Runge–Kutta Physics-Informed Neural Networks for Evolutionary PDEs

@ School of Mathematical and Computer Sciences (MACS), Heriot-Watt University

Costas Smaragdakis

Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean
& Institute of Applied and Computational Mathematics, FORTH

joint work with G. Akrivis & C. Makridakis

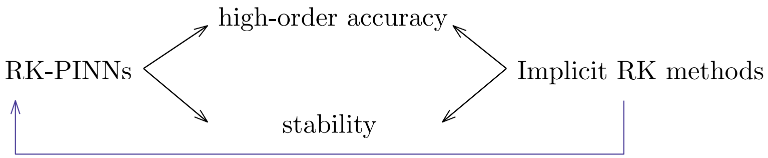
Motivation - ANNs as PDE Solvers

The Paradigm Shift: Transitioning from classical mesh-based methods to mesh-free, Neural Network approximations.

Key Motivations:

- High-Dimensionality: Standard meshing is prohibitive.
- Unified framework: Same infrastructure for *forward* and *inverse* problems.
- Optimization-Based: PDE solving becomes an optimization problem.

Challenge: Overcoming the stability & optimization issues of standard ANN PDE solvers.



Introduction

The Problem Setting: We consider the evolution equation for $u(t,x)$:

$$\begin{cases} \partial_t u(t,x) + \mathcal{A}u(t,x) = f(t,x), & (t,x) \in (0,T] \times \Omega \\ u(0,\cdot) = u_0(\cdot) \end{cases}$$

$\mathcal{A} : \mathcal{D}(\mathcal{A}) \subset L^2(\Omega) \to L^2(\Omega)$ is a self-adjoint, positive definite spatial operator.

The Continuous Loss:

$$\mathcal{G}(u) \coloneqq \int_0^T \left\| \partial_t u(t,\cdot) + \mathcal{A}u(t,\cdot) - f(t,\cdot) \right\|_{L^2(\Omega)}^2 dt + \left\| u(0,\cdot) - u_0(\cdot) \right\|_{H^1(\Omega)}^2.$$

- $\mathcal{G}(u) \geq 0$ for all admissible functions u.
- Well-posedness implies the exact solution $u^\star$ is the unique global minimizer ($\mathcal{G}(u^\star) = 0$).

Network Space & Discretization Setup

Neural Network Space: We define the NN space over a NN architecture:

$$\mathcal{N}_\ell \coloneqq \{U_\theta(\cdot,\cdot) : [0,T] \times \Omega \longrightarrow \mathbb{R}^M \mid \theta \in \mathbb{R}^{N_{params}(\ell)}\},$$

where ℓ is the capacity parameter of the ANNs within the space.

Time Discretization: We associate with ℓ a time partition $\mathcal{T}_\ell = \{t_n\}_{n=0}^{N(\ell)}$ of $(0,T]$.

- Subintervals: $J_n \coloneqq (t_n, t_{n+1}] \quad (k_n \coloneqq t_{n+1} - t_n)$.
- Mesh size: $k(\ell) \coloneqq \max_n k_n \quad (\, k(\ell) \longrightarrow 0$ as $\ell \longrightarrow \infty)$.

Auxiliary Nodes: Let $\{\tilde{c}_j\}_{j=0}^{q} \subset [0,1]$ be fixed nodes satisfying:

$$0 = \tilde{c}_0 < \cdots < \tilde{c}_q = 1$$

Local Auxiliary Nodes on J_n:

$$\tilde{t}_{nj} \coloneqq t_n + \tilde{c}_j k_n, \quad j = 0, \ldots, q.$$

The Approximation Space V_ℓ

The Interpolation Operator $\hat{I}_q$: Let $\{\tilde{\ell}_j\}_{j=0}^q$ be the standard Lagrange polynomials on $[0,1]$ satisfying $\tilde{\ell}_j(\tilde{c}_i) = \delta_{ji}$.

