MOMENTUM-ACCELERATED STRUCTURED PRECONDITIONING FOR PHYSICS-INFORMED NEURAL NETWORKS

Indra Priyadarsini, Onur Boyar & Sina Klampt
IBM Research - Tokyo
{indra.ipd, onur.boyar, sina.klampt}@ibm.com

ABSTRACT

Physics-Informed Neural Networks (PINNs) solve PDEs by training neural surrogates that minimize governing-equation residuals alongside boundary and initial conditions. Despite their flexibility, PINNs are notoriously difficult to optimize due to stiffness, loss imbalance, and gradient noise induced by stochastic collocation. In this work, we study a structured preconditioned optimizer inspired by SOAP and propose a lightweight extension based on a momentum accelerated method - MoQ, which uses gradient extrapolation to approximate Nesterov-style acceleration. We conduct preliminary studies comparing Adam, SOAP, and the proposed SOAP+MoQ on three PDEs problems, namely, viscous Burgers equation, 2D Poisson, and steady Navier–Stokes (Kovasznay flow). Across multiple random seeds, proposed SOAP+MoQ consistently improves accuracy over both Adam and SOAP, with potential strong gains on Poisson and Navier–Stokes. We further probe the role of gradient noise via batch reuse and observe that MoQ gains can increase in reduced-noise regimes. These preliminary results suggest MoQ is an effective stabilization mechanism for structured preconditioners in PINN optimization.

1 INTRODUCTION

Physics-Informed Neural Networks (PINNs) approximate solutions to partial differential equations by minimizing the residuals of governing equations and boundary/initial constraints, typically computed via automatic differentiation. Despite their broad applicability, PINNs often suffer from slow convergence and limited solution accuracy, which restricts their reliability as practical forward PDE solvers. Prior work has addressed these challenges through architectural innovations Wang et al. (2021); Sitzmann et al. (2020); Fathony et al. (2020); Moseley et al. (2021); Kang et al. (2023); Wang et al. (2024), improved activations Abbasi and Østebø Andersen (2024); Jagtap et al. (2020) and coordinate embeddings Sahli Costabal et al. (2024); Zeng et al. (2024); Huang and Alkhalifah (2024), refined collocation sampling strategiesWu et al. (2023); Daw et al. (2022); Nabian et al. (2021), alternative objective formulations, and enhanced training paradigms such as sequential training and transfer learning Cao and Zhang (2023); Urbán et al. (2025); Song et al. (2024); Jnini et al. (2024); Müller and Zeinhofer (2023).

A central bottleneck, however, remains the difficulty of optimizing PINN objectives, which are often ill-conditioned and exhibit severe gradient pathologies Wang et al. (2021; 2022). While Adam is widely adopted due to its robustness, it can converge slowly in stiff PINN regimes. Recent structured preconditioners such as Shampoo Gupta et al. (2018) and SOAP Vyas et al. (2024) improve conditioning by approximating curvature through Kronecker-factored second-moment statistics. Motivated by these advances, we investigate SOAP-style structured preconditioning for PINN training. Empirically, we find that SOAP alone can be sensitive in PINN settings, sometimes yielding high losses and worse accuracy than Adam.

To improve stability and accelerate convergence, we introduce a lightweight extension based on Momentum Quasi-Newton (MoQ) acceleration Mahboubi et al. (2021); Indrapriyadarsini et al. (2021), yielding the proposed SOAP+MoQ optimizer. We benchmark Adam, SOAP, and SOAP+MoQ across three representative PDEs and report consistent multi-seed accuracy improvements. Finally, we study the role of stochastic gradient noise through collocation batch reuse experiments, highlighting conditions under which MoQ provides the largest gains.

