interaction (FSI) from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	Terminal Rel. $L_2 \downarrow$	H1 Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
Bottleneck Houlsby et al. (2019)	0.0671	0.1022	0.3171	6.9447	5.097×10^{-5}
LoRA Hu et al. (2021)	0.0655	0.1015	0.3133	6.5355	5.031×10^{-5}
DoRA Liu et al. (2024)	0.0661	0.1010	0.3134	6.6155	5.033×10^{-5}
GraLoRA Jung et al. (2025)	**0.0654**	**0.1006**	**0.3130**	6.5957	5.032×10^{-5}
RandLoRA Albert et al. (2025)	0.0663	0.1012	0.3140	6.4579	5.038×10^{-5}
PIVOT	0.0658	0.1008	0.3132	**5.6591**	$\mathbf{4.500 \times 10^{-5}}$

Table 27: CNO-based adapter comparison on Fluid–Structure Interaction (FSI) from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	Terminal Rel. $L_2 \downarrow$	H1 Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
Bottleneck Houlsby et al. (2019)	0.0633	0.0962	0.3001	**5.1780**	4.837×10^{-5}
LoRA Hu et al. (2021)	0.0629	0.0958	0.2987	5.3955	4.775×10^{-5}
DoRA Liu et al. (2024)	**0.0618**	**0.0948**	**0.2938**	5.3828	4.700×10^{-5}
GraLoRA Jung et al. (2025)	0.0621	0.0950	0.2951	5.3136	4.724×10^{-5}
RandLoRA Albert et al. (2025)	0.0633	0.0963	0.3002	5.6891	4.854×10^{-5}
PIVOT	0.0620	0.0949	0.2954	5.2708	$\mathbf{4.532 \times 10^{-5}}$

Table 28: Adapter comparison using Transolver Wu et al. (2024) as the backbone on Fluid–Structure Interaction (FSI) from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	Terminal Rel. $L_2 \downarrow$	H1 Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
Bottleneck Houlsby et al. (2019)	0.0576	0.0893	0.2680	6.4653	4.348×10^{-5}
LoRA Hu et al. (2021)	0.0569	0.0891	0.2678	<u>6.3651</u>	4.352×10^{-5}
DoRA Liu et al. (2024)	<u>0.0566</u>	0.0890	0.2678	6.7711	4.352×10^{-5}
GraLoRA Jung et al. (2025)	0.0568	<u>0.0890</u>	<u>0.2676</u>	6.4996	4.347×10^{-5}
RandLoRA Albert et al. (2025)	0.0570	0.0891	0.2677	6.6283	<u>4.345×10^{-5}</u>
PIVOT	**0.0564**	**0.0888**	**0.2674**	**5.6695**	$\mathbf{4.252 \times 10^{-5}}$

Table 29: Strong neural-operator comparison on Fluid–Structure Interaction (FSI) from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
DPOT Hao et al. (2024)	0.1095	9.6350	6.543×10^{-5}
F-FNO Tran et al. (2023)	<u>0.0578</u>	6.1579	<u>4.431×10^{-5}</u>
MemNO Buitrago Ruiz et al. (2025)	0.0983	12.8920	6.987×10^{-5}
Poseidon Herde et al. (2024)	0.0698	7.4390	5.306×10^{-5}
Transolver-3 Zhou et al. (2026)	0.0691	6.8516	5.407×10^{-5}
FNO+PIVOT	0.0658	<u>5.6591</u>	4.500×10^{-5}
CNO+PIVOT	0.0620	**5.2708**	4.532×10^{-5}
Transolver+PIVOT	**0.0564**	5.6695	$\mathbf{4.252 \times 10^{-5}}$

Table 30: Backbone comparison with and without PIVOT on Fluid–Structure Interaction (FSI) from RealPDEBench Hu et al. (2026).

Method	Rel. $L_2 \downarrow$	Terminal Rel. $L_2 \downarrow$	H1 Rel. $L_2 \downarrow$	FE $\downarrow$	KE $\downarrow$
FNO	0.0686	0.1031	0.3175	8.7019	5.114×10^{-5}
FNO+PIVOT	**0.0658**	**0.1008**	**0.3132**	**5.6591**	$\mathbf{4.500 \times 10^{-5}}$
CNO	0.0662	0.0996	0.3043	10.3278	4.974×10^{-5}
CNO+PIVOT	**0.0620**	**0.0949**	**0.2954**	**5.2708**	$\mathbf{4.532 \times 10^{-5}}$
Transolver Wu et al. (2024)	0.0567	0.0891	0.2678	7.1668	4.360×10^{-5}
Transolver+PIVOT	**0.0564**	**0.0888**	**0.2674**	**5.6695**	$\mathbf{4.252 \times 10^{-5}}$

B Additional Experimental Results

B.1 Expanded Analysis of Predictive Performance and Rollout Fidelity

We extend the main-paper analysis across equations and backbone architectures to assess PIVOT's predictive accuracy and physical fidelity. PIVOT keeps the pretrained operator frozen throughout adaptation, following the same dataset-specific training and evaluation protocols as in the main paper. We examine predictive accuracy alongside complementary physical diagnostics to determine whether accurate field reconstruction also preserves the behavior of the underlying dynamics.