The operator $\hat{I}_q$ maps any continuous function w to a piecewise polynomial:

$$(\hat{I}_q w)(t,\cdot)\big|_{J_n} \coloneqq \sum_{j=0}^{q} \tilde{\ell}_{nj}(t) w(\tilde{t}_{nj},\cdot), \qquad \tilde{\ell}_{nj}(t) \coloneqq \tilde{\ell}_j\left(\frac{t-t_n}{k_n}\right), \quad t \in J_n.$$

The Approximation Space V_ℓ

The space $V_\ell \subset C([0,T]; L^2(\Omega))$ is defined as follows:

$$V_\ell \coloneqq \left\{\hat{u} : \hat{u} = \hat{I}_q U_\theta, \quad U_\theta \in \mathcal{N}_\ell\right\}.$$

$$\text{explicitly on } J_n : \quad \forall \hat{u} \in V_\ell, \quad \hat{u}(t,\cdot) = \sum_{j=0}^{q} \tilde{\ell}_j\left(\frac{t-t_n}{k_n}\right) U_\theta(\tilde{t}_{nj},\cdot).$$

Continuity: Since $\tilde{c}_0 = 0$ and $\tilde{c}_q = 1$, we have $\hat{u}(t_n^-,\cdot) = U_\theta(t_n,\cdot) = \hat{u}(t_n^+,\cdot)$.

Density of Approximation Space

Function Approximation: Let $v \in C([0,T]; L^2(\Omega))$ be a target function.

For any $\epsilon > 0$, there exists a sufficiently large capacity parameter ℓ such that there exists a function $\hat{u} \in V_\ell$ satisfying:

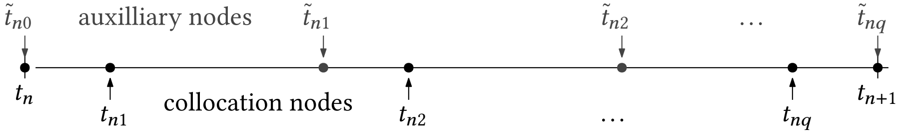
$$\|v - \hat{u}\|_{C([0,T];L^2(\Omega))} < \epsilon.$$

Two sources of error:

- Interpolation Error
- Network Approximation Error

$$\|v - \hat{u}\| \leq \underbrace{\|v - \hat{I}_q v\|}_{\substack{\text{Interpolation Error} \\ (\to 0 \text{ since } k(\ell) \to 0)}} + \underbrace{\|\hat{I}_q (v - U_\theta)\|}_{\substack{\text{Network Approximation} \\ (\to 0 \text{ as } \ell \to \infty)}}$$

Collocation Runge–Kutta Formulation



Collocation Nodes: Let the Runge–Kutta method be characterized by q distinct collocation nodes $0 \leq c_1 < \cdots < c_q \leq 1$. On each time interval $J_n = (t_n, t_{n+1}]$ of size k_n, we map these nodes to:

$$t_{ni} \coloneqq t_n + c_i k_n, \quad i = 1, \dots, q.$$

R-K Interpolation Operator I_{q-1}: For any function $w(t, x)$, the operator interpolates in time onto polynomials of degree $q - 1$:

$$(I_{q-1}w)(t, x) \coloneqq \sum_{i=1}^{q} \ell_i \left(\frac{t - t_n}{k_n} \right) w(t_{ni}, x), \quad t \in J_n,\ x \in \Omega$$

where $\{\ell_i\}_{i=1}^{q}$ are the Lagrange basis polynomials satisfying $\ell_i(c_j) = \delta_{ij}$.