2 PROPOSED METHOD: SOAP+MOQ FOR PINN OPTIMIZATION

PINN objective. Let $u_\theta(\mathbf{x})$ denote a neural surrogate for the PDE solution $u(\mathbf{x})$ on a spatiotemporal domain Ω with boundary $\partial\Omega$. A standard PINN trains θ by minimizing a weighted sum of residual and constraint losses:

$$\mathcal{L}(\theta) = w_f\, \mathbb{E}_{\mathbf{x}_f\sim\Omega}\left[\|f_\theta(\mathbf{x}_f)\|_2^2\right] + w_{bc}\, \mathbb{E}_{\mathbf{x}_{bc}\sim\partial\Omega}\left[\|u_\theta(\mathbf{x}_{bc}) - g_{bc}(\mathbf{x}_{bc})\|_2^2\right] + w_{ic}\, \mathbb{E}_{\mathbf{x}_{ic}}\left[\|u_\theta(\mathbf{x}_{ic}) - g_{ic}(\mathbf{x}_{ic})\|_2^2\right], \tag{1}$$

where f_θ is the PDE residual computed via automatic differentiation, and g_{bc}, g_{ic} denote the prescribed boundary/initial targets.

2.1 SOAP-STYLE STRUCTURED PRECONDITIONING

Let $g^{(k)} = \nabla_\theta \mathcal{L}(\theta^{(k)})$ denote the stochastic gradient at iteration k. For each parameter tensor θ_l, denote its gradient tensor by $G_l^{(k)}$. SOAP constructs a structured second-moment preconditioner by reshaping $G_l^{(k)}$ into a matrix in $\mathbb{R}^{m\times n}$ (e.g., flattening all but the first axis), and maintaining Kronecker-factored exponential moving averages:

$$A_l \leftarrow \beta_2 A_l + (1-\beta_2)\, G_l G_l^\top, \qquad B_l \leftarrow \beta_2 B_l + (1-\beta_2)\, G_l^\top G_l, \tag{2}$$

where $\beta_2 \in (0,1)$ and we omit the iteration superscript for readability. The preconditioned gradient for tensor l is

$$P_l(G_l) = (A_l + \lambda I)^{-1/2}\, G_l\, (B_l + \lambda I)^{-1/2}, \tag{3}$$

with damping $\lambda > 0$ for numerical stability. The inverse square roots are computed via eigendecomposition and updated periodically. For very small tensors or tensors exceeding a maximum dimension threshold, we fall back to a diagonal second-moment preconditioner.

2.2 MOMENTUM ACCELERATION AND SOAP+MOQ UPDATE

Inspired by the Momentum Quasi-Newton (MoQ) gradient extrapolation in Mahboubi et al. (2021), we introduce SOAP-MoQ. SOAP alone can be sensitive in PINN training due to stochastic collocation and non-stationary residual gradients. We therefore introduce MoQ-based gradient extrapolation before applying SOAP preconditioning. SOAP+MoQ combines Kronecker-factored second-moment preconditioning with momentum-based gradient extrapolation. For each tensor l, MoQ forms

$$G_{l,\mathrm{MoQ}}^{(k)} = (1+\mu)\, G_l^{(k)} - \mu\, G_l^{(k-1)}, \tag{4}$$

where $\mu \in [0,1)$ is the momentum coefficient and $G_l^{(k-1)}$ is the previous gradient tensor.

SOAP+MoQ applies SOAP preconditioning to $G_{l,\mathrm{MoQ}}^{(k)}$ and updates parameters with momentum:

$$V_l^{(k+1)} = \mu V_l^{(k)} - \eta\, P_l\!\left(G_{l,\mathrm{MoQ}}^{(k)}\right), \qquad \theta_l^{(k+1)} = \theta_l^{(k)} + V_l^{(k+1)}, \tag{5}$$

where η is the learning rate and V_l is the velocity buffer for tensor l.

In our implementation, we update the SOAP factors (A_l, B_l) using the same tensor gradient that is passed to the preconditioner (i.e., $G_{l,\mathrm{MoQ}}^{(k)}$), which empirically improves stability under stochastic collocation.

SOAP+MoQ combines structured second-moment preconditioning with a momentum based gradient extrapolation mechanism. This yields a lightweight optimizer that remains first-order in computational cost, while empirically improving robustness and convergence in stiff PINN optimization regimes.

3 RESULTS AND DISCUSSION

To evaluate the effectiveness of SOAP-style structured preconditioning and the proposed SOAP+MoQ optimizer, we consider three widely used PINN benchmark PDEs spanning time-dependent nonlinear dynamics, elliptic boundary value problems, and coupled nonlinear systems.

