Learning Physical Models That Can Respect Conservation Laws

Derek Hansen[1]*, **Danielle C. Maddix**[2], **Shima Alizadeh**[2], **Gaurav Gupta**[2], **and Michael W. Mahoney**[3]
[1]University of Michigan, [2]AWS AI Labs, [3]Amazon Supply Chain Optimization Technologies
[1]dereklh@umich.edu, [2,3]{dmmaddix,alizshim,gauravaz,zmahmich}@amazon.com

Abstract

Recent work in scientific machine learning (SciML) has focused on incorporating partial differential equation (PDE) information into the learning process. Most of this work has focused on relatively "easy" PDE operators (e.g., elliptic and parabolic), with less emphasis on relatively "hard" PDE operators (e.g., hyperbolic). Within numerical PDEs, the latter need to maintain a type of volume element or conservation constraint for a desired physical quantity, which is known to be challenging. Delivering on the promise of SciML requires seamlessly incorporating both types of problems into the learning process. To address this issue, we propose ProbConserv, a framework for incorporating constraints into a black-box probabilistic deep-learning architecture. To do so, ProbConserv combines the integral form of a conservation law with a Bayesian update. We demonstrate the effectiveness of ProbConserv via a case study of the Generalized Porous Medium Equation (GPME), a parameterized family of equations that includes both easier and harder PDEs. On the challenging Stefan variant of the GPME, we show that ProbConserv seamlessly enforces physical conservation constraints, maintains probabilistic uncertainty quantification (UQ), and deals well with shocks and heteroscedasticity. In addition, it achieves superior predictive performance on downstream tasks.

1 Introduction

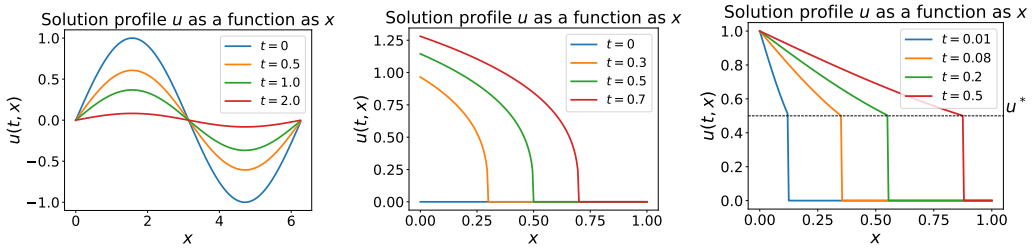
Partial differential equations (PDEs) that model physical phenomena ranging from heat transfer to flow dynamics are classified into three groups: parabolic, elliptic, and hyperbolic. While solutions to simple parabolic and elliptic problems are smooth and diffusive (Figure 1(a)), solutions to hyperbolic problems can be sharp (Figure 1(b)) or even discontinuous (Figure 1(c)) (Evans, 2010).

The Generalized Porous Medium Equation (GPME) is a family of equations that exhibit both parabolic (easy) and hyperbolic (hard) behaviors. Originally developed to model the density of fluids through porous media, the GPME has been used in several applications, such as underground flow transport, non-linear heat transfer, and water desalination (Vázquez, 2007). For a particular non-linear pressure function $k(u)$, the PDE of the density function u is given as:

$$u_t - \nabla \cdot (k(u)\nabla u) = 0. \tag{1}$$

Figure 1 illustrates the differences between the solution profiles of the GPME with different $k(u)$. Figure 1(a) shows the solution for $k(u) = 1$ with a sine initial condition. Figure 1(b) shows a solution to the Porous Medium Equation (PME) subclass of the GPME where $k(u) = u^m$. Here, we show $m = 3$, where a constant zero initial condition develops a sharp gradient that does not dissipate over time. Lastly, Figure 1(c) illustrates an example of the hard Stefan problem sub-class of the GPME where $k(u)$ is a nonlinear discontinuous step-function of the unknown u defined by the unknown value $u^\star = u(t, x^\star(t)) = 0.5$ at the discontinuity $x^\star(t)$, where the solution evolves as a rightward moving shock over time.

*Work completed during internship at AWS AI Labs.