APEBench: balance preservation across equations and backbones. Across Burgers, KS, and Kolmogorov Flow, PIVOT's clearest recurring advantage is substantially stronger balance preservation (Tables 8, 11, and 16). Balance Error consistently reaches the 10^{-8} scale or below across FNO, CNO, and Transolver, while standard adapters typically remain several orders of magnitude higher. These gains accompany competitive or improved rollout accuracy in most settings, showing that stronger balance preservation can coexist with accurate forecasting. The complete comparisons in Tables 4 through 14 establish that this pattern extends across both equations and backbone families.

PDE Defect measures how these balance gains relate to equation satisfaction. Across all nine equation–backbone pairs, PIVOT lowers both Balance Error and PDE Defect relative to the frozen backbone. The two metrics capture different properties: global balance and local equation residuals. Their changes also differ in magnitude. On Burgers with FNO, for example, Balance Error falls from 8.781×10^{-5} to 5.510×10^{-9}, while PDE Defect decreases from 0.0042 to 0.0039. Reporting both diagnostics makes clear how the correction changes each aspect of the predicted dynamics.

RealPDEBench: temporal and energetic fidelity. Across Controlled Cylinder Flow, Cylinder Flow, and FSI, PIVOT reduces both FE and KE for every evaluated backbone and system relative to the corresponding frozen model (Tables 21, 25, and 30). These physical gains frequently occur alongside closely matched Rel. L_2, terminal, and H^1 errors across adapters. This separation is central to evaluating adaptation: similar field accuracy can conceal substantial differences in frequency and energy behavior.

These results extend the APEBench findings to complementary properties of flow dynamics. Together, the two benchmark suites show that PIVOT's physical advantages extend beyond a single diagnostic, while also demonstrating why conventional prediction metrics provide an incomplete account of adaptation quality.

Comparison with strong neural operators. The independently trained operator comparisons further clarify this advantage. On APEBench, some operators achieve lower prediction error or PDE Defect, while PIVOT exhibits markedly stronger balance preservation (Tables 7 and 15). On RealPDEBench, PIVOT-equipped backbones achieve the lowest FE in each reported comparison, with FNO+PIVOT leading on Controlled Cylinder Flow and CNO+PIVOT leading on Cylinder Flow and FSI (Tables 20, 24, and 29), although other operators lead on selected field or energy metrics. These differing metric-wise rankings reinforce the need to evaluate physical adaptation through complementary diagnostics and to distinguish which properties each method preserves most effectively.

Imp lications for physical adaptation. The expanded evidence supports the central premise of PIVOT: local responses retained by a frozen operator can support corrections that improve physical fidelity while maintaining competitive predictive accuracy. Together with the response analysis and ablations in the main paper, these results support the complementary roles of extracting informative responses, learning their spatial combination, and assessing the physical effect of the assembled correction. The recurring gains across equations and architectures show the breadth of this approach, while the differing metric rankings reveal how its effects depend on the system and backbone. Physical consistency provides essential evidence of adaptation quality beyond field accuracy alone.

B.2 Additional Analysis of Local Response Fidelity and Probe Quality

We examine the physical information that PIVOT extracts from a frozen operator and the conditions under which that information supports correction. The analysis connects three aspects of the method: fidelity to numerical-solver responses, selection of informative admissible probes, and the need to assess the physical effect of the assembled update. Cross-backbone and probe-scale evaluations test the robustness of retained responses, while a higher-order audit checks whether their fidelity extends to second-order local information.

Response fidelity and physical diagnostics. For an input x, probe direction q, and perturbation magnitude $h > 0$, we query the frozen operator F_θ and numerical solver $\mathcal{S}$ at the same perturbed inputs $x \pm hq$ and compare their symmetric directional responses:

$$r_{G,h}(x, q) = \frac{G(x + hq) - G(x - hq)}{2h}, \qquad G \in \{F_\theta, \mathcal{S}\}. \tag{19}$$

All response comparisons use outputs before projection or physics filtering. We measure agreement through relative L_2 error, cosine similarity, and amplitude ratio. The relative error divides the norm of the response difference by the solver-response norm; the amplitude ratio divides the frozen-response norm by the solver-response norm. Better agreement means lower relative error, with cosine similarity and amplitude ratio both closer to one. To assess sensitivity to probe magnitude, we also replace h with $h/2$ and $2h$ and measure the L_2 change in the frozen response, normalized by its norm at h. Table 31 reports solver agreement and these scale changes, with percentages marked

(a) Frozen response accuracy on Burgers

Backbone	Rel. response error (%) ↓	Response cosine ↑	Amplitude ratio
CNO	4.392	0.999	0.994
Transolver	2.065	1.000	0.989

(b) Probe-scale stability of FNO on Cylinder

Response change at $h/2$ (%) ↓	Response change at $2h$ (%) ↓
0.063	0.248

Table 31: Frozen response validation. (a) Response accuracy of CNO and Transolver on Burgers against the numerical solver. (b) Probe-scale stability of FNO on Cylinder at $h/2$ and $2h$.

Table 32: Higher-order local-structure audit across datasets and backbones.