The Runge–Kutta PINN Formulation

The collocation RK–PINN method requires the discrete solution $\hat{u} \in V_\ell$ to satisfy the PDE at the collocation nodes.

$$\partial_t \hat{u}(t,x) + \mathcal{A}(I_{q-1}\hat{u})(t,x) = (I_{q-1}f)(t,x), \quad \forall t \in J_n,\ x \in \Omega.$$

The Discrete Loss Functional $\mathcal{G}_\ell$

We minimize the norm of this residual. For any $u \in V_\ell$, we define:

$$\mathcal{G}_\ell(u) \coloneqq \int_0^T \left\| \partial_t u(t,\cdot) + \mathcal{A} I_{q-1} u(t,\cdot) - I_{q-1} f(t,\cdot) \right\|_{L^2(\Omega)}^2 dt + \left\| u(0,\cdot) - u_0(\cdot) \right\|_{H^1(\Omega)}^2$$

We set $\mathcal{G}_\ell(u) = +\infty$ if $u \notin V_\ell$.

Optimization Problem:

$$\hat{u}_\ell \leftarrow \operatorname{argmin}_{u \in V_\ell} \mathcal{G}_\ell(u)$$

Theoretical Goals:

1. Stability
2. Convergence

Sufficient Condition for Stability

To guarantee stability, we require the numerical method to respect the energy dynamics of the PDE.

The Continuous Energy Law:

$$\underbrace{\int_{J_n} (\partial_t u, \mathcal{A}u)\, dt}_{\text{Continuous cross-term}} = \int_{J_n} \frac{1}{2}\frac{d}{dt}\|\mathcal{A}^{1/2}u\|^2 dt = \underbrace{\frac{1}{2}\left(\|\mathcal{A}^{1/2}u(t_{n+1},\cdot)\|^2 - \|\mathcal{A}^{1/2}u(t_n,\cdot)\|^2\right)}_{\text{Energy Change }(\Delta\mathcal{E})}$$

The Discrete Energy Condition: In RK-PINN, we approximate $\partial_t u \approx \partial_t \hat{u}$ and $\mathcal{A}u \approx \mathcal{A}I_{q-1}\hat{u}$.

$$\underbrace{\int_{J_n} (\partial_t \hat{u}, \mathcal{A}I_{q-1}\hat{u})\, dt}_{\text{Model cross-term}} \geq \underbrace{\frac{1}{2}\left(\|\mathcal{A}^{1/2}\hat{u}(t_{n+1},\cdot)\|^2 - \|\mathcal{A}^{1/2}\hat{u}(t_n,\cdot)\|^2\right)}_{\text{Model Energy Change }(\Delta\hat{\mathcal{E}})}$$

From Energy Condition to Maximal Regularity

If the energy condition holds, the numerical method is stable.

$$\int_{J_n} \|\partial_t \hat{u}\|^2 + \int_{J_n} \|\mathcal{A} I_{q-1} \hat{u}\|^2 + 2 \int_{J_n} (\partial_t \hat{u}, \mathcal{A} I_{q-1} \hat{u}) = \int_{J_n} \|I_{q-1} f\|^2$$

$$\Big\downarrow \text{(Energy Condition)}$$

$$\int_{J_n} \|\dots\|^2 + \int_{J_n} \|\dots\|^2 + \left(\|\mathcal{A}^{1/2} \hat{u}_{n+1}\|^2 - \|\mathcal{A}^{1/2} \hat{u}_n\|^2 \right) \leq \int_{J_n} \|\dots\|^2$$

$$\Big\downarrow \sum_{n=0}^{N-1} \cdot$$

$$\|\mathcal{A}^{1/2} \hat{u}_N\|^2 + \|\partial_t \hat{u}\|_{L^2}^2 + \|\mathcal{A} I_{q-1} \hat{u}\|_{L^2}^2 \leq \|\mathcal{A}^{1/2} u_0\|^2 + \|I_{q-1} f\|_{L^2}^2$$

We obtain the Maximal Regularity property.

Maximal Regularity: Gauss & Radau IIA

- Weights $b_i > 0$.
- The stability matrix $M = (m_{ij})$ with $m_{ij} = b_i a_{ij} + b_j a_{ji} - b_i b_j$ is positive semi-definite.

