OPT2025: 17th Annual Workshop on Optimization for Machine Learning

Empirical-Bayes XTFC for Inverse Parameter Estimation

Vikas Dwivedi VIKAS.DWIVEDI@CREATIS.INSA-LYON.FR
Monica Sigovan MONICA.SIGOVAN@CREATIS.INSA-LYON.FR
Bruno Sixou BRUNO.SIXOU@INSA-LYON.FR
Creatis Biomedical Imaging Laboratory, INSA-Lyon, France

Abstract

We introduce Empirical-Bayes XTFC for inverse parameter estimation in boundary-value problems. The method embeds Dirichlet boundary conditions exactly via the Extreme Theory of Functional Connections (XTFC) and combines noisy measurements with physics residuals in a lightweight linear model built from fixed Gaussian RBF basis functions. We target the PDE parameter ν and select it by empirical Bayes: for each candidate ν we maximize the marginal likelihood (evidence) with respect to a single prior precision η placed on the XTFC weights (the linear coefficients in the constrained RBF basis), then choose the $\hat{\nu}$ achieving the highest evidence. This yields posterior means and uncertainty bands without iterative PDE solves or boundary-penalty tuning. On a 1D convection–diffusion benchmark with sharp boundary layers, the approach accurately recovers the diffusion coefficient ν from only 50 measurements with modest compute, and remains robust in the small-ν regime where identifiability is challenging. The formulation is simple, stable, and easy to implement. For reproducibility, our source code is available at https://github.com/vikas-dwivedi-2022/OPT_Bayesian_XTFC.

Keywords: Empirical-Bayes, Inverse Parameter Estimation, Extreme Theory of Functional Connections, Convection-Diffusion Equation

1. Introduction

Physics-informed neural networks (PINNs) [6] have emerged as a popular tool for solving forward and inverse problems governed by differential equations. Their appeal lies in unifying data and physics within a single optimization problem. However, standard PINN training relies on iterative backpropagation through deep architectures, which can be slow and memory intensive. Moreover, PINNs are known to struggle on boundary-layer and sharp-gradient problems due to spectral bias [8], often requiring specialized architectures [1] and loss functions [4].

To improve both speed and accuracy, recent work has explored alternatives, such as physics-informed extreme learning machines (PIELM) [3], Extreme Theory of Functional Connections (XTFC) [7] that avoid heavy gradient-based training. XTFC provides constrained expressions that satisfy boundary conditions exactly. In parallel, Bayesian variants of PIELM [5] have been developed to treat both forward and inverse settings within a linear-Gaussian framework, enabling evidence-based model selection and calibrated uncertainty without deep backpropagation.

Our Work. We combine these lines into a single method, Empirical-Bayes XTFC, for inverse parameter estimation. The key idea is simple: embed boundary conditions exactly via TFC; represent the solution in a fixed feature space (Gaussian RBFs); and stack a data block with a physics-residual block into a whitened linear model. Placing a Gaussian prior on the feature coefficients yields a

closed-form marginal likelihood (evidence). We scan the physical parameter (e.g., diffusivity) and perform empirical Bayes to tune the prior precision, selecting the parameter that maximizes the evidence. This delivers posterior means and uncertainty bands with a single Cholesky factorization—no boundary penalties, no deep networks, and no iterative backpropagation.

Contributions. (i) An Empirical-Bayes XTFC formulation for inverse problems that enforces boundary conditions exactly and couples data with physics in a linear-Gaussian model; (ii) an evidence-driven estimator that provides calibrated uncertainty and avoids hand-tuned loss weights; (iii) a practical recipe for sharp-gradient (boundary-layer) regimes via right-clustered collocation and RBF feature design; and (iv) empirical results on a convection–diffusion benchmark showing accurate recovery from as few as 50 measurements with modest runtime, and graceful degradation in the singularly perturbed regime.