Dataset	Backbone	Cached q_2 error (%) ↓	FP64 straight-path q_2 error (%) ↓	AD q_2 error (%) ↓
Burgers	FNO	434.351	434.012	434.123
Burgers	CNO	6160.445	6160.374	100.000
Burgers	Transolver	341.134	341.585	341.935
KS	FNO	869.401	868.994	869.094
Kolmogorov Flow	FNO	50.645	50.643	50.647

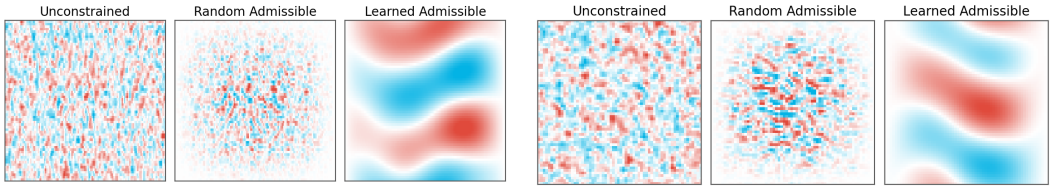


Figure 8: Probe fields on Cylinder (left) and FSI (right). The comparison illustrates the spatial structure induced by probe learning on the same underlying states.

explicitly. The two comparisons assess fidelity to the target dynamics and stability with respect to the probing scale, respectively.

We further examine second-order local information through the centered response

$$q_{2,G,h}(x,q) = \frac{G(x+hq) - 2G(x) + G(x-hq)}{h^2}, \qquad G \in \{F_\theta, \mathcal{S}\}. \tag{20}$$

Its relative discrepancy is the L_2 norm of the difference between the frozen and solver responses, divided by the solver-response norm. Table 32 reports percentage errors for the cached computation, FP64 straight-path recomputation, and an AD-based audit. This comparison tests whether the fidelity of the retained first-order responses extends to second-order information; PIVOT uses only the first-order responses for correction.

For the probe comparison, *Physics Violation* is the task-specific physical residual evaluated at the perturbed state. Together with response fidelity, it assesses both the physical quality of the queried perturbation and the accuracy of the operator's response to it.

Retained responses provide a physical basis for correction. Table 31(a) extends the main-text FNO analysis to CNO and Transolver on Burgers. Both closely reproduce the numerical response in direction and amplitude. Together with the main-text results on Burgers, KS, and Kolmogorov Flow, this supports local response retention across the evaluated equations and architectures. The agreement before projection or filtering is central to PIVOT's motivation: useful information about physical changes remains accessible within the frozen operator, providing candidate correction directions when its absolute predictions require adaptation.

Table 31(b) further shows that FNO responses on Cylinder change little when the nominal probe magnitude is halved or doubled. The extracted signal is therefore stable across the tested scales. This complements solver-based validation by showing that response extraction is not narrowly dependent on one perturbation magnitude.

Table 33: Probe construction on Burgers. Learned admissible probes improve solver-response agreement together with the physical residual of the perturbed states. Response error is reported as a ratio.

Probe Strategy	Response Error ↓	Response Cosine ↑	Physics Violation ↓
Unconstrained	0.5830	0.8979	1.039×10^{-2}
Random Admissible	0.5843	0.8976	1.083×10^{-2}
Learned Admissible	**0.1900**	**0.9997**	$\mathbf{2.796 \times 10^{-4}}$

Higher-order response fidelity. Table 32 shows substantially weaker agreement for the audited higher-order quantity. FP64 straight-path recomputation closely matches the cached errors throughout: higher precision leaves the discrepancy essentially unchanged. The AD audit also yields substantial errors, although its agreement with finite differences is architecture dependent. In particular, CNO exhibits a large discrepancy between the two evaluations: its AD error is 100%, compared with approximately 6160% for finite differences. These evaluations therefore require separate interpretation for this backbone.

The directional responses are substantially more faithful to the solver than the audited higher-order quantity. PIVOT uses this first-order signal to construct correction directions. The audit explains this choice: the frozen operator retains useful information about how outputs change along a probe, while its second-order finite response can differ sharply from the solver's.

Probe learning couples informative responses with physical quality. Table 33 compares unconstrained probes optimized without feasibility restrictions, random probes drawn from the admissible space, and probes learned within that space. Learned admissible probing achieves closer solver-response agreement together with a substantially smaller physical residual. Thus, its advantage extends beyond directional alignment to the physical quality of the perturbed states used to query the operator.

Random admissible and unconstrained probes have similar response errors and physical residuals in this comparison. Learned admissible probes improve both measures, combining feasibility restrictions with task-driven direction selection. Learning within that construction identifies directions that better satisfy both objectives. These results support the combined probing design: admissibility defines the search space, and learning selects useful directions within it. They also motivate evaluating response fidelity and physical violation jointly, since neither alone characterizes probe quality.

Probe structure on RealPDEBench. The original high-fidelity simulator cannot be queried for arbitrary virtually perturbed states on Cylinder and FSI, so direct solver-based response validation is unavailable for these datasets. Figure 8 provides complementary qualitative evidence: on the same underlying states, learned admissible probes exhibit more coherent spatial organization than the unconstrained and random alternatives. This illustrates how learning shapes the queried perturbations. Their effect on the predicted dynamics is assessed through the quantitative prediction and physics-sensitive evaluations reported elsewhere in the paper.