1. **1D viscous Burgers equation.** We consider the 1D viscous Burgers equation

$$u_t + uu_x = \nu u_{xx}, \tag{6}$$

defined on $(x,t) \in [-1,1] \times [0,1]$ with initial condition $u(x,0) = -\sin(\pi x)$ and Dirichlet boundary conditions $u(-1,t) = u(1,t) = 0$. Burgers is a standard nonlinear benchmark that exhibits sharp gradients and shock-like behavior for small viscosity, making optimization challenging for PINNs.

2. **2D Poisson equation.** We evaluate a steady elliptic Poisson problem on a unit square domain $\Omega = [0,1]^2$ with prescribed Dirichlet boundary conditions.

$$-\Delta u(x,y) = f(x,y), \quad (x,y) \in \Omega, \tag{7}$$

Poisson problems are widely used to test smooth solution approximation and constraint satisfaction, while still exhibiting stiffness due to second-order derivatives in the residual.

3. **2D steady incompressible Navier–Stokes (Kovasznay flow).** We consider the steady incompressible Navier–Stokes equations on $\Omega = [0,1]^2$ with Reynolds number $Re = 40$:

$$uu_x + vu_y + p_x - \nu\Delta u = 0, \tag{8}$$

$$uv_x + vv_y + p_y - \nu\Delta v = 0, \tag{9}$$

$$u_x + v_y = 0, \tag{10}$$

where $\nu = 1/Re$. We use the Kovasznay exact solution to impose boundary conditions and compute evaluation error over the interior.

Evaluation metric. We report relative L_2 error on a uniform evaluation grid:

$$\mathrm{Rel}\, L_2 = \frac{\|u_\theta - u_{\mathrm{ref}}\|_2}{\|u_{\mathrm{ref}}\|_2}. \tag{11}$$

For Navier–Stokes, $\mathrm{Rel}\, L_2$ is computed jointly over (u,v,p) by concatenating the fields. All experiments use an MLP with tanh activations, Xavier initialization, width 128, depth 4. The MoQ momentum term is set to 0.85.

3.1 Performance Comparison

Table 1 summarizes the relative L_2 error across five random seeds for Adam, SOAP, and SOAP+MoQ on all three PDE benchmarks. Across all problems, SOAP+MoQ achieves the lowest mean relative error, demonstrating that MoQ acceleration consistently improves the effectiveness of SOAP-style preconditioning in PINN optimization.

PDE Problem	Adam	SOAP	SOAP+MoQ
Burger	0.3536 ± 0.0072	0.3064 ± 0.0284	$\mathbf{0.1477 \pm 0.0665}$
Navier-Stokes	0.0137 ± 0.0049	0.0240 ± 0.0053	$\mathbf{0.0082 \pm 0.0027}$
Poisson	0.0374 ± 0.0105	0.0506 ± 0.0178	$\mathbf{0.0278 \pm 0.0239}$

Table 1: Summary of results Relative L2 Error (Mean $\pm$ standard deviation of over 5 seeds.

Across all three PDEs, SOAP+MoQ achieves the best mean relative error, suggesting that MoQ provides a robust improvement over SOAP-style preconditioning when applied to PINNs. A key empirical finding is that SOAP alone can be brittle in PINN settings, sometimes yielding significantly higher losses and worse accuracy than Adam, whereas SOAP+MoQ consistently mitigates this issue.

3.2 Effect of Gradient Noise via Collocation Batch Reuse

A major difficulty in PINN optimization is that gradients are typically estimated using randomly resampled collocation points, which introduces stochastic gradient noise and non-stationarity into the objective. To isolate the impact of gradient noise, we perform a controlled sweep where collocation batches are reused for K consecutive optimization steps before resampling. Larger reuse factors K reduce gradient variance and yield more consistent descent directions.

Table 2 reports the MoQ gain over SOAP, computed as $\mathrm{err}_{\mathrm{SOAP}}/\mathrm{err}_{\mathrm{SOAP+MoQ}}$, for reuse factors $K \in \{1, 5, 10, 20\}$. Across all PDEs, SOAP+MoQ consistently outperforms SOAP for all reuse factors, and the advantage is often amplified under larger reuse values.

Reuse K	Burger	Poisson	Navier–Stokes
1	1.37×	3.07×	2.87×
5	2.08×	4.42×	2.60×
10	1.08×	1.58×	6.46×
20	4.45×	5.81×	1.68×

Table 2: MoQ gain over SOAP measured as $\mathrm{err}_{\mathrm{SOAP}}/\mathrm{err}_{\mathrm{SOAP+MoQ}}$ for different reuse factors K.