Expanding the norm of the update yields:

$$\|\mathcal{A}^{1/2}\hat{u}_{n+1}\|^2 = \|\mathcal{A}^{1/2}\hat{u}_n\|^2 + 2k_n \sum_{i=1}^{q} b_i(\partial_t \hat{u}_i, \mathcal{A}\hat{u}_i) - \underbrace{k_n^2 \sum_{i,j=1}^{q} m_{ij}(\mathcal{A}^{1/2}\partial_t \hat{u}_i, \mathcal{A}^{1/2}\partial_t \hat{u}_j)}_{\geq 0 \quad \text{(since } M \text{ is P.S.D.)}}$$

Using the exactness of quadrature for the degree $2q-2$:

$$k_n \sum_{i=1}^{q} b_i(\partial_t \hat{u}_i, \mathcal{A}\hat{u}_i) = \int_{J_n} \underbrace{(\partial_t \hat{u}, \mathcal{A} I_{q-1}\hat{u})}_{\text{Integrand} \in \mathbb{P}_{2q-2}} \, dt.$$

$$\Rightarrow \quad \int_{J_n} (\partial_t \hat{u}, \mathcal{A} I_{q-1}\hat{u}) \, dt \geq \frac{1}{2}\left(\|\mathcal{A}^{1/2}\hat{u}_{n+1}\|^2 - \|\mathcal{A}^{1/2}\hat{u}_n\|^2\right).$$

Maximal Regularity: Lobatto IIIA

The Lobatto polynomial $\hat{u}$ (degree q) does not satisfy the energy condition directly.

The reconstruction $\tilde{u} = I_{q-1}\hat{u}$ satisfies the exact energy balance:

$$\int_{J_n} (\partial_t \tilde{u}, \mathcal{A}\tilde{u})\, dt = \frac{1}{2}\left(\|\mathcal{A}^{1/2}\tilde{u}(t_{n+1},\cdot)\|^2 - \|\mathcal{A}^{1/2}\tilde{u}(t_n,\cdot)\|^2\right)$$

This works because the “error” in the time derivative is invisible to the physics of the problem:

$$\int_{J_n} (\partial_t \hat{u} - \partial_t \tilde{u}, \underbrace{\mathcal{A}\tilde{u}}_{\text{Test Function}})\, dt = 0 \implies \int_{J_n} (\partial_t \tilde{u}, \mathcal{A}\tilde{u})\, dt = \int_{J_n} (\partial_t \hat{u}, \mathcal{A} I_{q-1}\hat{u})\, dt$$

- Integration by Parts: Shifts the derivative to $\mathcal{A}\tilde{u}$ (degree $q-1$).
- Boundary Terms: Vanish because collocation nodes include endpoints.
- Interior Integral: Vanishes because quadrature rule is exact for the resulting polynomial.

$$\tilde{u}(t_n,\cdot) = \hat{u}(t_n,\cdot),\ \forall t_n \implies \frac{1}{2}\left(\|\mathcal{A}^{1/2}\tilde{u}(t_{n+1},\cdot)\|^2 - \|\mathcal{A}^{1/2}\tilde{u}(t_n,\cdot)\|^2\right) = \frac{1}{2}\left(\|\mathcal{A}^{1/2}\hat{u}(t_{n+1},\cdot)\|^2 - \|\mathcal{A}^{1/2}\hat{u}(t_n,\cdot)\|^2\right)$$

Maximal Regularity of $\tilde{u}$ $\implies$ **Maximal Regularity of $\hat{u}$**

Convergence of Minimizers

Let $\{\hat{u}_\ell\}$ be the sequence of minimizers of the discrete energy $\mathcal{G}_\ell$. As the network capacity $\ell \to \infty$, the sequence converges to the exact solution u:

$$\hat{u}_\ell \to u^\star \quad \text{in } L^2((0,T); H^1(\Omega))$$

The De Giorgi Framework (Γ-convergence):