Implications for physically consistent adaptation. These analyses support a specific mechanism for PIVOT: retained responses supply physically informative candidate directions, and admissible probe learning improves which directions are exposed. Reliable individual responses can interact when combined with spatially varying weights. Checking the assembled correction assesses these interactions and the resulting physical behavior of the predicted field. Together with the physical-consistency results and acceptance ablation, the evidence connects the extraction of retained local knowledge to controlled use of that knowledge in adaptation.

B.3 Cross-Domain Matching under Distribution Shift

We evaluate cross-domain matching under a controlled diffusion shift on Burgers and simulation-to-real transfer with Reynolds extrapolation on Cylinder Flow. These settings test adaptation to altered dynamics and to combined physical and observational shifts. All FNO-based methods share

Table 34: Cross-domain performance on Cylinder. All adapters use FNO as the backbone

Method	Rel-L2 ↓	Terminal Rel-L2 ↓	H1 Rel-L2 ↓	FE ↓	KE ↓
MemNO (Buitrago Ruiz et al., 2025)	0.0991	0.1376	0.6052	157.82	1.107×10^{-3}
FNO (Li et al., 2021)	0.1015	0.1395	0.5919	161.84	3.005×10^{-3}
Bottleneck (Houlsby et al., 2019)	0.0986	0.1362	0.5857	127.79	1.950×10^{-3}
LoRA (Hu et al., 2021)	0.0936	0.1230	0.5982	80.05	9.149×10^{-4}
GraLoRA (Jung et al., 2025)	0.0935	0.1222	0.6024	78.35	9.145×10^{-4}
F-Adapter (Zhang et al., 2025)	0.0914	0.1214	0.5870	71.69	9.425×10^{-4}
POTT (Yang & Ren, 2025)	0.1128	0.1404	0.6905	118.88	1.000×10^{-3}
PIVOT	**0.0899**	**0.1204**	**0.5836**	**64.65**	$\mathbf{7.969 \times 10^{-4}}$

Table 35: Cross-domain performance on Burgers. All adapters use FNO as the backbone

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
MemNO (Buitrago Ruiz et al., 2025)	**0.4549**	**0.7973**	**1.4461**	6.389×10^{-3}	6.977×10^{-6}
FNO (Li et al., 2021)	0.7084	1.5225	1.9254	5.767×10^{-3}	1.216×10^{-4}
Bottleneck (Houlsby et al., 2019)	0.6949	1.2371	1.9053	8.375×10^{-3}	7.882×10^{-5}
LoRA (Hu et al., 2021)	0.6901	1.2368	1.8950	7.417×10^{-3}	1.178×10^{-4}
GraLoRA (Jung et al., 2025)	0.6875	1.2316	1.8923	7.363×10^{-3}	1.089×10^{-4}
F-Adapter (Zhang et al., 2025)	0.6918	2.7392	1.9071	7.696×10^{-3}	1.390×10^{-4}
POTT (Yang & Ren, 2025)	0.7026	2.7072	2.1138	2.535×10^{-2}	6.193×10^{-4}
PIVOT	0.6791	1.2090	1.8666	$\mathbf{5.470 \times 10^{-3}}$	$\mathbf{6.945 \times 10^{-9}}$

the source checkpoint, target observations, normalization, and evaluation protocol. The frozen FNO measures performance before adaptation, while MemNO provides a separate architectural reference.

Burgers: controlled diffusion shift. We use the two-dimensional conservative Burgers configuration from APEBench (Koehler et al., 2024). The target domain reduces diffusion while preserving the convection parameter, periodic domain, spatial resolution, initial-condition distribution, and saved timestep. Source and target data use the same numerical solver configuration, so the comparison isolates the change in diffusion strength.

Source training, source validation, and all target splits use independently generated initial conditions with disjoint generation keys. Adaptation uses short autoregressive windows, whereas validation and testing use progressively longer rollouts. Test rollouts begin with an observed context and then use model predictions as inputs. This tests generalization to unseen initial conditions over horizons longer than those used for adaptation and checkpoint selection.

Cylinder Flow: simulation-to-real extrapolation. We use numerical and experimental velocity fields from the Cylinder dataset in RealPDEBench (Hu et al., 2026). Source training and source-model selection use exclusively low-Reynolds CFD trajectories. Target adaptation and evaluation use experimental trajectories at Reynolds numbers above the source range. Numerical trajectories at target Reynolds settings are excluded from source training, checkpoint selection, and normalization.

The selected target trajectories are assigned to adaptation, validation, and testing in order of increasing Reynolds number. Each complete trajectory and Reynolds setting belongs to a single split, and all test Reynolds values exceed those used for adaptation and validation. Windows are extracted after this assignment. The resulting evaluation combines simulation-to-real transfer with parameter extrapolation, testing whether adaptation transfers to higher Reynolds numbers.

Preprocessing and model selection. Cylinder fields retain the shared velocity components (u, v) and use fixed spatial resampling to align the physical observation region and grid. Temporal sampling follows the released timestamps, with the same physical sampling interval in both domains. Normalization statistics are computed from the corresponding source training split and remain fixed

Table 36: Cross-domain matching protocols. Cylinder Reynolds intervals summarize the selected discrete conditions, with complete trajectories assigned to disjoint splits. Burgers horizons count predicted steps after the input context.

