VNIVERSITAT DE VALÈNCIA
Escola Tècnica Superior d'Enginyeria (ETSE-UV)

Kerr quasinormal modes without separating variables

A two-dimensional hyperboloidal Teukolsky solver with physics-informed neural networks

Antonio Ferrer-Sánchez

IDAL, Department of Electronic Engineering, ETSE-UV, Universitat de València

With D. D. Doneva, J. D. Martín-Guerrero, S. S. Yazadjiev, R. Ruiz de Austri-Bazan, Y. Vives-Gilabert and J. A. Font

November 2026

Outline

1. Introduction
2. State of the art
3. Methods
4. Results
5. Conclusions

Speaker and collaborators



Antonio Ferrer-Sánchez

PhD candidate, Doctoral Programme in Electronic Engineering, Universitat de València. Supervisors: J. D. Martín-Guerrero and R. Ruiz de Austri. Thesis: PINNs for relativistic hydrodynamics, gravitational collapse and quantum control.

Machine learning · IDAL

Ferrer-Sánchez, Martín-Guerrero, Vives-Gilabert. Dept. of Electronic Engineering, ETSE-UV; ValgrAI.

Gravitational physics · UV

Doneva, Font. Dept. of Astronomy and Astrophysics; Tübingen; Observatori Astronòmic.

Sofia and IFIC

Yazadjiev, Ruiz de Austri-Bazan. Sofia University and IMI, Bulgarian Academy of Sciences; IFIC, CSIC-UV.

Code, configurations and the 336 trained networks: doi:10.5281/zenodo.23656788.

Part 1

Introduction

1. Introduction
2. State of the art
3. Methods
4. Results
5. Conclusions

Quasinormal modes

A perturbed black hole rings down in quasinormal modes whose complex frequencies depend only on **M and a** in general relativity.

$$h(t) = \sum_{\ell mn} A_{\ell mn}\, e^{-(t-t_0)/\tau_{\ell mn}} \cos\left[2\pi f_{\ell mn}(t-t_0)+\phi_{\ell mn}\right], \qquad \omega_{\ell mn} = 2\pi f_{\ell mn} - \frac{i}{\tau_{\ell mn}}$$

Eigenvalues

Outgoing at null infinity, ingoing at the horizon: a **non-Hermitian** problem with $\mathrm{Im}\,\omega < 0$.

Labels

ℓ, m angular indices, n the overtone. We follow **seven branches** (ℓ, m, n) in spin.

Frequency and damping time

$f = \mathrm{Re}\,\omega/2\pi$; $\tau = 1/|\mathrm{Im}\,\omega|$, the time for the amplitude to fall by e.

[Kokkotas and Schmidt 1999; Berti, Cardoso and Starinets 2009] Frequencies are quoted as $M\omega$ against a/M.

Teukolsky equation

Gravitational perturbations of Kerr obey one master equation for Ψ_s, $s = -2$, which **separates** into a radial and an angular ODE.

$$\left[\frac{(r^2+a^2)^2}{\Delta} - a^2\sin^2\theta\right]\partial_t^2\Psi_s + \frac{4Mar}{\Delta}\partial_t\partial_\phi\Psi_s - \Delta^{-s}\partial_r\left(\Delta^{s+1}\partial_r\Psi_s\right) - \frac{1}{\sin\theta}\partial_\theta\left(\sin\theta\,\partial_\theta\Psi_s\right) + \cdots = 0$$

Separation

$\Psi_s = e^{-i\omega t}e^{im\phi}R(r)S(\theta)$, coupled through ω and a constant $A_{\ell m}$.

Reference spectra

Leaver's continued fractions, through the **qnm** package, give the values we compare with.

Limitation

Separability is a property of Kerr; it is **not guaranteed** beyond it.

[Teukolsky 1973; Leaver 1985; Stein 2019] $\Delta = r^2 - 2Mr + a^2$.

Non-separable perturbations beyond Kerr

In modified gravity and for Kerr–Newman the perturbation equations **do not separate**; only the t and ϕ dependence factor out.

$$\Psi_s(t, r, \theta, \phi) = e^{-i\omega t}\, e^{im\phi}\, \Phi(r, \theta), \qquad \text{solve for } \Phi \text{ and } \omega \text{ together}$$

Scalar–Gauss–Bonnet (sGB)

A scalar coupled to $R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$; hairy black holes, rotating ones solved spectrally.