(a) easy: heat equation ($k = 1$) (b) medium: PME ($k(u) = u^3$) (c) hard: Stefan (discont. $k(u)$)

Figure 1: Paradigm of easy to hard PDEs as illustrated from the GPME family of equations: (a) easy parabolic smooth heat equation solutions with constant parameter $k = 1$; (b) degenerate parabolic Porous Medium Equation (PME) solutions with nonlinear monomial coefficient $k(u) = u^3$ with parameter $m = 3$; (c) hard hyperbolic-like (degenerate parabolic) sharp Stefan solutions with nonlinear step-function parameter defined by constant u^* at various times.

Writing the flux as $F(u) = -k(u)\nabla u$, Equation 1 can be written as a conservation law $u_t + \nabla \cdot F(u) = 0$, which implies conservation of mass. Respecting conservation laws is critical to obtain a solution to challenging hyperbolic problems. For example, finite volume methods have been developed for hyperbolic conservation laws (LeVeque, 1990; 2007). In hard degenerate parabolic problems that can have hyperbolic-like behavior such as Stefan, it is critical to accurately estimate the shock speed, position, and be volume conservative (Maddix et al., 2018).

Recently, there have been several works in scientific machine learning (SciML) focused on studying partial differential equation (PDEs) using deep learning. These works include Physics-Informed Neural Networks (PINNs) (Raissi et al., 2019), which uses a neural network to approximate the PDE solution and Neural Operators (NOs) (Li et al., 2021; Gupta et al., 2021) that aim to learn the underlying continuous function map. These works have mainly focused on solving smooth and nicely-behaved PDEs, e.g., elliptic and parabolic operators, for which finite difference and finite element methods are appropriate with less emphasis on relatively "hard" PDE operators e.g., hyperbolic operators, for which more sophisticated finite volume methods are necessary. In contrast, these deep learning solutions have no guarantee that the physical property of volume conservation is satisfied, and can violate the governing conservation law.

To address the lack of guarantee of satisfying conservation by deep learning models, we propose ProbConserv, a framework for incorporating physical constraints into a black-box deep-learning architecture. ProbConserv consists of two steps. In the first step, we choose the Attentive Neural Process (ANP) (Kim et al., 2019) as our black-box, data-driven model, and we refer to this instantiation of our framework as ProbConserv-ANP. In particular, given sparse context points, the ANP outputs a conditional probability distribution on the remaining target points. In the second step, we treat this distribution as a prio and formulate conservation of mass as a Bayesian update, leading to a conservative posterior predictive distribution. We illustrate our method on the Stefan variant of the GPME by enforcing global conservation of mass, and show reduction of mean squared error (MSE), improved adherence to the global mass, and better estimation of the shock position compared to the other baselines.

2 ProbConserv

ProbConserv consists of two steps. First, a data-driven ML model predicts the mean and covariance of the unknown function at a discrete set of points. Second, this mean and covariance are updated to respect the conservation law via a Bayesian update. These steps are summarized in Figure 2, where the Attentive Neural Process (ANP) is used as the Step 1 model.

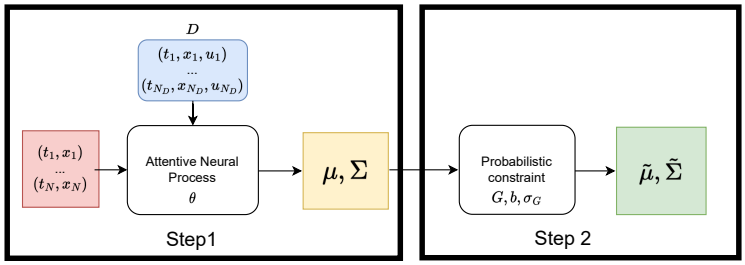


Figure 2: Schematic of each component of ProbConserv-ANP.