Item	Burgers	Cylinder Flow
Source domain	$\gamma = 1.5$	CFD, $Re \in [1800, 6000]$
Target domain	$\gamma = 0.75$	Experimental (u, v) fields
Model grid	160×160	64×128
Target adaptation	8 trajectories; 50-step horizon	8 trajectories; $Re \in [9000, 9843]$
Target validation	4 trajectories; 100-step horizon	4 trajectories; $Re \in [9937, 10218]$
Target testing	30 trajectories; 200-step horizon	11 trajectories; $Re \in [10312, 11812]$
Training windows	4 input frames; 20 predicted steps	20 input frames; 20 predicted frames
Adaptation sampling	All eligible training windows	8 windows per trajectory
Validation sampling	Full validation rollouts	8 windows per trajectory
Test protocol	Autoregressive rollout	Fixed windows; stride 40

during adaptation and testing. Physical diagnostics are evaluated after restoring physical units, using consistent grid coordinates and a shared valid-fluid mask.

Cylinder adaptation and validation windows are evenly distributed over each assigned trajectory; their complete input–output spans do not overlap. Trajectory assignments and window indices are identical across methods. Checkpoint selection uses target validation prediction error: mean rollout nRMSE for Burgers and relative L_2 error for Cylinder. Training hyperparameters follow Appendix C. Metrics are averaged within each test trajectory and then equally across trajectories. All methods use the same training seed.

Cylinder Flow: joint predictive and physical gains. PIVOT achieves the lowest errors across all reported metrics in Table 34, combining improved field reconstruction with lower physical errors under the simulation-to-real and Reynolds shifts. Its advantage over both the frozen FNO and competing adapters extends to independent experimental trajectories beyond the adaptation Reynolds range. Physical diagnostics further distinguish methods with similar reconstruction errors.

POTT exhibits a contrasting outcome: its physical diagnostics improve over the frozen FNO, while its prediction errors increase. Prediction errors and physical diagnostics thus capture different aspects of forecast quality. PIVOT improves both categories, yielding more accurate fields together with lower physical errors.

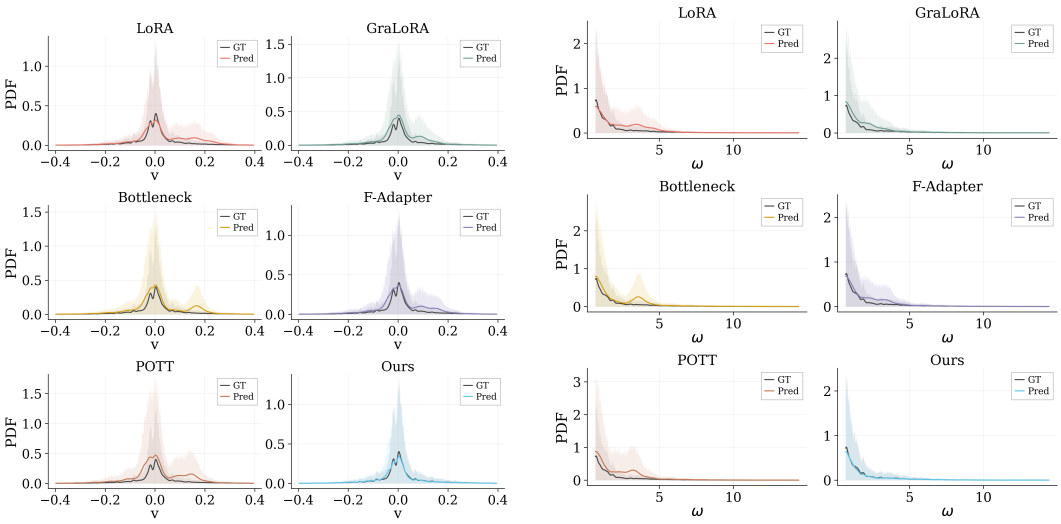
Burgers: adaptation quality over extended rollouts. PIVOT provides the strongest predictive adaptation of the shared FNO in Table 35 and achieves the lowest PDE Defect and Balance Error among all compared methods. MemNO remains the most accurate predictive architecture overall. PIVOT therefore combines the strongest adaptation of the shared source operator with the lowest physical errors.

Mean and terminal errors reveal different rollout behavior: F-Adapter and POTT improve mean prediction error over the frozen FNO while worsening terminal prediction and physical consistency. PIVOT improves average and terminal accuracy together with the physical diagnostics. These gains extend across unseen target trajectories and forecasting horizons longer than those used for adaptation.

Implications. The two settings provide complementary evidence for response-based output correction under controlled dynamical change and combined simulation-to-real and parameter shifts. Across both, PIVOT couples the strongest predictive adaptation of the shared FNO with leading physical diagnostics. Field reconstruction, late-horizon prediction, and physical consistency together reveal how these gains extend across complementary aspects of adaptation quality.

Table 37: Distributional consistency after FNO adaptation to experimental Cylinder Flow.