- Step 1: Using the uniform bounds from Maximal Regularity and lower semicontinuity:

$$\mathcal{G}(\hat{u}) \le \liminf_{\ell\to\infty} \mathcal{G}_\ell(\hat{u}_\ell)$$

- Step 2: We construct a *recovery sequence* w_ℓ:

$$w_\ell \to u^\star \implies \lim_{\ell\to\infty} \mathcal{G}_\ell(w_\ell) = \mathcal{G}(u^\star) = 0$$

Conclusion:

$$0 \le \mathcal{G}(\hat{u}) \le \liminf \mathcal{G}_\ell(\hat{u}_\ell) \le \limsup \mathcal{G}_\ell(\hat{u}_\ell) \le \limsup \mathcal{G}_\ell(w_\ell) = 0 \implies \mathcal{G}(\hat{u}) = 0 \implies \hat{u} = u^\star$$

RK-PINNs: Towards Applications

Step 1: Select the Underlying Collocation RK Method ($q = 3$)

RK Method	Collocation Nodes (c_i)	Key Features
Gauss	$\frac{1}{2} - \frac{\sqrt{15}}{10}, \quad \frac{1}{2}, \quad \frac{1}{2} + \frac{\sqrt{15}}{10}$	Energy Conservation
Lobatto IIIA	$0, \quad \frac{1}{2}, \quad 1$	Energy Stable
Radau IIA	$\frac{4-\sqrt{6}}{10}, \quad \frac{4+\sqrt{6}}{10}, \quad 1$	L-Stable

Step 2: Instantiate the Approximation Space

The approximation space V_ℓ is fully determined once we fix the following components:

- A Network Space $\mathcal{N}_\ell$,
- A fixed time partition $\mathcal{T}_\ell$,
- The auxiliary nodes $0 = \tilde{c}_0 < \tilde{c}_1 < \cdots < \tilde{c}_q = 1$.

Problem Formulation & Collocation Residual

(for simplicity $f = 0$) Find $u : \Omega \times [0, T] \to \mathbb{R}^M$ such that:

$$u_t + \mathcal{A}u = \mathbf{0}, \quad u(x, 0) = u_0(x).$$

+Boundary Conditions

The Collocation Runge-Kutta Residual

We define the residual $\zeta(x, t)$ via the RK stages. Within $J_n = (t_n, t_{n+1}]$:

$$\zeta(t_{nj}, x) = \sum_{i=0}^{q} \left(k_n^{-1} \tilde{\ell}_i'(c_j) U_\theta(\tilde{t}_{ni}, x) - \tilde{\ell}_i(c_j) \mathcal{A} U_\theta(\tilde{t}_{ni}, x)\right)$$

$$\zeta(t, x) = \sum_{j=1}^{q} \ell_j \left(\frac{t - t_n}{k_n}\right) \zeta(t_{nj}, x), \quad x \in \Omega.$$

Applications:

- 2D Heat Equation
- 2D Wave Equation

Discrete Loss Functionals & Optimization

We minimize the total cost $C_\Omega + C_0 + \sum_s C_{\partial\Omega_s}$.

Interior Cost: (Sobol sampling $\{x_r\}$ + RK Time Integration)

$$C_\Omega[\theta] = \frac{\mathrm{Vol}(\Omega)}{R} \sum_{m,r,n} \underbrace{k_n \sum_{j=1}^{q} w_j \zeta_m(t_{nj}, x_r)^2}_{=\int_{J_n} \zeta_m^2 dt} .$$

Initial & Boundary Costs:

$$C_0[\theta] = \frac{\mathrm{Vol}(\Omega)}{R} \sum_{m,r} (u_m(0, x_r; \theta) - u_{m0}(x_r))^2$$

$$C_{\partial\Omega_s}[\theta] = \frac{\mathrm{Vol}(\partial\Omega_s)}{R} \sum_{m,r',n} (u_m(t_n, x_{r'}) - u_{ms}(x_{r'}))^2 \quad \text{(Dirichlet condition)}$$

