Proceeding Series of the Brazilian Society of Computational and Applied Mathematics
Preprint

Adaptive PINN-SP1: NTK-Based Penalties for Noisy Inverse Problems in Radiative Transfer

Renato A. S. Klein[1], Pedro H. A. Konzen[2]
PPGMAp/IME/UFRGS, Porto Alegre, RS

Abstract. We consider the inverse problem of estimating the absorption coefficient in a two-dimensional participating medium governed by the SP_1 approximation of the particle transport equation. While Physics-Informed Neural Networks (PINNs) are promising for such tasks, estimating a piecewise constant absorption from discrete boundary detectors introduces severe loss imbalances that destabilize training. To overcome this, our methodology relies on adaptive penalties to dynamically weight the loss components, ensuring a stable and robust optimization process. We evaluate this adaptive framework using radiative flux data from discrete boundary detectors with noise. Our results demonstrate that incorporating adaptive penalties successfully mitigates training instabilities, allowing the model to accurately estimate both the location and magnitude of the absorption coefficient even in noisy scenarios.

Keywords. Inverse Problems, Radiative Transfer Equation, Physics-Informed Neural Networks, Adaptive Penalties

1 Introduction

The modeling of physical problems involves solving differential equations whose parameters characterize the properties of the system. Directly measuring these parameters can sometimes be impractical, or even the ultimate goal, as in the case of tumor imaging in optical tomography [1]. In this context, the development of methods capable of estimating parameters through observed data becomes essential. Several works have tackled this problem with different methodologies, such as [10], which establishes the formulation for inverse problems with Bayesian inference in finite-dimensional spaces, and [8], which used the analytical discrete ordinates method to solve the adjoint particle transport equation to address the inverse source reconstruction problem. Previous work was done using PINN-$SP1$ for absorption parameter estimation [4]. In inverse problems, Physics-Informed Neural Networks (PINNs) often struggle to accurately estimate unknown physical parameters due to a phenomenon known as loss imbalance. Because the PINN loss function typically aggregates multiple competing terms, such as data mismatch from discrete detectors, boundary conditions, and the underlying partial differential equations, the gradients originating from these distinct components can vary significantly in magnitude [11]. The present work expands upon previous work [4] by incorporating adaptive penalties from Neural Tangent Kernel (NTK) theory [11] to stabilize loss imbalance during training, leading to better results and faster convergence. This approach also mitigates the need for empirically tuning hyperparameters, which has a high computational cost due to multiple runs needed to find suitable values. In this work, we deal with the radiative transfer equation, which describes how electromagnetic radiation is absorbed, emitted, and scattered through a medium. This equation appears in various applications such as

[1] renato.klein@ufrgs.br
[2] pedro.konzen@ufrgs.br

molten glass cooling, heat transfer in gas turbines [6], and optical tomography applications [1]. We consider the SP_1 approximation of this equation [2, 5], aiming to solve the inverse problem of estimating the absorption coefficient in a two-dimensional participating medium using a PINN method [9]. The SP_1 approximation is given by

$$-\varepsilon^2 \nabla \left(\frac{1}{3(\sigma + \kappa)} \nabla \Phi(x, y) \right) + \kappa \Phi = \kappa (4\pi B), \quad (x, y) \in \mathcal{D} = [0, 1]^2 \tag{1}$$

with Robin boundary conditions

$$\Phi(x, y) + \left(\frac{1 + 3r_2}{1 - 2r_1} \frac{2\varepsilon}{3(\sigma + \kappa)} \right) \mathbf{n} \cdot \nabla \Phi = 4\pi B_b, \quad (x, y) \in \partial \mathcal{D} \tag{2}$$

where the absorption coefficient is given by

$$\kappa(x, y) = \begin{cases} a, & (x, y) \in \mathcal{R} \subset \mathcal{D}, \\ b, & (x, y) \in \mathcal{D} \backslash \mathcal{R}, \end{cases} \tag{3}$$

with $\mathcal{R}$ being a small region within the domain $\mathcal{D}$ and $a > b$. Φ is the total incident radiation flux, ε describes the distance to an optically thick medium, B is the internal source, B_b is the boundary source, $\mathbf{n}$ is a normal vector pointing towards the exterior of the boundaries, r_1 and r_2 are coefficients associated with the reflectivity of the boundaries, and σ is the scattering coefficient.