Paper outline. The Empirical-Bayes XTFC model and inference method are described in Section 2. Section 3 details the numerical setup and presents the results. The conclusion is provided in Section 4, with supplementary derivations included in the Appendix.

2. Methodology

2.1. Inverse Problem Setup

We study the steady one–dimensional convection–diffusion BVP

$$\mathcal{R}_\nu[u](x) \;:=\; u'(x) - \nu\, u''(x) = 0, \qquad x \in (0,1),$$

with Dirichlet conditions $u(0) = BL = 0$, $u(1) = BR = 1$. The diffusion parameter $\nu > 0$ is *unknown* and is inferred from noisy observations

$$y_n = u(x_n;\nu) + \varepsilon_n, \qquad \varepsilon_n \sim \mathcal{N}(0,\sigma^2), \quad n = 1,\dots,N_{\text{data}}.$$

For controlled experiments we generate synthetic data from the analytic solution and add Gaussian noise of known standard deviation σ. To resolve the boundary layer when $\nu \ll 1$, both the data locations and the PDE collocation points are right–clustered near $x = 1$.

2.2. Trial Solution and Whitened Linear Model

Boundary conditions are enforced *exactly* using a constrained expression (XTFC):

$$u(x;c) = g(x) + \sum_{i=1}^{N_s} c_i\, \psi_i(x), \qquad g(x) = (1-x)\, BL + x\, BR, \qquad \psi_i(0) = \psi_i(1) = 0.$$

We adopt Gaussian RBF features $\phi(z) = e^{-z^2}$ with affine arguments $z_i(x) = m_i x + b_i$ and define

$$\psi_i(x) = \phi\big(z_i(x)\big) - (1-x)\,\phi\big(z_i(0)\big) - x\,\phi\big(z_i(1)\big).$$

Centers $\alpha_i^\star$ and widths $\sigma_{x,i}$ are chosen in physical space (linearly spaced); the affine parameters follow

$$m_i = \frac{1}{\sqrt{2}\,\sigma_{x,i}}, \qquad b_i = -m_i\,\alpha_i^\star,$$

so $z_i(x) = m_i(x - \alpha_i^\star)$. With $\mathcal{R}_\nu[v] = v' - \nu v''$, the residual of the trial is

$$\mathcal{R}_\nu[u](x; c) = f_g + \sum_{i=1}^{N_s} c_i \, r_i(x; \nu), \qquad f_g := \mathcal{R}_\nu[g] = BR - BL,$$

and $r_i(x; \nu) = \mathcal{R}_\nu[\psi_i](x)$ admits closed form (Appx. A).

We combine data fidelity and physics consistency in a *whitened* linear–Gaussian model:

$$\underbrace{\begin{bmatrix} \sqrt{\beta_{\text{data}}}\, H_{\text{data}} \\ \sqrt{\beta_{\text{pde}}}\, R_{\text{pde}}(\nu) \end{bmatrix}}_{\Phi(\nu) \in \mathbb{R}^{N \times M}} c = \underbrace{\begin{bmatrix} \sqrt{\beta_{\text{data}}}\,\big(y - g(x)\big) \\ \sqrt{\beta_{\text{pde}}}\,(-f_g\, \mathbf{1}) \end{bmatrix}}_{y(\nu) \in \mathbb{R}^N} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I),$$

where $H_{\text{data}}(n, i) = \psi_i(x_n)$, $R_{\text{pde}}(k, i) = r_i(x_k^{\text{pde}}; \nu)$, $\beta_{\text{data}} = \sigma^{-2}$ and $\beta_{\text{pde}} > 0$ are user–set precisions. Here $M = N_s$ (number of features) and $N = N_{\text{data}} + N_{\text{pde}}$. Note that $\Phi(\nu)$ and $y(\nu)$ depend on ν only through the PDE block.