Kerr–Newman

Gravitational and electromagnetic perturbations couple and do not separate.

Our test bed

Kerr solved **as if** it did not separate, where the answer is known.

[Chung and Yunes 2024; Blázquez-Salcedo et al. 2025; Dias, Godazgar and Santos 2015] EdGB: Einstein–dilaton–Gauss–Bonnet.

Hyperboloidal formulation

On slices that end on $\mathscr{I}^+$ and $\mathscr{H}^+$ the QNM conditions become **regularity** of a field $h(\sigma, u)$, and the problem is quadratic in ω.

$$\sigma = \frac{r_+}{r} \in [0,1], \quad u = \cos\theta, \qquad \mathcal{L}_1[h] + \lambda\,\mathcal{L}_2[h] - w\,\lambda^2 h = 0, \qquad \lambda = -i\,r_+\omega$$

Domain

$\sigma = 0$ is null infinity, $\sigma = 1$ the horizon: both are **inside** the square.

No boundary data

The slicing carries the outgoing and ingoing behaviour; no Leaver prefactor is needed.

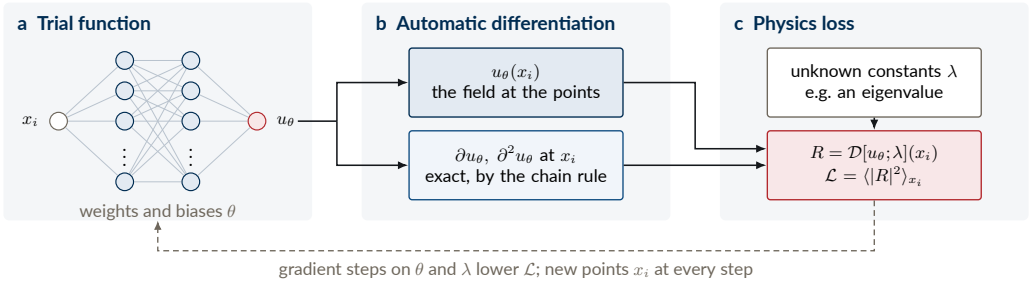
Generality

It uses the horizon structure, not a closed-form asymptotic solution.

[Zenginoğlu 2008; Ansorg and Panosso Macedo 2016; Panosso Macedo 2020; Assaad and Panosso Macedo 2025] $\mathcal{L}_{1,2}$: second- and first-order operators in (σ, u).

PINN definition

A PINN represents the unknown by a neural network and **minimises the residual of the equation**; no solution data are used.



$\mathcal{D}$: the differential operator of the equation, so $R = 0$ for an exact solution. [Lagaris, Likas and Fotiadis 1998; Raissi, Perdikaris and Karniadakis 2019] Autodiff: [Baydin et al. 2018].

PINN strengths and limitations

PINNs suit scattered points and numerical coefficients, but are **less precise and slower** than spectral methods on smooth problems.

Strengths: flexibility	Limitations: precision and cost	Remedies used here
Points anywhere, redrawn each step. Coefficients known only on a grid. Parameters as inputs, inverse problems.	Errors of 10^{-3} to 10^{-6}. Smooth parts learned first. Hours of training.	A curriculum in the spin. Relative residuals. A second-order polish.

[Rahaman et al. 2019; Wang, Teng and Perdikaris 2021; Krishnapriyan et al. 2021]

Part 2

State of the art

1 Introduction

2 State of the art

3 Methods

4 Results

5 Conclusions

Previous PINN studies of QNMs

Separated PINNs reached errors of **0.1 to 0.9 %** for the Kerr (2,0,0) mode, with one network for $R(r)$ and one for $S(\theta)$.

a/M	Re error	Im error
Luna et al. (2023), separated PINN		
0	0.330 %	0.108 %
0.8	0.233 %	0.853 %
0.98	0.564 %	0.337 %
0.9998	0.511 %	0.073 %
This work, worst of 16 spins	**0.0001 %**	**0.0012 %**

Modified gravity (2024)

Axial (odd-parity) QNMs of static scalarised sGB black holes; below 0.1 % over most of the range.