Method	$\mathbf{JSD}_u$ $(\times 10^{-3})\downarrow$	$\mathbf{JSD}_v$ $(\times 10^{-3})\downarrow$	$\mathbf{JSD}_\omega$ $(\times 10^{-3})\downarrow$
LoRA (Hu et al., 2021)	32.114	<u>11.198</u>	25.035
Bottleneck (Houlsby et al., 2019)	31.274	15.752	29.524
GraLoRA (Jung et al., 2025)	<u>28.634</u>	11.433	<u>23.297</u>
F-Adapter (Zhang et al., 2025)	30.955	18.668	25.553
POTT (Yang & Ren, 2025)	60.805	15.201	25.753
PIVOT	**27.228**	**10.486**	**22.569**



(a) Distribution of velocity component v.

(b) Distribution of vorticity magnitude $|\omega|$.

Figure 9: Velocity and vorticity distributions after FNO adaptation to experimental Cylinder Flow.

B.4 Distributional Consistency under Cross-Domain Matching

We extend the Cylinder Flow analysis to assess whether cross-domain adaptation recovers the marginal statistics of the target flow. Following the JSD protocol defined in the main text, we report separate scores for the velocity components u and v and the vorticity magnitude $|\omega|$ in Table 37. These comparisons complement reconstruction errors and aggregate physical diagnostics by revealing how adaptation changes the relative frequency of observable values.

Evaluation details. All methods use identical held-out target windows, with source normalization reversed before computing the observables. Statistics are pooled across evaluated windows, forecast frames, and spatial locations before computing JSD. The reported vorticity magnitude uses coordinates normalized independently along the two grid axes, with second-order centered differences in the interior and second-order one-sided differences at the boundaries. This coordinate convention is shared by all methods and the reference for the JSD diagnostic; the preceding physical-error metrics use physical grid coordinates.

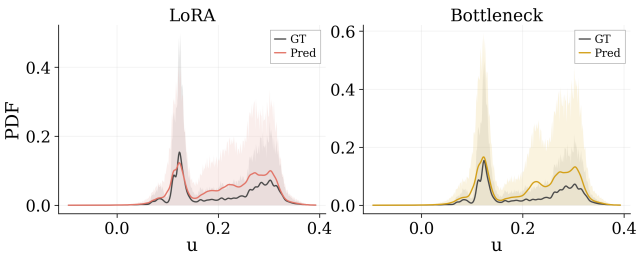


Figure 10: Distribution of velocity component u after FNO adaptation to experimental Cylinder Flow.

Table 38: PIVOT ablation on Cylinder Flow with CNO.

Method	Rel.-L_2 ↓	Terminal Rel.-L_2 ↓	H^1 Rel.-L_2 ↓	FE ↓	KE ↓
w/o Spatial Response Combination	0.0649	0.0887	0.4807	42.10	5.76×10^{-4}
w/o Error-Reconstruction Alignment	0.0645	0.0883	0.4773	38.80	5.48×10^{-4}
w/o Coefficient Distillation	0.0646	0.0884	0.4782	39.60	5.55×10^{-4}
w/o Physical Acceptance	0.0642	0.0880	0.4758	44.60	5.88×10^{-4}
Full	**0.0631**	**0.0875**	**0.4752**	**34.75**	$\mathbf{4.21 \times 10^{-4}}$

Table 39: PIVOT ablation on Cylinder Flow with FNO.

Method	Rel.-L_2 ↓	Terminal Rel.-L_2 ↓	H^1 Rel.-L_2 ↓	FE ↓	KE ↓
w/o Spatial Response Combination	0.0669	0.0900	0.5048	44.70	5.44×10^{-4}
w/o Error-Reconstruction Alignment	0.0665	0.0895	0.5024	41.70	5.28×10^{-4}
w/o Coefficient Distillation	0.0666	0.0896	0.5030	42.40	5.32×10^{-4}
w/o Physical Acceptance	0.0663	0.0893	0.5017	46.10	5.49×10^{-4}
Full	**0.0652**	**0.0882**	**0.5013**	**38.41**	$\mathbf{5.14 \times 10^{-4}}$

Consistent agreement across complementary observables. Table 37 shows that PIVOT achieves the lowest reported JSD for all three observables, outperforming both general-purpose and physics-oriented adapters. Its distributional advantage spans both velocity components and a derivative-based statistic. Together, these results show that improved prediction accompanies closer agreement with the target flow's velocity and rotational-magnitude statistics.

Figure 9 identifies the distributional differences behind these rankings. For v, PIVOT closely follows the central reference distribution and reduces the excess density at positive velocities visible in competing adapters. For $|\omega|$, it more closely follows the displayed reference decay and reduces the intermediate-range excess exhibited by several baselines. These comparisons show improved agreement in the frequency of velocity values and local rotational magnitudes, connecting the aggregate scores to specific distributional discrepancies.

Figure 10 provides a complementary example: LoRA and Bottleneck overestimate density in the higher-velocity region of u. This comparison locates the discrepancy within the observable range: these adapters assign too much probability mass to higher velocities. The distribution plots thus make the aggregate JSD scores interpretable.