Step 1: Unconstrained Probability Distribution In Step 1 of ProbConserv, we use a supervised black-box model to infer the mean and covariance Σ of the unknown function u from observed data D. For example, D can include values of the function u observed at a small set of points. Over a set of N input points $(t_1, x_1), \ldots, (t_N, x_N)$, the probability distribution of $u := [u(t_1, x_1), \ldots u(t_N, x_N)] \in \mathbb{R}^N$ conditioned on data D has mean $\mu := \mathbb{E}(u|D)$ and covariance $\Sigma := \mathrm{Cov}(u|D)$ given by the black-box model f_θ, i.e.,

$$\mu, \Sigma = f_\theta\left((t_1, x_1), \ldots, (t_N, x_N); D\right). \tag{2}$$

This framework is general, and there are possible choices for the model in Equation 2. Gaussian Processes (Rasmussen & Williams, 2006) are a natural choice, assuming that one has chosen an appropriate mean and kernel function for the specific problem. The ANP model (Kim et al., 2019), which uses a transformer architecture to encode the mean and covariance, is another choice. A third option is to perform repeated runs, e.g., with different initial seeds, of non-probabilistic black-box NN models to compute empirical estimates of mean and variance parameters.

Step 2: Enforcing Conservation Constraint In Step 2 of ProbConserv, we incorporate knowledge of the total mass in the spatial domain Ω at time t, written $b(t)$. This is related to the function u via a linear operator $\mathcal{G}u$:

$$\mathcal{G}u(t, x) = \int_\Omega u(t, x) d\Omega = \int_\Omega h(x) d\Omega - \int_0^t \int_\Gamma F(u) \cdot n d\Gamma dt = b(t). \tag{3}$$

Here, $h(x) := u(0, x)$ is the initial condition, $F(u)$ is the flux, and Γ is the boundary of the spatial domain Ω. Since $\mathcal{G}$ is a linear operator, Equation 3 can be discretized to match the output of Step 1:

$$b = Gu + \sigma_G \epsilon, \tag{4}$$

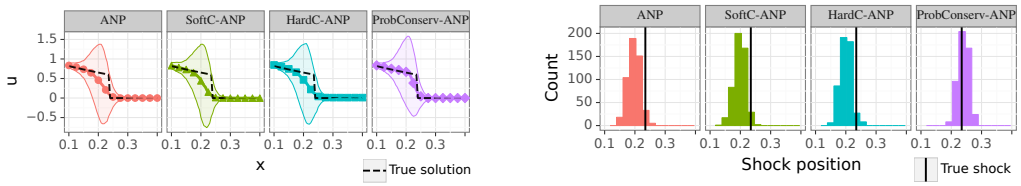
where G denotes a matrix approximating the linear operator $\mathcal{G}$, b denotes a vector of observed constraint values, and ϵ denotes a noise term, where each component has unit variance. The parameter $\sigma_G \geq 0$ controls how much the conservation constraint can be violated, with $\sigma_G = 0$ enforcing exact adherence. Step 2 outputs the following updated mean $\tilde{\mu}$ and covariance $\tilde{\Sigma}$ that respect conservation, given as:

$$\tilde{\mu} = \mu - \Sigma G^T (\sigma_G^2 I + G \Sigma G^T)^{-1} (G\mu - b), \tag{5a}$$

$$\tilde{\Sigma} = \Sigma - \Sigma G^T (\sigma_G^2 I + G \Sigma G^T)^{-1} G \Sigma, \tag{5b}$$

where μ and Σ denote the mean and covariance matrix, respectively, from Step 1 (Equation 2).

The update rule given in Equation 5 can be justified from two complementary perspectives. From a Bayesian probabilistic perspective, Equation 5 is the posterior mean and covariance of the predictive distribution of u after incorporating the information given by the conservation constraint via Equation 4. From an optimization perspective, Equation 5 is the solution to a least-squares problem that places a binding inequality constraint on the conserved quantity Gu (i.e., $\|Gu - b\|_2 \leq c$ for some $c \in (0, \|G\mu - b\|_2)$).



(a) Solution profile.

(b) Posterior of the shock position.

Figure 3: Solution profile and posterior shock position for the Stefan problem.

Table 1: Mean and standard error for CE $\times 10^{-2}$ (should be zero), LL (higher is better), and MSE $\times 10^{-3}$ (lower is better) for the Stefan problem at time $t = 0.05$.