2 Methodology

In this section, we introduce the methodology used to estimate the absorption parameter of equation (3) in the presence of noisy detectors. The estimation is performed via the PINN method [9], which simultaneously estimates the total radiation flux and the absorption parameter through the residuals of equations (1) and (2), given data from detectors at the boundaries of domain $\mathcal{D}$. To deal with the PINNs' loss-imbalance issue, we employed adaptive penalties computed as the trace of NTK matrices.

2.1 PINN for Absorption Parameter Estimation

We introduce two multi-layer perceptrons (MLPs) [7] to estimate the total incident radiation and absorption coefficient

$$\tilde{\kappa} = \mathcal{N}_\kappa(\theta_\kappa) \approx \kappa(x, y), \tag{4}$$

$$\tilde{\Phi} = \mathcal{N}_\Phi(\theta_\Phi) \approx \Phi(x, y), \tag{5}$$

where $\mathcal{N}_\kappa$ and $\mathcal{N}_\Phi$ are MLPs that estimate the absorption parameter and the total incident radiation, θ_κ and θ_Φ are the weights and biases of the MLPs, and (x, y) are random samples with uniform distribution generated at each epoch within the domain $\mathcal{D}$. The training of the MLPs is performed by minimizing the mean squared error (MSE) of the residuals of equations (1) and (2), as well as the MSE of the total incident radiation at the detectors.

$$\mathcal{R}_{in}(\theta_\Phi, \theta_\kappa) = -\frac{\varepsilon^2}{3(\sigma + \tilde{\kappa})} \nabla^2 \tilde{\Phi} + \tilde{\kappa} \tilde{\Phi} - \kappa 4\pi B, \tag{6}$$

$$\mathcal{R}_{bc}(\theta_\Phi, \theta_\kappa) = \tilde{\Phi} + \left(\frac{1 + 3r_2}{1 - 2r_1} \frac{2\varepsilon}{3(\sigma + \tilde{\kappa})} \right) n \cdot \nabla \tilde{\Phi} - 4\pi B_b, \tag{7}$$

$$\mathcal{R}_d(\theta_\Phi) = \tilde{\Phi} - \mathbf{d}, \tag{8}$$

where

$$\mathbf{d} = [\Phi(0, y_i), \Phi(1, y_i), \Phi(x_i, 0), \Phi(x_i, 1)], \quad (x_i, y_i) \in \partial[0,1]^2. \tag{9}$$

are the values of the total incident radiation fluxes at the detectors present on the boundaries at positions (x_i, y_i). Derivatives are computed using automatic differentiation. The total loss is given by

$$\begin{aligned}\mathcal{L}(\theta_\Phi, \theta_\kappa) =& \frac{\lambda_{in}}{n_{in}} \sum_{(x_i,y_i)\in(0,1)^2} |\mathcal{R}_{in}(\theta_\Phi, \theta_\kappa)|^2 + \\ & \frac{\lambda_{bc}}{n_{bc}} \sum_{(x_i,y_i)\in\partial[0,1]^2} |\mathcal{R}_{bc}(\theta_\Phi, \theta_\kappa)|^2 + \frac{\lambda_d}{n_d} \sum_{(x_i,y_i)\in\partial[0,1]^2} |\mathcal{R}_d(\theta_\Phi)|^2.\end{aligned} \tag{10}$$

where n_{in} is the number of internal samples, n_{bc} is the number of boundary samples, and n_d is the number of detectors. Furthermore, λ_{in}, λ_{bc}, and λ_d are penalties. The MLPs are trained for n_e epochs. The optimizer used is Adam [3].

2.2 NTK Adaptive Penalties

Inverse problems, due to the multi-objective nature of the loss function, tend to have discrepancies in the convergence rates associated with the PDE residual, boundary conditions, and data. In this case, it is common for the PDE residual to dominate training, leading to spurious local minima. To mitigate this problem, we employ adaptive penalties based on the Neural Tangent Kernel (NTK) theory proposed by [11], where we use the trace of the NTK matrices to balance the influence of the PDE and boundary-condition terms. Every $p_{penalties}$ epochs, we compute

$$\lambda_{bc} = \frac{Tr(\boldsymbol{K}(n))}{Tr(\boldsymbol{K}_{uu}(n))} \tag{11}$$