Implications for target-domain correction. Distributional fidelity measures how often observable values occur, while trajectory errors assess where and when the predicted fields agree with observations. Vorticity magnitude summarizes the unsigned curl in the normalized grid coordinates. PIVOT's lower distributional discrepancies accompany its leading predictive and physical results on Cylinder Flow. Together, these results support response-based correction under the evaluated domain shift, with improvements extending from field reconstruction to the marginal statistics of velocity and local rotational variation.

B.5 Expanded Component Analysis across Systems and Backbones

We extend the component analysis in Figure 7 across Burgers and Cylinder Flow using FNO, CNO, and Transolver backbones. All variants share the same pretrained backbone, optimization protocol, and evaluation setting, with exactly one component removed at a time. Complete results are reported in Tables 41–43 and Tables 38–40. Across systems and architectures, the ablations distinguish two roles: constructing an effective correction and controlling its physical effect.

Table 40: PIVOT ablation on Cylinder Flow with Transolver.

Method	Rel.-L_2 ↓	Terminal Rel.-L_2 ↓	H^1 Rel.-L_2 ↓	FE ↓	KE ↓
w/o Spatial Response Combination	0.0705	0.0938	0.4630	146.80	2.60×10^{-4}
w/o Error-Reconstruction Alignment	0.0701	0.0934	0.4613	136.50	2.53×10^{-4}
w/o Coefficient Distillation	0.0702	0.0935	0.4618	139.60	2.55×10^{-4}
w/o Physical Acceptance	0.0699	0.0932	0.4603	149.30	2.63×10^{-4}
Full	**0.0698**	**0.0931**	**0.4600**	**125.76**	$\mathbf{2.47 \times 10^{-4}}$

Table 41: PIVOT ablation on Burgers with CNO.

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H^1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
w/o Spatial Response Combination	0.4130	1.3310	1.2980	0.011800	2.80×10^{-7}
w/o Error-Reconstruction Alignment	0.3470	1.2420	1.0880	0.007600	1.30×10^{-7}
w/o Coefficient Distillation	0.3650	1.2640	1.1460	0.008300	1.70×10^{-7}
w/o Physical Acceptance	0.3100	1.1830	0.9450	0.014600	3.60×10^{-5}
Full	**0.2990**	**1.1647**	**0.9150**	**0.005587**	$\mathbf{5.12 \times 10^{-9}}$

Physical acceptance reveals a clear prediction–physics separation. Removing physical acceptance most clearly separates field accuracy from physical fidelity. Across all three backbones, predictive errors remain close to those of full PIVOT, while the physical metrics deteriorate substantially more. On Burgers, Balance Error rises from the 10^{-9} scale under the full model to the 10^{-5} scale after removing physical acceptance for all three backbones, with much smaller relative changes in rollout error. PDE Defect also increases for every backbone. Cylinder Flow exhibits a parallel pattern: Rel.-L_2, terminal error, and H^1 error change only slightly, while FE and KE consistently worsen.

These results distinguish physical acceptance from the components that improve prediction accuracy. Responses that individually reduce prediction error can combine into a spatial update that increases physical residuals. Physical acceptance evaluates this assembled correction directly, after the learned spatial weights have combined the response contributions. The ablation shows that comparable field accuracy can coexist with markedly different physical fidelity.

Spatial response combination is critical to effective correction. Removing spatial response combination causes the strongest predictive degradation among the structural ablations, particularly on Burgers with CNO and Transolver, while physical diagnostics deteriorate concurrently across both systems. Spatially varying response contributions allow the correction to follow errors that differ across the field. Retained responses provide local directions of change, and spatial combination gives the readout the flexibility to adjust their contributions at different locations in the predicted field. Restricting this mechanism weakens both reconstruction quality and the physical fidelity of the resulting rollout.

Alignment determines response relevance; distillation supports their use. Error-reconstruction alignment and coefficient distillation affect a different stage of the correction process. Removing alignment weakens both predictive and physical performance in every setting, supporting its role in training the probed responses to explain the baseline error. Removing coefficient distillation produces larger errors than removing alignment across all six settings and all reported metrics. Distillation transfers the locally fitted response-combination structure into the learned readout. Together, these results support two linked requirements: the retained responses should explain the error, and the inference-time readout should learn how to combine them.

The component roles persist across operator families. The same qualitative behavior recurs for spectral FNOs, convolutional CNOs, and Transformer-based Transolvers. Spatial response combi-

Table 42: PIVOT ablation on Burgers with FNO.

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H^1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
w/o Spatial Response Combination	0.4865	1.3498	1.3045	0.004112	2.40×10^{-7}
w/o Error-Reconstruction Alignment	0.4832	1.3415	1.2935	0.004038	1.10×10^{-7}
w/o Coefficient Distillation	0.4845	1.3445	1.2978	0.004067	1.50×10^{-7}
w/o Physical Acceptance	0.4807	1.3368	1.2846	0.004103	1.20×10^{-5}
Full	**0.4798**	**1.3352**	**1.2818**	**0.003938**	$\mathbf{5.51 \times 10^{-9}}$

Table 43: PIVOT ablation on Burgers with Transolver.