2.3. Evidence–Driven Empirical-Bayes Estimation of ν

We place a zero–mean isotropic Gaussian prior on the XTFC coefficients, $c \sim \mathcal{N}(0, \eta^{-1} I)$ with precision $\eta > 0$. Given (ν, η), the posterior over c is

$$S_N(\nu, \eta) = \big(\eta I + \Phi^\top \Phi\big)^{-1}, \qquad m_N(\nu, \eta) = S_N \, \Phi^\top y.$$

The marginal log-likelihood (evidence) (Chapter 3 of PMRL book [2]) is

$$\log p(y \mid \nu, \eta) = \frac{M}{2} \log \eta - \frac{1}{2} \log \det\big(\eta I + \Phi^\top \Phi\big) - \frac{1}{2} \big\| y - \Phi m_N \big\|_2^2 - \frac{\eta}{2} \|m_N\|_2^2 - \frac{N}{2} \log(2\pi),$$

with $M = N_s$ the number of basis functions and N the number of stacked rows. For each ν on a user-specified logarithmic grid $\nu \in [\nu_{\min}, \nu_{\max}]$ with G points, we optimize the coefficient prior precision η by a damped fixed-point update initialized at a user-specified η_0 and constrained to $[\eta_{\min}, \eta_{\max}]$:

$$\gamma = \sum_{i=1}^{M} \frac{\sigma_i^2}{\sigma_i^2 + \eta}, \qquad \eta \leftarrow \frac{\gamma}{\|m_N\|_2^2}.$$

where $\{\sigma_i^2\}$ are the eigenvalues of $\Phi^\top \Phi$ (squared singular values of Φ). Unless stated otherwise, we use a log grid $[\nu_{\min}, \nu_{\max}] = [10^{-3}, 1]$ with $G = 500$ points, and empirical-Bayes update for η initialized at $\eta_0 = 10^{-7}$ with bounds $[\eta_{\min}, \eta_{\max}] = [10^{-12}, 10^{-2}]$, damping $\tau = 0.3$, tolerance 10^{-6}, and at most 50 iterations. We then select

$$\hat{\nu} = \arg\max_{\nu} \; \log p\big(y \mid \nu, \hat{\eta}(\nu)\big).$$

2.4. Prediction

For any $x_\star$, define $h_\star = (\psi_i(x_\star))_{i=1}^{N_s}$. The predictive mean and standard deviation are

$$\hat{u}(x_\star) = g(x_\star) + h_\star m_N, \qquad \text{Std}\big[u(x_\star)\big] = \sqrt{h_\star S_N h_\star^\top}\,.$$

Table 1: Inverse estimation of the diffusion parameter ν for the 1D convection–diffusion boundary-value problem using Empirical-Bayes XTFC. Each row reports collocation points, tunable features, ground-truth ν, estimated $\hat{\nu}$, and wall-clock time; $N_{\text{data}} = 50$ with 1% noise in all cases.

Collocation Points	Tunable Parameters	ν_{true}	ν_{pred}	Time (s)
200	100	0.5	0.512	2
200	100	0.05	0.049	2
600	300	0.005	0.006	32

3. Results and Discussion

Table 1 summarizes inverse estimates of the diffusion parameter ν using the proposed Empirical-Bayes XTFC formulation with Gaussian RBF features. The number of data points is fixed at $N_{\text{data}} = 50$ across all cases, while the number of PDE collocation points and tunable features is varied. Figure 1 shows the corresponding evidence distributions over ν and the posterior predictive fits.

Accuracy. For moderate values, the estimator recovers ν with small relative error: for $\nu_{\text{true}} = 0.5$ we obtain $\hat{\nu} = 0.512$ (absolute error 0.012, relative error 2.4%), and for $\nu_{\text{true}} = 0.05$ we obtain $\hat{\nu} = 0.049$ (absolute error 0.001, relative error 2.0%). The most challenging regime is the sharp boundary-layer case $\nu_{\text{true}} = 0.005$, for which $\hat{\nu} = 0.006$ (absolute error 0.001, relative error 20%). This upward bias at very small ν is expected due to increased ill-conditioning and the need for finer resolution of the layer.