Final Optimization Problem:

$$\theta^\star \leftarrow \min_{\theta\in\Theta} \left(C_\Omega[\theta] + C_0[\theta] + \sum_s C_{\partial\Omega_s}[\theta] \right) .$$

Application: Heat Equation (Discontinuous Data)

Problem Setup

Diffusion on $\Omega = (0,1)^2$ with $k = 0.02$ and Neumann BCs:

$$\begin{cases} u_t - k(u_{xx} + u_{yy}) = 0, \\ \partial_n u = 0 \quad \text{on } \partial\Omega. \end{cases}$$

Discontinuous Initial Value

The initial state is a characteristic function χ_D on a disk D (center $(0.6, 0.7)$, radius 0.1):

$$u(0,x,y) = \begin{cases} 1 & \text{if } (x,y) \in D \\ 0 & \text{otherwise} \end{cases}$$

Analysis Objectives

1. Smoothing Property

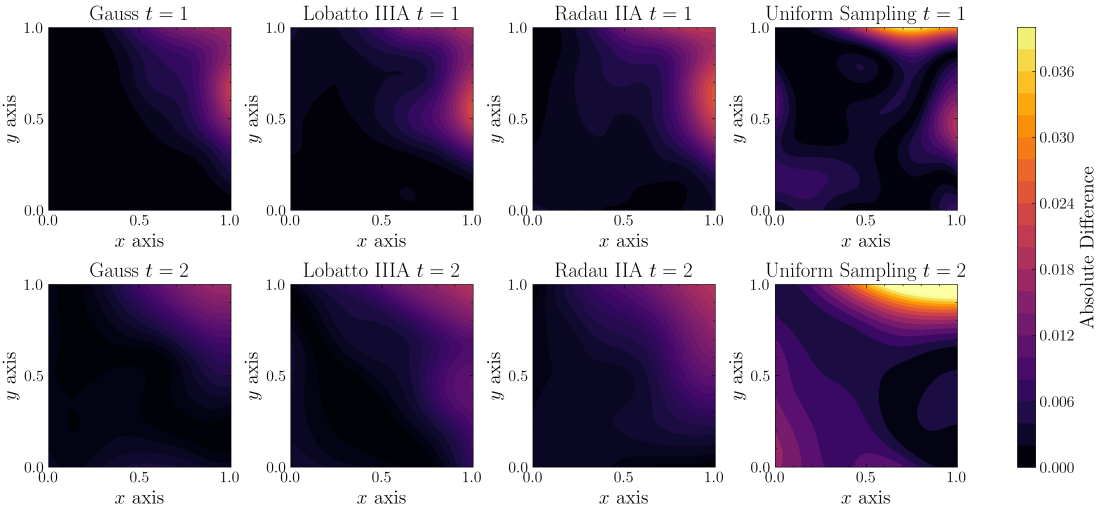
- The discontinuity leads to oscillations.
- Radau IIA damps high frequencies (L-stable), whereas Gauss and Lobatto IIIA may exhibit oscillations near $t = 0$.

2. Heat Conservation

- Due to Neumann BCs, total heat must be invariant:

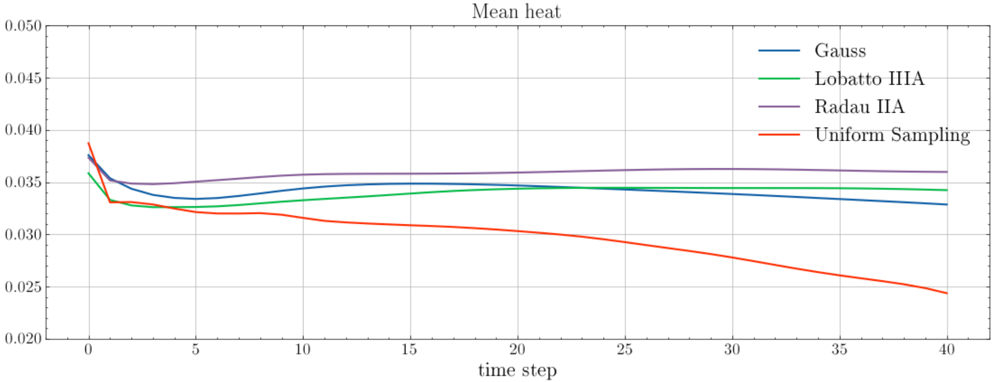
$$\frac{d}{dt}\int_\Omega u(t,x,y)d\Omega = 0$$

Application: Heat Equation (Discontinuous Data)

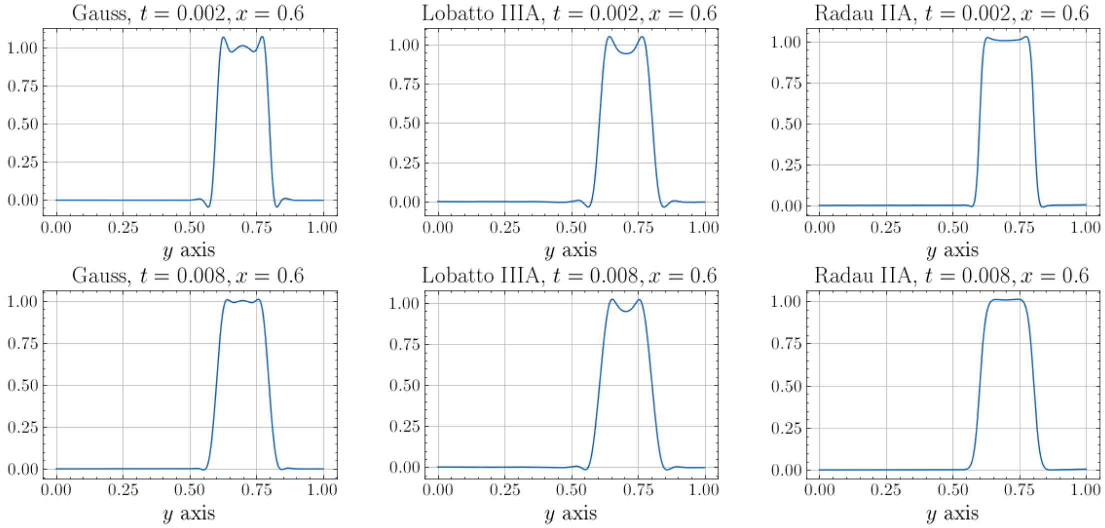


Application: Heat Equation (Discontinuous Data)

$$\int_{\Omega} u(t, x, y)\, dx = \int_{\Omega} u(0, x, y)\, d\Omega\,.$$



Application: Heat Equation (Discontinuous Data)



Application: Wave Equation

Problem Setup

We consider an initial value wave propagation problem on $\Omega := (0,1)^2$ with homogeneous Dirichlet boundary conditions:

$$
\begin{aligned}
u_{tt} - c^2(u_{xx} + u_{yy}) &= 0, \quad t \in (0,1],\ (x,y) \in \Omega,\ c = 0.5,\\
u(0,x,y) &= \left(0.5 + 0.5\cos\left(4\pi\sqrt{(x-0.3)^2 + (y-0.5)^2}\right)\right)\chi_D(x,y),\\
u_t(0,x,y) &= 0, \quad (x,y) \in \Omega,\\
u &= 0, \quad \text{on } (0,1) \times \partial\Omega
\end{aligned}
$$