$$\lambda_{in} = \frac{Tr(\boldsymbol{K}(n))}{Tr(\boldsymbol{K}_{rr}(n))} \tag{12}$$

where

$$\boldsymbol{K}(t) = \begin{bmatrix} \boldsymbol{K}_{uu}(t) & \boldsymbol{K}_r(t) \\ \boldsymbol{K}_{ru}(t) & \boldsymbol{K}_{rr}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{J}_u(t) \\ \boldsymbol{J}_r(t) \end{bmatrix} \left[\boldsymbol{J}_u^T(t), \boldsymbol{J}_r^T(t)\right] = \boldsymbol{J}(t)\boldsymbol{J}^T(t). \tag{13}$$

and J_u is the Jacobian of the MLP prediction at the boundaries computed with respect to the parameters θ_Φ, J_r is the Jacobian of the residual $\mathcal{R}_{in}$ in equation (6) computed with respect to the parameters θ_Φ, and t is the t-th iteration. The computation of the Jacobians, J_r and J_u, is done in smaller random batches, η_{in} and η_{bc}, respectively, of the n_{in} and n_{bc} points in equation (10) to reduce computational cost.

2.3 Data Generation and Setup

Synthetic data were generated by solving the forward problem (1)-(2) using a central and second-order finite difference scheme on a uniform mesh. Data reliability was ensured through an *a posteriori* mesh convergence analysis. To mimic sparse data acquisition, the continuous field Φ was sampled at discrete boundary detectors. We determined the minimum number of homogeneously distributed detectors required for accurate absorption parameter estimation by evaluating the Mean Squared Error (MSE) [4]. Model robustness was further assessed by introducing Gaussian white noise to the exact detector data $\mathbf{d}$, modeled as $\mathbf{d}^\delta = \mathbf{d} + \delta\sigma(\mathbf{d})\boldsymbol{\xi}$, where δ is the noise level, $\sigma(\mathbf{d})$ the standard deviation, and $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

2.4 Network Architecture and Training Strategy

Two multilayer perceptrons (MLPs) were designed to estimate the total incident radiation and the absorption parameter, with hyperparameters optimized via an exploratory study. During training, we employed an adaptive learning-rate strategy that reduces the current rate l_r by a decay factor γ if the loss $\mathcal{L}$ fails to decrease significantly over p epochs. A significant decrease is defined as $\mathcal{L}_{current} < \mathcal{L}_{best}(1-\tau)$, where τ is a given threshold. To prevent convergence to spurious local minima, early stopping is triggered whenever l_r falls below a specified tolerance.

3 Results

Let us consider the case study of equations (1) and (2), where we take $\varepsilon = 0.1$, $r_1 = r_2 = 0$ (non-reflective boundaries), $B = 0$, $B_b = 1/\pi$, and $\sigma + \kappa = 1$, with

$$\kappa(x,y) = \begin{cases} 0.9, & (x,y) \in \mathcal{R} \subset D, \\ 0.1, & (x,y) \in D \backslash \mathcal{R}, \end{cases} \tag{14}$$

where $\mathcal{D} = [0,1]^2$ and $\mathcal{R} = [0.4, 0.6]^2$. Throughout all simulations, we fixed the hyperparameters for training the MLPs as shown in Table 1. We took $p = 2000$ and $\gamma = 0.5$ as the decay factor; we also used $\tau = 10^{-4}$, and the tolerance for early stopping was $l_{r_t} < 5 \times 10^{-6}$. In Table 1, *tol* is the minimum value of the loss for the method to stop. Also, notice that the data penalty, λ_d, is fixed.

n_{in}	n_{bc}	n_e	tol	l_{r_0}	λ_d
2500	50	2×10^5	5×10^{-5}	10^{-3}	1.0

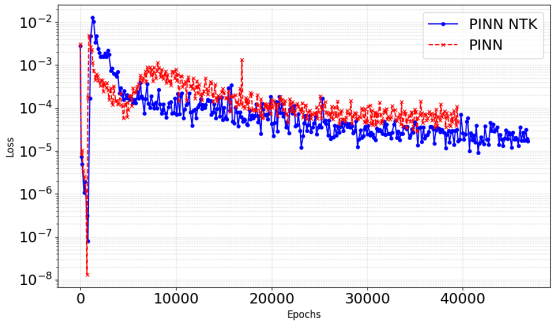
Table 1: Fixed hyperparameters for MLP training.

