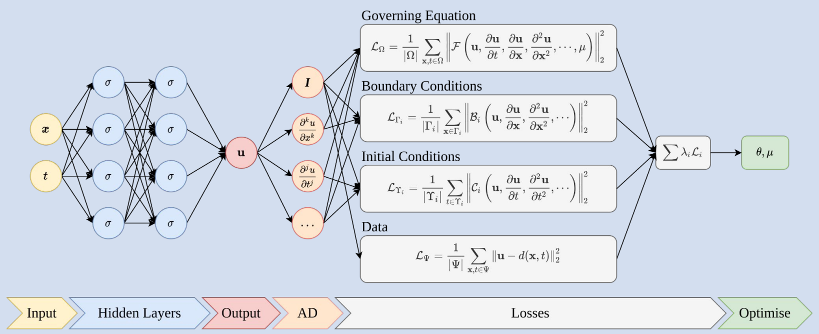
Multi-Objective Loss Balancing for Physics-Informed Deep Learning

R. Bischof, Dr. M. A. Kraus
E-mail: rabischof@ethz.ch; kraus@ibk.bau.ethz.ch

1 Introduction

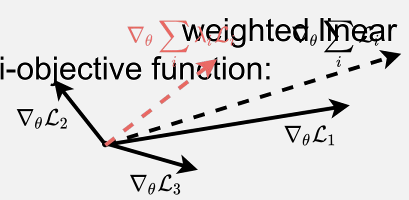
Physics-Informed Neural Networks (PINN) leverage physical laws by including partial differential equations (PDE) together with a respective set of boundary and initial conditions (BC / IC) as penalty terms into their loss function [1]. The physical prior was shown to accelerate convergence and increase networks' ability to generalise. With emerging frameworks like SciML or NVIDIA SimNet considerably reducing the technical hurdles for deploying PINNs, these networks are set to attract lots of attention in coming years and find applications in a variety of engineering problems.



2. Problem

As the terms in the loss function stem from physical laws and their IC resp. BC, their numerical magnitude may vary significantly due physical units. Consequently, the signal strengths of backpropagated gradients potentially differs from term to term, leading to pathologies that were shown to impede proper training and cause imbalance solutions [2] with unsatisfactory training efficiency and accuracy.

As a remedy, one may interpret the physics-induced loss function as the weighted linear scalarization of a multi-objective function:



$$\mathcal{L}(\theta) = \sum_{i=1}^{k} \lambda_i \mathcal{L}_i(\theta), \quad \lambda_i \in \mathbb{R}_{>0}$$

3 Adaptive Loss Balancing

In order to avoid finding optimal λ_i manually during training of PINNs, employing adaptive loss balancing techniques is suggested. We implemented and evaluated three state-of-the-art methods (Learning Rate Annealing, GradNorm and SoftAdapt) and propose a novel, self-adaptive heuristic called **ReLoBRaLo** (Relative Loss Balancing with Random Lookbacks). The performance of these techniques was assessed on three benchmark problems for PINNs: Burgers' equation, Kirchhoff's plate bending equation and Helmholtz's equation.

4 ReLoBRaLo

We propose a novel method and implementation **ReLoBRaLo** [3] for balancing multiple terms in the scalarised multi-objective loss function for training of PINNs:

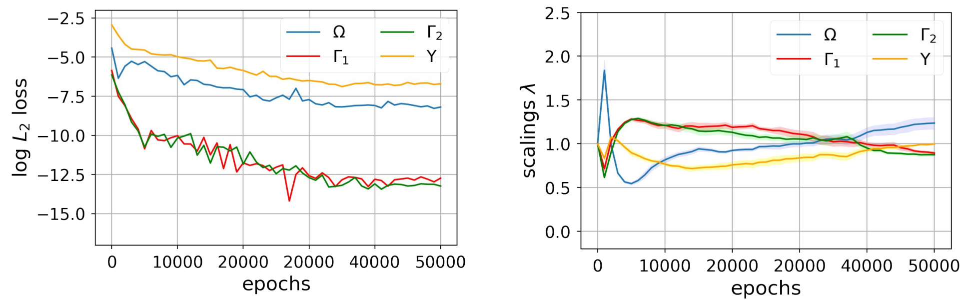
$$\hat{\lambda}_i^{(t;t')} = m \cdot \frac{exp\left(\frac{\mathcal{L}_i^{(t)}}{\mathcal{T}\mathcal{L}_i^{(t')}}\right)}{\sum_{j=1}^{m} exp\left(\frac{\mathcal{L}_j^{(t)}}{\mathcal{T}\mathcal{L}_j^{(t')}}\right)}, \quad i \in \{0, \ldots, m\}$$

$$\lambda_i^{(t)} = \alpha(\rho\lambda_i^{(t-1)} + (1-\rho)\hat{\lambda}_i^{(t;0)}) + (1-\alpha)\hat{\lambda}_i^{(t;t-1)}$$

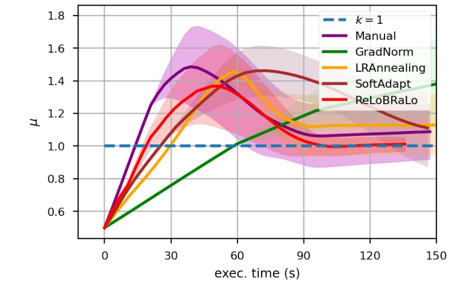
- SoftAdapt's balancing method is employed, using the rate of change between consecutive training steps and normalising them through a softmax function.
- Similarly to Learning Rate Annealing, the scalings are updated using an exponential decay of rate α in order to utilise loss statistics from more than just one training step in the past.
- In addition, a random lookback is introduced through a Bernoulli random variable ρ (called *saudade* factor) which defines, whether to use the previous steps' loss statistics, or to look all the way back to the start of training.

Our studies empirically suggest to chose the *saudade* Bernoulli random variable ρ such, that the expectation E[ρ] is close to 1.

5 Results and Discussion



(right) PINNs L2 Convergence, (left) ReLoBRaLo Loss Scaling for Burgers' equation



PINNs Parameter Estimation Convergence (Helmholtz)

Median execution times (in s) per 1'000 optimisation steps

[s] per 1'000 it.	Burgers	Kirchhoff	Helmholtz
Manual	3.700	17.30	4.800
GradNorm	14.30	128.6	10.40
LR annealing	7.200	139.7	6.700
SoftAdapt	4.100	20.20	5.000
ReLoBRaLo	4.300	22.50	5.200

Our computations empirically prove, that ReLoBRaLo is able to consistently outperform the baseline of existing scaling methods (GradNorm, Learning Rate Annealing, SoftAdapt or manual scaling) in terms of accuracy, while also being up to six times more computationally efficient (training epochs or wall-clock time) [3]. Using Bayesian Optimisation instead of GridSearch turned out to be an effective tool to reduce the laborious work of finding optimal hyperparameters for the PINNs. Finally, we showed that the adaptively chosen scalings λ_i can be inspected to learn about the PINNs training process and identify weak points. This allows to take informed decisions in order to improve the framework.

References

1. Raissi, Maziar, Paris Perdikaris, and George 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 (2019): 686-707.
2. Wang, Sifan, Yujun Teng, and Paris Perdikaris. "Understanding and mitigating gradient flow pathologies in physics-informed neural networks." SIAM Journal on Scientific Computing 43.5 (2021): A3055-A3081.
3. Bischof, Rafael, and Michael Kraus. "Multi-Objective Loss Balancing for Physics-Informed Deep Learning." arXiv preprint arXiv:2110.09813 (2021).

GITHUB **PAPER** **RESEARCHGATE**

NVIDIA GTC 2021, San Jose, CA, US, March 2021

ETH zürich