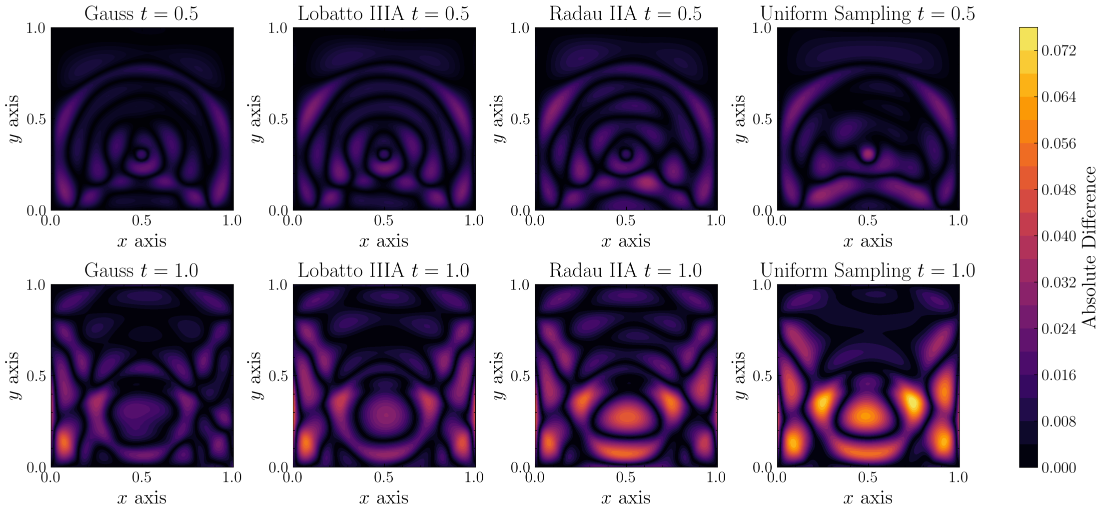
where χ_D is the characteristic function of the disk D (center $(0.3, 0.5)$, radius 0.25).

Reformulation: To apply the RK-PINN, we introduce velocity $v := u_t$ to obtain the system:

$$
\begin{pmatrix} u \\ v \end{pmatrix}_t + \begin{pmatrix} 0 & -I \\ -c^2\Delta & 0 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
$$

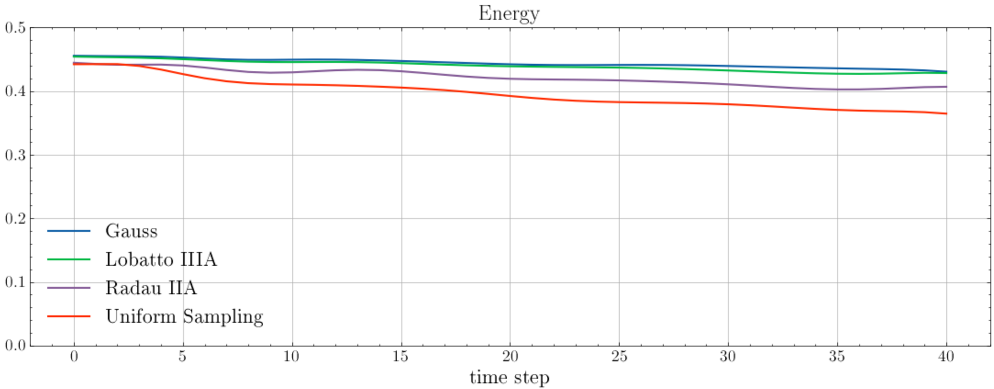
Analysis Objective: Whether the numerical methods preserve the total energy.

Application: Wave Equation



Application: Wave Equation

$$E(t) := \frac{1}{2}\|u_t(t,\cdot)\|_{L^2}^2 + \frac{1}{2}c^2\|\nabla u(t,\cdot)\|_{L^2}^2 = E(0)$$



Thank you

G. Akrivis, C. G. Makridakis, C. Smaragdakis: Runge-Kutta Physics Informed Neural Networks: Formulation and Analysis. Numerische Mathematik, 2025.