In the exploratory study, by fixing the size of one MLP and varying the other, we found that $2-5\times 64-1$ for both MLPs, $\mathcal{N}_\kappa$ and $\mathcal{N}_\Phi$, gave good qualitative estimates. For detector selection, we ran three experiments, with 10, 50, and 100 detectors per boundary. For $n_d = 10$ detectors, the method was not able to reconstruct the shape of region $\mathcal{R}$ in equation (14), while for $n_d = 50$ and $n_d = 100$ the method yielded good qualitative estimates. In this case, we chose $n_d = 50$ as the minimum number of detectors. Also, $N = 320$ points were used in the mesh for data generation in the forward problem using finite differences [4]. The adaptive penalties, λ_{in} and λ_{bc}, were computed every $p_{penalties} = 3000$ epochs with batches $\eta_{in} = 128$ and $\eta_{bc} = 16$.

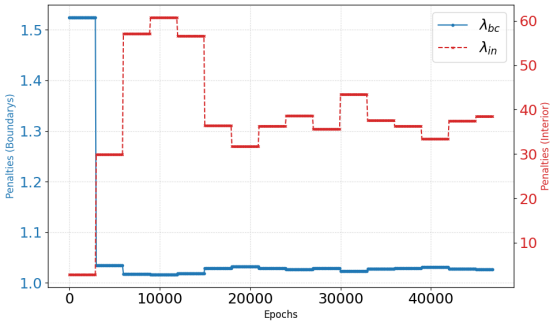
In Figure 1, (1a) shows a comparison of the interior-point loss, i.e., the loss with respect to the residual $\mathcal{R}_{in}$ in equation (6), between a case with constant penalties ($\lambda_{bc} = 0.01$ and $\lambda_{in} = 1.0$, chosen in an exploratory study [4]) and a case using NTK adaptive penalties for λ_{bc} and λ_{in}. We can observe that the loss with adaptive penalties decreases faster and remains more stable during the initial epochs compared to constant penalties. Furthermore, (1b) illustrates the values taken by λ_{bc} and λ_{in} throughout the training process for the adaptive case.

To assess the computational cost of PINN with NTK compared to PINN with constant penalties, we performed five runs of each method with 10000 epochs and computed the average computational time. PINN with NTK yielded an average of 195.69 seconds, while PINN with constant penalties yielded an average of 189.14 seconds over five runs and 10000 epochs, meaning that PINN with NTK is about 10% slower for the same number of epochs. This higher computational time for the same number of epochs is due to the computation of the NTK matrix in equation (13) every 3000 epochs for a batch of $\eta_{in} = 126$ and $\eta_{bc} = 16$ samples. Although PINN with NTK takes more time for the same number of epochs, the loss converges faster than PINN with constant penalties.

As seen in Figure (2a), PINN with constant penalties achieves in 40000 epochs approximately the same loss that PINN with NTK achieves in about 30000 epochs, and PINN with NTK keeps decreasing the loss while PINN with constant penalties reaches its early-stopping condition.

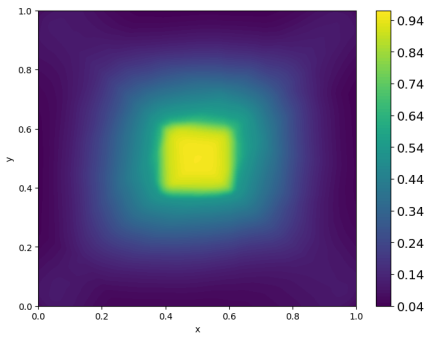


(a) PDE loss for PINN (with $\lambda_{bc} = 0.01$ and $\lambda_{in} = 1.0$) and PINN using NTK penalties. Source: author.

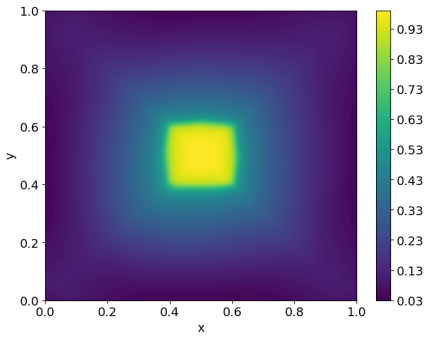


(b) NTK adaptive penalties for λ_{bc} and λ_{in}. Source: author.