These results suggest that the benefits of MoQ are closely tied to gradient noise and non-stationarity. SOAP preconditioning relies on estimating structured second-moment statistics, which can be unreliable when gradients fluctuate significantly across steps. MoQ mitigates this effect by using an extrapolated gradient direction that incorporates information from previous updates, yielding improved stability and more consistent progress. This is consistent with the observation that SOAP alone can occasionally diverge or converge to poor solutions in PINN regimes, while SOAP+MoQ more reliably attains low error solutions.

4 Conclusion

We proposed SOAP+MoQ, a simple optimizer extension that combines SOAP-style structured preconditioning with a momentum accelerated gradient extrapolation step. Across three PDE benchmarks (Burgers, Poisson, and Navier–Stokes) and multiple random seeds, SOAP+MoQ consistently improves solution accuracy over both Adam and SOAP, with particularly strong gains on Poisson and Navier–Stokes where SOAP alone can be unstable. Additional collocation batch reuse experiments suggest that MoQ benefits from reduced gradient noise, supporting the interpretation that MoQ acts as an effective stabilizer for structured preconditioned PINN optimization. Overall, SOAP+MoQ provides a lightweight and practical modification for improving PINN training robustness, and motivates future work on variance-aware and curvature-informed optimizers for physics-informed learning.

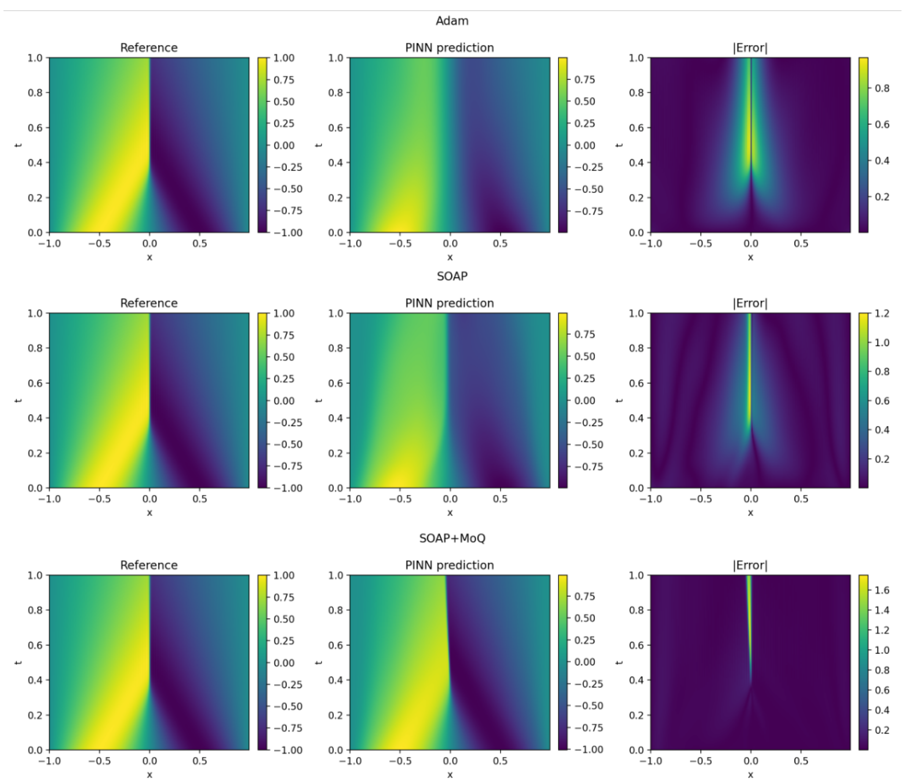

Figure 1: Burgers equation solution heatmaps. From top to bottom: Adam, SOAP, and SOAP+MoQ. Each row shows the reference solution, PINN prediction, and absolute error. SOAP+MoQ produces the sharpest shock structure and lowest error concentration near the discontinuity region.

REFERENCES

Jassem Abbasi and Pål Østebø Andersen. Physical activation functions (pafs): An approach for more efficient induction of physics into physics-informed neural networks (pinns). *Neurocomputing*, 608:128352, 2024. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2024.128352. URL https://www.sciencedirect.com/science/article/pii/S0925231224011238.