Missing

A PINN for the **rotating, non-separated** problem.

[Luna et al. 2023, Tables I and II; Luna et al. 2024] About 2 min per mode on a laptop CPU. Error definitions differ: compare orders of magnitude only.

Spectral and hybrid spectral–PINN methods

On the same fundamentals and spins, the PINN is **at the level of spectral solvers**: $\leq 4 \times 10^{-6}$ to $a/M = 0.95$.

Method	Reported, Kerr	Spins	This work, same spins
Metric perturbations, Chung, Wagle and Yunes 2024	$< 10^{-5}$	$a/M \leq 0.95$	$\mathbf{3.6 \times 10^{-6}}$
Metric perturbations, Blázquez-Salcedo et al. 2024	$\leq 10^{-6}$	$a/M < 0.8$	$\mathbf{2.6 \times 10^{-6}}$
	$< 10^{-3}$	$a/M < 0.98$	$\mathbf{4.9 \times 10^{-5}}$
Hybrid spectral–PINN, Pombo and Pizzuti 2026	Re + Im: 0.1 to 0.001 %	–	Re + Im: **0.005 %** to 0.98
2D hyperboloidal eigensolver, Assaad and Panosso Macedo 2025	exponential convergence	0, 0.5, 0.9	no Leaver comparison

This work: the larger of the Re and Im relative errors, worst of the five fundamentals, 4×84. The overtones reach 2.3×10^{-3} only at $a/M = 0.9998$; 97 % of all 112 errors are below 0.1 %. The eigensolver needs no initial guess and returns every mode of an m sector at once. Error definitions differ.

Advantages of a standalone PINN

A standalone PINN removes six prerequisites of spectral solvers; they matter on **numerical, rotating, non-separable** backgrounds.

1 · Numerical backgrounds

Grid data enter the residual **point by point**; no global spectral fit.

2 · Any ω dependence

ω is trainable; no polynomial eigenproblem is needed.

3 · Points where needed

Points are redrawn where the field is steep; no matrix to rebuild.

4 · Little pre-processing

A new theory is a new residual; autodiff gives the derivatives.

5 · Inverse problems

Spin or coupling as inputs give a differentiable $\omega(a, \alpha)$.

6 · Higher dimension

A dense spectral solve costs $\sim (N_1 N_2)^3$; relevant in 3D. A prospect.

Advantages 1 and 4 echo [Luna et al. 2024]. None shows on Kerr, where the spectral steps are cheap.

PINN and spectral eigensolver compared

The PINN gains **flexibility**; it gives up the full spectrum, speed and a built-in error estimate.

Question	Spectral eigensolver	This work, PINN	For the PINN
Field representation	Chebyshev basis, tensor grid	network, evaluated **anywhere**	**+ advantage**
Derivatives	differentiation matrices	**automatic differentiation**	**+ advantage**
Numerical background	smooth global fit	coefficients **point by point**	**+ advantage**
How ω enters	polynomially	**any way**; ω is trained	**+ advantage**
Initial guess	**none**	a start, then continuation	**− drawback**
Modes per run	**all of an m sector**	one branch	**− drawback**
Accuracy to $a/M = 0.95$	10^{-5} to 10^{-6}	3.6×10^{-6} (fundamentals)	**comparable**
Cost	one eigensolve per spin	2.6 to 7.5 h per branch on an H100	**− drawback**
Error estimate	**convergence in resolution**	network sizes only	**− drawback**

Accuracy: worst of the five fundamentals, 4×84; the overtones reach 2.3×10^{-3} only at $a/M = 0.9998$.

Kerr as a benchmark

Kerr has exact reference spectra and a hard near-extremal regime; the references stay **out of the computation**.

Exact references

Leaver continued fractions, through the **qnm** package, far more precise than our errors.

Hard regime included

Up to $a/M = 0.9998$: damping goes to zero and a steep near-horizon layer forms.

Same machinery beyond Kerr

The 2D operator never uses separability; only its coefficients would change.