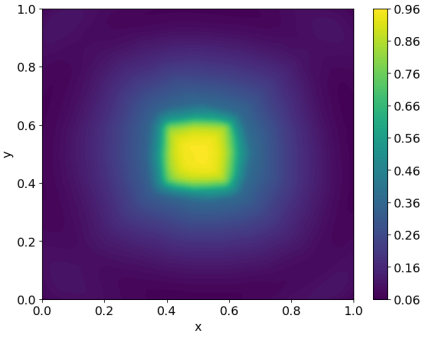
Figure 1: Comparison between PINN and PINN with NTK penalties in (1a), and NTK adaptive penalties along the epochs in (1b), with all results for 0% noise. Source: author.



(a) $\delta = 0.01$. Source: author.



(b) $\delta = 0.05$. Source: author.



(c) $\delta = 0.10$. Source: author.

Figure 2: Estimation of the absorption coefficient for different noise levels. (2a) 1% noise, (2b) 5% noise and (2c) 10% noise. Source: author.

The estimation of the absorption coefficient with 1% noise in Figure (2a), 5% noise in Figure (2b), and 10% noise in Figure (2c) provided a good estimate of the shape, location, and magnitude of region $\mathcal{R}$ (from equation (3)), although it tends to estimate the absorption value in region $\mathcal{R}$ higher than the expected value. This shows that the method is stable for noise levels up to 10%.

4 Final Considerations

In this work, we employed a methodology based on Adaptive PINNs to solve the inverse problem of estimating the absorption coefficient in a two-dimensional medium modeled by the SP_1 approximation of the radiative transfer equation. We used two MLPs to approximate the total incident radiation and the absorption coefficient. Our results show that an architecture of $2 - 5 \times 64 - 1$ and 50 detectors provides balanced accuracy with low computational cost for estimating the absorption coefficient in the presence of noise levels up to 10%, leveraging NTK adaptive penalties, which show a consistently lower loss when compared to the PINN method with constant penalties. This work is under development; we aim to extend our results to more physically accurate data by solving the forward problem using the full transport equation via the method of characteristics and evaluating the capability of the SP_1 approximation to estimate the absorption coefficient from these high-fidelity detector data.

References

[1] G. S. Abdoulaev. "Three-dimensional optical tomography with the equation of radiative transfer". In: **Journal of Electronic Imaging** 12 (2003), p. 594.

[2] M. Frank, M. Seaid, A. Klar, R. Pinnau, G. Thommes, and J. Janicka. "A comparison of approximate models for radiation in gas turbines". In: **Progress in Computational Fluid Dynamics** 4 (2004), pp. 191–197.

[3] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". In: **arXiv preprint arXiv:1412.6980** (2014).

[4] R. A. S. Klein and P. H. A. Konzen. "PINN-SP1 for Absorption Estimation in Radiative Transfer: Discrete Detectors and Sensitivity Analisys". In: **III Encontro Regional de Matemática Aplicada e Computacional (ERMAC-SC 2026)**. To appear. 2026.

[5] E. W. Larsen, G. Thommes, A. Klar, M. Seaid, and T. Gotz. "Simplified PN approximations to the equations of radiative heat transfer and applications". In: **J. Comput. Phys.** 183 (2002), pp. 652–675.

[6] M. F. Modest. **Radiative Heat Transfer**. Elsevier Science, 2003.

[7] F. Murtagh. "Multilayer perceptrons for classification and regression". In: **Neurocomputing** 2 (1991), pp. 183–197.

[8] C. B. Pazinatto. "ADO formulation for the one-dimensional adjoint transport problem and application to an inverse source reconstruction problem". Master dissertation. Universidade Federal do Rio Grande do Sul, 2015.

[9] 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". In: **Journal of Computational Physics** 378 (2019), pp. 686–707.

[10] A. M. Stuart. "Inverse problems: A Bayesian perspective". In: **Acta Numerica** 19 (2010), pp. 451–559.

[11] Sifan Wang, Xinling Yu, and Paris Perdikaris. "When and why PINNs fail to train: A neural tangent kernel perspective". In: **Journal of Computational Physics** 449 (Jan. 2022), p. 110768. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2021.110768. URL: http://dx.doi.org/10.1016/j.jcp.2021.110768.