	CE	LL	MSE
ANP	-1.30 (0.01)	3.53 (0.00)	5.38 (0.01)
SoftC-ANP	-1.72 (0.04)	**3.57** (0.01)	6.81 (0.15)
HardC-ANP	**0** (0.00)	2.33 (0.06)	5.18 (0.02)
ProbConserv-ANP	**0** (0.00)	**3.56** (0.00)	**1.89** (0.01)

3 Experiments: Stefan Problem

We evaluate the performance of ProbConserv-ANP on the Stefan variant of the Generalized Porous Medium Equation (GPME) in Equation 1 with nonlinear discontinuous coefficient $k(u) = \mathbf{1}_{u \geq u^\star}$ for indicator function $\mathbf{1}$ and shown in Figure 1(c). We compare to the unconstrained ANP, and two additional physics-based models. SoftC-ANP includes the differential form as a soft constraint as done in PINNs (Raissi et al., 2019). HardC-ANP projects the solution from the end of Step 1 to the nearest globally-conservative solution in L_2, inspired by Négiar et al. (2023).

We train the ANP from Kim et al. (2019) on $N = 10,000$ independent function samples with $u^\star \in [0.55, 0.7]$. Each function sample has $N_D = 100$ context points and $N_{\text{train}} = 100$ target points. After training, ProbConserv-ANP is utilized to set G to conserve global mass. The global mass is taken from the true solution profile. We set the noise to $\sigma_G^2 = 0$; i.e. exact conservation is enforced. At test-time, we evaluate our ProbConserv-ANP and the baselines over $n_{\text{test}} = 50$ independent context sets, each with $N_D = 100$ points, drawn from the solution at parameter value $u^\star = 0.6$, which is inside the training set. We reconstruct the solution over a 201×201 grid of equally-spaced target points in time and space. Compared to the baseline ANP, ProbConserv-ANP leads to an increase in log-likelihood and a $\mathbf{3}\times$ decrease in MSE.

In addition to MSE and LL, we also explore the impact of enforcing global conservation on the downstream task of shock point estimation. Define the shock point at time t as the first spatial point (left-to-right) where the function equals zero: $x^\star(t) \equiv \inf_x\{u(t,x) = 0\}$. To practically compare the estimated shock point on discrete estimates, we replace the infimum with a minimum over the discrete grid points. We can directly estimate the posterior distribution of $x^\star(t)$ by drawing samples from the posterior distributions from ProbConserv-ANPand the baselines.

We show histograms of the posteriors of the shock position for $t = 0.05$ in Figure 3(b). Even though the underlying parameter is inside the training range ($u^\star = 0.6$), the ANP and other baselines underestimate the true shock position. After enforcing global conservation of mass, the ProbConserv-ANP posterior is centered around the true mass value.

4 Conclusion

ProbConserv-ANP harnesses the flexibility of black-box neural processes while maintaining physical viability. This is accomplished by framing the problem probabilistically, offering a coherent way to combine information coming from both data and knowledge of underlying conservation laws.

References

L.C. Evans. *Partial Differential Equations*, volume 19 of *Graduate studies in mathematics*. American Mathematical Society, 2nd edition, 2010.

Gaurav Gupta, Xiongye Xiao, and Paul Bogdan. Multiwavelet-based Operator Learning for Differential Equations. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. Attentive Neural Processes. *arXiv preprint arXiv:1901.05761*, 2019.

Randall J. LeVeque. *Numerical Methods for Conservation Laws*. Lectures in mathematics ETH Zürich. Birkhäuser Verlag, 1990.

Randall J. LeVeque. *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*. SIAM, 2007.

Zongyi Li, Nikola 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.

Danielle C. Maddix, Luiz Sampaio, and Margot Gerritsen. Numerical artifacts in the discontinuous Generalized Porous Medium Equation: How to avoid spurious temporal oscillations. *Journal of Computational Physics*, 368:277–298, 2018.

Geoffrey Négiar, Michael W. Mahoney, and Aditi S. Krishnapriyan. Learning differentiable solvers for systems with hard constraints. In *International Conference on Learning Representations*, 2023.

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

C.E. Rasmussen and C.K. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

J.L. Vázquez. *The Porous Medium Equation: Mathematical Theory*. The Clarendon Press, Oxford University Press, Oxford, 2007.