Inside the loop

PDE residual, the accepted PINN history of the branch, field regularity. **Target-free.**

Held out

Reference values: never in the loss, the starts after $a = 0$, selection, stopping or the frequency solve.

Pass criterion

Re and Im both below 1 % at every spin, on all three networks: **336 of 336** pass.

In fairness: the method was developed on these branches, so the result validates the final method, not a blind test. References scored after each 16-spin branch is complete. [Stein 2019; Leaver 1985]

Part 3

Methods

1 Introduction

2 State of the art

3 Methods

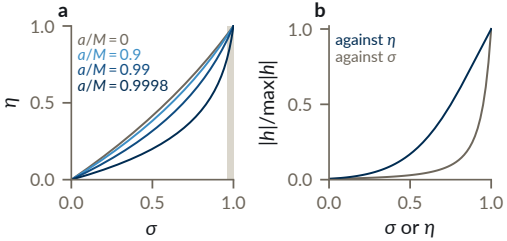
4 Results

5 Conclusions

Unknowns, normalisation and network inputs

$\eta(\sigma)$ is a second radial input that **spreads the near-horizon layer**, so the network sees a gentle field as $a \to M$.

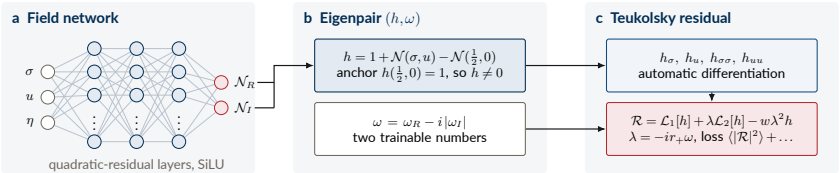
Inputs σ, u
$\sigma = r_+/r$: null infinity at 0, horizon at 1; $u = \cos\theta$.
Third input $\eta(\sigma)$
$\eta = 1 - \ln[1 + (1-\sigma)/\delta]/\ln(1 + 1/\delta)$;
$\delta = (r_+ - r_-)/r_+$ is the width of the near-horizon layer, $\to 0$ as $a \to M$.
Why
A network learns slow, domain-wide changes first; a layer only **4 %** wide is learned last and worst. Measured in η, the same layer spans **21 %** and is easy to fit. The equation is still solved in σ.
Unknowns and anchor
h from two outputs, scale fixed by $h(\frac{1}{2}, 0) = 1$; ω: two trainable numbers.



a: shaded, the layer at $a/M = 0.9998$. b: stored (2,1,0) field, $a/M = 0.9998$, $u = +1$: its 10–90 % rise spans 23 % of σ, 55 % of η.

PINN architecture and network sizes

One network and one residual; the whole calculation is repeated with **three network sizes** as a resolution test.



Loss function

Only $\mathcal{L}_{\rm PDE}$ has unit weight; the other terms are weak numerical aids, and **none uses reference values**.

$$\mathcal{L} = \mathcal{L}_{\rm PDE} + w_{\rm rel}\,\mathcal{L}_{\rm rel} + \lambda_{\rm mode}(e)\,\mathcal{L}_{\rm mode} + \mathbf{1}_{n>0}\left[\lambda_{\rm br}(e)\,\mathcal{L}_{\rm br} + \lambda_{\rm sh}(e)\,\mathcal{L}_{\rm sh}\right] + \lambda_{\omega}(e)\,\mathcal{L}_{\omega} + \lambda_{\nabla}\mathcal{L}_{\nabla}$$

$\mathcal{L}_{\rm PDE} = \langle|\mathcal{R}|^2\rangle$ the Teukolsky equation on the whole square, boundaries included. **The only physical constraint**, with unit weight.

$\mathcal{L}_{\rm rel}$ the residual relative to the size of its own terms, so the horizon, where coefficients vanish, is not ignored; $w_{\rm rel} = 10^{-2}$.

$\mathcal{L}_{\rm mode}$ angular smoothness and coherence of h: keeps the branch on its ℓ.

$\mathcal{L}_{\nabla} = \langle|\nabla\mathcal{R}|^2\rangle$ sharpens the ω dependence near extremality; $\lambda_{\nabla} = 10^{-2}$ for $m \neq 0$, $a/M \geq 0.96$.