Evidence profiles. The evidence curves in Figure 1 are unimodal and well peaked around the true parameter for $\nu_{\text{true}} \in \{0.5,\ 0.005\}$, indicating good identifiability.

Runtime and scaling. Wall-clock time increases from 2 s to 32 s when both the feature count grows from $N_s = 100$ to $N_s = 300$ and the collocation set from 200 to 600. This trend is consistent with the computational costs of assembling the stacked system Φ (roughly linear in the number of rows) and solving the regularized normal equations via Cholesky (dominant cost roughly cubic in N_s). In practice, the observed $16\times$ increase reflects both effects as well as cache and implementation overheads.

Qualitative behavior. Posterior predictive means closely track the analytic solution across the domain, with narrow uncertainty bands in the bulk and mild widening near the right boundary where gradients are steepest. Right-clustered sampling near $x = 1$ and exact BC satisfaction via XTFC contribute to stable training without boundary penalties.

4. Conclusion

We presented Empirical-Bayes XTFC for inverse parameter estimation in boundary-value problems, combining exact boundary satisfaction with an evidence-driven linear model that fuses data and physics. On a 1D convection–diffusion benchmark with sharp gradients, the method accurately recovers the diffusion coefficient from only 50 measurements, runs in modest time, and returns

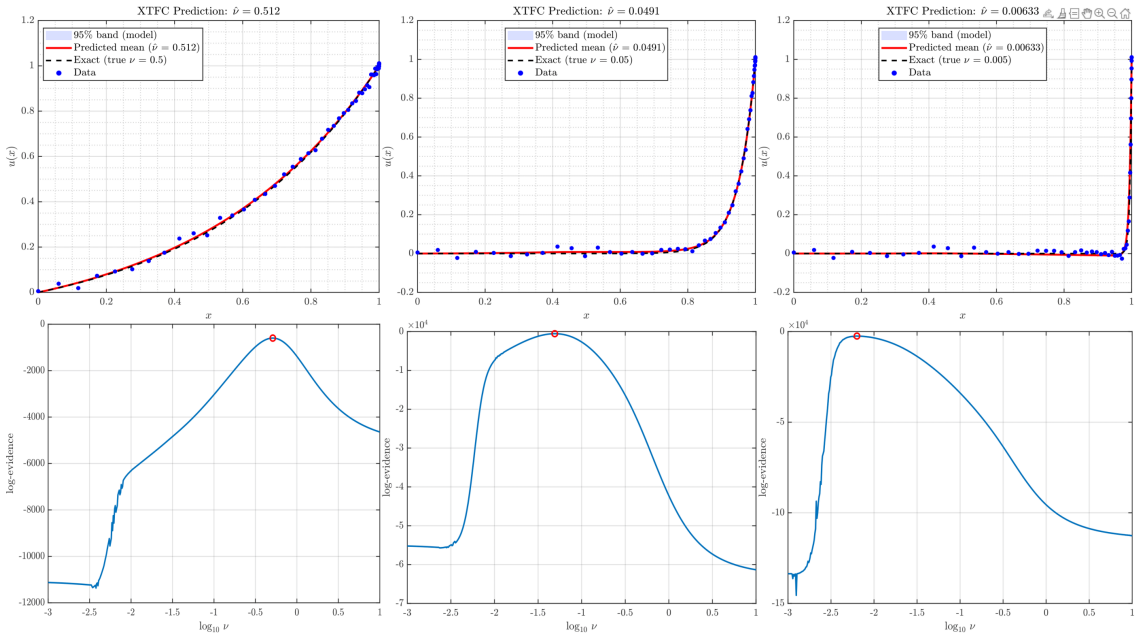

Figure 1: Inverse parameter estimation and corresponding evidence distribution for viscosity parameter ν_{true} at values 0.5, 0.05, and 0.005.

calibrated uncertainty; performance degrades gracefully in the small-ν regime where identifiability is hardest.