Method	Mean nRMSE ↓	Terminal nRMSE ↓	H^1 nRMSE ↓	PDE Defect ↓	Balance Error ↓
w/o Spatial Response Combination	0.6610	2.7080	3.0930	0.07950	3.10×10^{-7}
w/o Error-Reconstruction Alignment	0.5900	2.5840	2.9070	0.07080	1.50×10^{-7}
w/o Coefficient Distillation	0.6120	2.6250	2.9740	0.07320	1.90×10^{-7}
w/o Physical Acceptance	0.5400	2.5080	2.7360	0.08480	2.40×10^{-5}
Full	**0.5266**	**2.4747**	**2.6891**	**0.06319**	$\mathbf{5.83 \times 10^{-9}}$

nation remains closely associated with reconstruction quality; alignment and distillation consistently improve the use of retained responses; and physical acceptance shows the strongest selective effect on physical fidelity. This pattern appears under different physical diagnostics: PDE Defect and Balance Error on Burgers, and FE and KE on Cylinder Flow. The same division of roles therefore appears across all six tested configurations, covering three operator families and two sets of physical diagnostics.

Implication for physically consistent adaptation. The ablations connect each component to a distinct stage of correction. Learned probing supplies informative local responses, error-reconstruction alignment connects them to the current prediction error, coefficient distillation teaches the readout how to combine them, spatial response combination distributes that information across the field, and physical acceptance assesses the physical effect of the assembled update.

The clearest separation appears when physical acceptance is removed: prediction quality can remain nearly unchanged while physical fidelity degrades sharply. The physical gains of PIVOT therefore cannot be explained by improved field reconstruction alone. They also depend on checking how the combined local responses change the physical residuals of the final prediction. This result supports the central design principle of PIVOT: adapting a frozen neural operator requires both extracting useful retained responses and governing their physical effect.

C Hyperparameters and Computational Cost

Training protocol. Within each controlled comparison, adaptation methods share the same data splits, pretrained backbone checkpoints, prediction protocols, and validation-based model selection. PIVOT optimizes its adaptation modules using AdamW with an effective batch size of 8, an initial learning rate of 10^{-3}, weight decay of 10^{-4}, and gradient-norm clipping at 1.0. The learning rate follows a cosine schedule over 100 epochs to 10^{-5} (1% of its initial value) and remains at this minimum thereafter, independently of the 1,000-epoch training ceiling. Validation is performed every epoch, and training stops after 20 consecutive checks without an improvement exceeding 10^{-6}. The checkpoint with the lowest validation prediction error is selected for evaluation, with the initial model also eligible. APEBench uses the geometric mean of channelwise trajectory-averaged normalized RMSE over the first 100 forecast steps (or the available horizon if shorter); RealPDEBench uses relative L^2 error over the forecast window. Test data are used only after model selection.

PIVOT configuration. PIVOT learns two admissible probe directions per input and queries the frozen operator at paired positive and negative perturbations. Including the unperturbed prediction, each prediction step uses five operator evaluations; autoregressive forecasting repeats this procedure at every rollout step. Only first-order local responses are retained. The probe controller has two hidden layers of width 64, and the local response readout uses a hidden width of 64 for APEBench and 32 for RealPDEBench to produce spatially varying combination weights. Normalized probe amplitudes lie in $[5 \times 10^{-4}, 10^{-1}]$, and the correction-budget ratio A/S is 0.5. The physics-consistency and amplitude-regularization weights are 0.05 and 0.01, respectively. The error-reconstruction alignment, coefficient distillation, and diversity weights are 0.1, 0.25, and 0.02; the coefficient-distillation multiplier follows a cosine decay to 0.2 of its initial value over the scheduler horizon. Coefficient fields undergo one fixed averaging step. One residual-based refinement step uses damping 0.1, a relative radius of 0.25, and a finite-difference step of 0.01. At inference, physical acceptance tests the full correction and up to four successive halvings, using relative and absolute tolerances of 10^{-5} and 10^{-8}; otherwise, it returns the frozen prediction. The pretrained backbone parameters remain frozen throughout adaptation.

Compute environment. The reported experiments run on a single NVIDIA RTX PRO 6000 Blackwell Server Edition GPU with 96 GB of memory. Computational measurements use Python 3.11, PyTorch 2.11, CUDA 12.8, and BF16 automatic mixed precision. Virtual response queries are evaluated in FP32. We report wall-clock training time to early stopping and peak GPU memory monitored externally during training.

Computational cost. On Cylinder Flow with CNO, the reported PIVOT measurements are 36.3 hours of training and 1,406 MB of peak GPU memory, compared with 9.0 hours and 12,060 MB for DoRA Liu et al. (2024). PIVOT therefore reduces measured peak memory by 88.3%, using only 11.7% of DoRA's peak memory at approximately $4.0\times$ the training time in this setting. Each prediction step requires five operator evaluations, compared with one for DoRA Liu et al. (2024). Training additionally incurs gradient computation and possible activation recomputation, while validation-based early stopping makes total wall-clock time depend on both per-step cost and the number of optimization steps. These measurements characterize the reported trade-off between training time and peak memory.

Code availability. The supplementary material includes the model architecture and core implementation of PIVOT, enabling direct inspection of the proposed method. The complete training and evaluation pipeline, experiment-specific configurations, and remaining implementation details will be released upon acceptance to support full reproducibility.