$\mathcal{L}_{\rm br}$ **branch**, overtones only: keeps ω near the extrapolation of its own accepted history.

$\mathcal{L}_{\rm sh}$ **shape**, overtones only: penalises a radially smooth, fundamental-like field.

$\mathcal{L}_{\omega}$ when ω is released at a new spin, a quadratic penalty keeps it near the value **predicted from the previous spins** of the branch; weighted like the residual, used only where ω could jump branch.

$\lambda(e)$: weights scheduled over the epochs e. Full definitions in Backup. $\mathcal{L}_{\nabla}$: [Yu et al. 2022]

Initial frequencies at a = 0

At $a = 0$ Kerr **is** Schwarzschild, so each branch starts from the Schwarzschild mode it continues: a physical seed, not a random guess.

Branch	Start $M\omega$	Taken from
(2,0,0), (2,1,0), (2,2,0)	$0.375 - 0.090i$	Schwarzschild $0.3737 - 0.0890i$, deliberately rounded
(2,2,1); (2,2,2)	$0.3467 - 0.2739i$; $0.3011 - 0.4783i$	published Schwarzschild $n = 1, 2$ values, 4 decimals
(3,3,0)	$0.675 - 0.095i$	eikonal formula; also 0.9 and 0.8 of its Re
(4,4,0)	$0.866 - 0.096i$	eikonal formula; same variants

Why Schwarzschild

Every Kerr QNM branch is the spin continuation of one Schwarzschild mode with the same (ℓ, n); at $a = 0$ the m values coincide.

Where the values come from

$\ell = 2$: classic tables. $\ell = 3, 4$: light-ring formula $M\omega \approx [\ell + \frac{1}{2} - i(n + \frac{1}{2})]/3\sqrt{3}$.

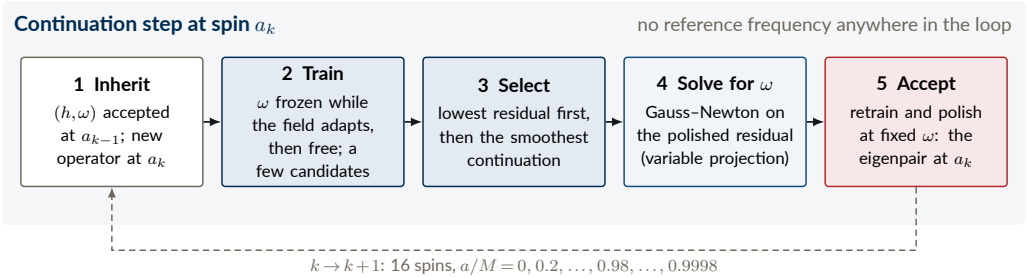
No known spectrum?

Beyond Kerr, start from the **GR limit** and continue in the coupling, as here in spin, or from the light ring of the new background.

Only seeds: the residual finds the eigenvalue. [Chandrasekhar and Detweiler 1975; Leaver 1985; Berti et al. 2009; Cardoso et al. 2009]

Spin continuation

Each branch is carried from Schwarzschild along **16 spins** to $a/M = 0.9998$; training finds the branch and a solve fixes ω.



Select: lowest residual first, then continuity of ω and of the field with the accepted history. A curriculum in PINN terms [Krishnapriyan et al. 2021].

Frequency solve

After training, ω is fixed by **stationarity of the polished residual**, not by where the optimiser stopped.

$\boldsymbol{r}(\Theta, \boldsymbol{\omega})$: real and imaginary parts of $\mathcal{R}$ on a 65×65 mesh; Θ: weights; $\boldsymbol{\omega} = (\omega_R, \omega_I)$.

1. **Polish the field** at fixed $\boldsymbol{\omega}$: $(J_\Theta^\intercal J_\Theta + \mu I)\,\Delta\Theta = -J_\Theta^\intercal \boldsymbol{r}$, giving $\hat{\Theta}(\boldsymbol{\omega})$.
2. **Update the frequency**: $\Delta\boldsymbol{\omega} = -\left[J_\omega^\intercal (I-P) J_\omega\right]^{-1} J_\omega^\intercal (I-P)\,\boldsymbol{r}$.