The approach is simple to implement, numerically stable, and free of boundary penalties, making it a practical alternative to heavier neural PDE solvers. Limitations include sensitivity to feature width/number and collocation density. Future work will address multi-parameter and space-varying coefficients, adaptive sampling guided by residuals or evidence, and extensions to higher-dimensional PDEs.

References

[1] Amirhossein Arzani, Kevin W. Cassel, and Roshan M. D'Souza. Theory-guided physics-informed neural networks for boundary layer problems with singular perturbation. *Journal of Computational Physics*, 473:111768, 2023. ISSN 0021-9991. URL https://www.sciencedirect.com/science/article/pii/S0021999122008312.

[2] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006. URL https://link.springer.com/book/9780387310732.

[3] Vikas Dwivedi and Balaji Srinivasan. Physics informed extreme learning machine (pielm)–a rapid method for the numerical solution of partial differential equations. *Neurocomputing*, 391:96–118, 2020. ISSN 0925-2312. URL https://doi.org/10.1016/j.neucom.2019.12.099.

[4] Seungchan Ko and Sanghyeon Park. Vs-pinn: A fast and efficient training of physics-informed neural networks using variable-scaling methods for solving pdes with stiff behavior. *Journal of Computational Physics*, 529:113860, 2025. ISSN 0021-9991. URL https://www.sciencedirect.com/science/article/pii/S0021999125001433.

[5] Xu Liu, Wen Yao, Wei Peng, and Weien Zhou. Bayesian physics-informed extreme learning machine for forward and inverse pde problems with noisy data. *Neurocomputing*, 549: 126425, 2023. ISSN 0925-2312. URL https://doi.org/10.1016/j.neucom.2023.126425.

[6] 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. ISSN 0021-9991. URL https://doi.org/10.1016/j.jcp.2018.10.045.

[7] Enrico Schiassi, Roberto Furfaro, Carl Leake, Mario De Florio, Hunter Johnston, and Daniele Mortari. Extreme theory of functional connections: A fast physics-informed neural network method for solving ordinary and partial differential equations. *Neurocomputing*, 457:334–356, 2021. ISSN 0925-2312. URL https://doi.org/10.1016/j.neucom.2021.06.015.

[8] 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. ISSN 0021-9991. URL https://www.sciencedirect.com/science/article/pii/S002199912100663X.

Appendix A. XTFC Basis and Residual (Gaussian RBFs)

Let $\phi(z) = e^{-z^2}$, $\phi'(z) = -2z\,e^{-z^2}$, $\phi''(z) = (4z^2 - 2)e^{-z^2}$, and $z_i(x) = m_i x + b_i$. Define $A_i = \phi(z_i(0))$, $B_i = \phi(z_i(1))$, and

$$\psi_i(x) = \phi(z_i(x)) - (1-x)A_i - xB_i, \qquad \psi_i(0) = \psi_i(1) = 0.$$

With $\mathcal{R}_\nu[v] = v' - \nu v''$ and $z_i'(x) = m_i$,

$$r_i(x;\nu) = \mathcal{R}_\nu[\psi_i](x) = (A_i - B_i) + \phi(z_i(x))\Big[-2m_i z_i(x) - \nu m_i^2\big(4z_i(x)^2 - 2\big)\Big].$$

Appendix B. Whitened Stacking and Evidence

Data rows: $H_{\text{data}}(n,i) = \psi_i(x_n)$, targets $y_{\text{data}} = y - g(x)$. PDE rows: $R_{\text{pde}}(k,i) = r_i(x_k^{\text{pde}};\nu)$, targets $y_{\text{pde}} = -f_g\mathbf{1}$, with $f_g = BR - BL$. The stacked, whitened system is $\Phi c \approx y$ as in the main text. Posterior, evidence, and the update for η follow directly; we evaluate $\log p(y \mid \nu, \eta)$ on a log grid of ν and pick the maximizer.