Wenbo Cao and Weiwei Zhang. Tsonn: Time-stepping-oriented neural network for solving partial differential equations. *arXiv preprint arXiv:2310.16491*, 2023.

Arka Daw, Jie Bu, Sifan Wang, Paris Perdikaris, and Anuj Karpatne. Rethinking the importance of sampling in physics-informed neural networks. *arXiv preprint arXiv:2207.02338*, 2022.

Rizal Fathony, Anit Kumar Sahu, Devin Willmott, and J Zico Kolter. Multiplicative filter networks. In *International conference on learning representations*, 2020.

Vineet Gupta, Tomer Koren, and Yoram Singer. Shampoo: Preconditioned stochastic tensor optimization. In *International Conference on Machine Learning*, pages 1842–1850. PMLR, 2018.

Xinquan Huang and Tariq Alkhalifah. Efficient physics-informed neural networks using hash encoding. *Journal of Computational Physics*, page 112760, 2024.

S Indrapriyadarsini, Shahrzad Mahboubi, Hiroshi Ninomiya, Takeshi Kamio, and Hideki Asai. A modified limited memory nesterov's accelerated quasi-newton. *arXiv preprint arXiv:2112.01327*, 2021.

Ameya D Jagtap, Kenji Kawaguchi, and George Em Karniadakis. Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. *Journal of Computational Physics*, 404:109136, 2020.

Anas Jnini, Flavio Vella, and Marius Zeinhofer. Gauss-newton natural gradient descent for physics-informed computational fluid dynamics. *arXiv preprint arXiv:2402.10680*, 2024.

Namgyu Kang, Byeonghyeon Lee, Youngjoon Hong, Seok-Bae Yun, and Eunbyung Park. Pixel: Physics-informed cell representations for fast and accurate pde solvers. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 8186–8194, 2023.

Shahrzad Mahboubi, Hiroshi Ninomiya, Hideki Asai, et al. Momentum acceleration of quasi-newton based optimization technique for neural network training. *Nonlinear Theory and Its Applications, IEICE*, 12(3):554–574, 2021.

Ben Moseley, Andrew Markham, and Tarje Nissen-Meyer. Finite basis physics-informed neural networks (fbpinns): a scalable domain decomposition approach for solving differential equations. *arXiv preprint arXiv:2107.07871*, 2021.

Johannes Müller and Marius Zeinhofer. Achieving high accuracy with pinns via energy natural gradient descent. In *International Conference on Machine Learning*, pages 25471–25485. PMLR, 2023.

Mohammad Amin Nabian, Rini Jasmine Gladstone, and Hadi Meidani. Efficient training of physics-informed neural networks via importance sampling. *Computer-Aided Civil and Infrastructure Engineering*, 2021.

Francisco Sahli Costabal, Simone Pezzuto, and Paris Perdikaris. δ-pinns: Physics-informed neural networks on complex geometries. *Engineering Applications of Artificial Intelligence*, 127:107324, 2024.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

Yongcun Song, Xiaoming Yuan, and Hangrui Yue. The admm-pinns algorithmic framework for nonsmooth pde-constrained optimization: A deep learning approach. *SIAM Journal on Scientific Computing*, 46(6):C659–C687, 2024.

Jorge F. Urbán, Petros Stefanou, and José A. Pons. Unveiling the optimization process of physics informed neural networks: How accurate and competitive can pinns be? *Journal of Computational Physics*, 523:113656, 2025.

Nikhil Vyas, Depen Morwani, Rosie Zhao, Mujin Kwun, Itai Shapira, David Brandfonbrener, Lucas Janson, and Sham Kakade. Soap: Improving and stabilizing shampoo using adam. *arXiv preprint arXiv:2409.11321*, 2024.

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

Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why pinns fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2022.

Sifan Wang, Bowen Li, Yuhan Chen, and Paris Perdikaris. Piratenets: Physics-informed deep learning with residual adaptive networks. *Journal of Machine Learning Research*, 25(402):1–51, 2024.

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

Chengxi Zeng, Tilo Burghardt, and Alberto M. Gambaruto. Rbf-pinn: Non-fourier positional embedding in physics-informed neural networks. *arXiv preprint arXiv:2402.08367*, 2024.