Jacobians, damping

$J_\Theta = \partial \boldsymbol{r}/\partial\Theta$, $J_\omega = \partial \boldsymbol{r}/\partial\boldsymbol{\omega}$; μ blends Gauss–Newton and gradient steps.

Projector P

Onto what the output layer can absorb; $I - P$ keeps what needs ω to move.

Result

All 336 points converged in at most **32 steps**.

[Levenberg 1944; Marquardt 1963] Variable projection: [Golub and Pereyra 1973].

Part 4

Results

1 Introduction

2 State of the art

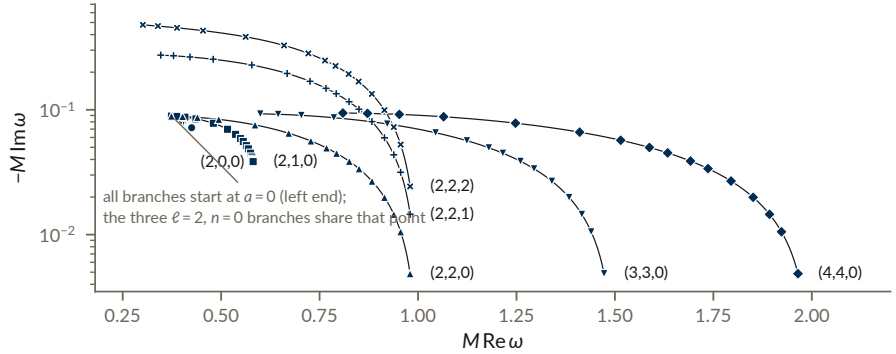
3 Methods

4 Results

5 Conclusions

QNM frequencies in the complex plane

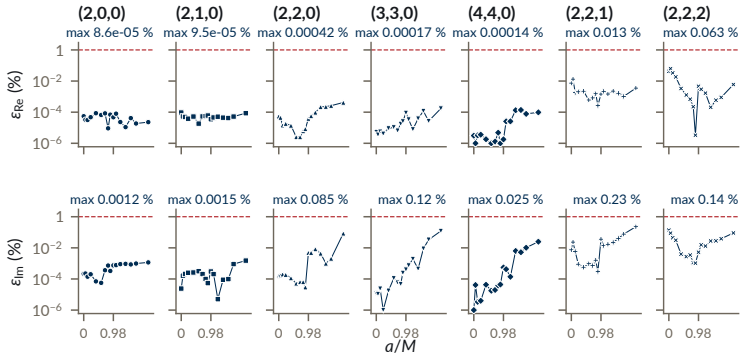
All seven branches follow the reference trajectories from $a = 0$ to $a/M = 0.9998$.



Markers: PINN, 4×84, 16 spins. Lines: Leaver reference.

Relative errors

All 112 errors are below **0.23 %**; the median is $2 \times 10^{-4}\,\%$.

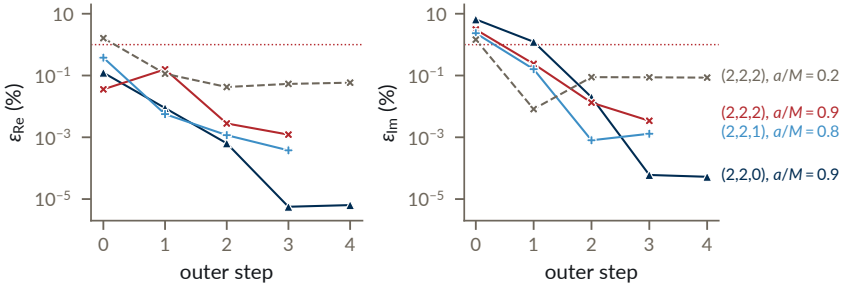


Top: real part; bottom: damping; dashed: 1 %. Damping errors grow near extremality; the largest is the (2,2,1) damping at $a/M = 0.9998$.

Effect of the frequency solve

Where training stopped far from the eigenvalue, up to **6.6 %** off, the frequency solve brings both components **below 0.1 % in two outer steps**.

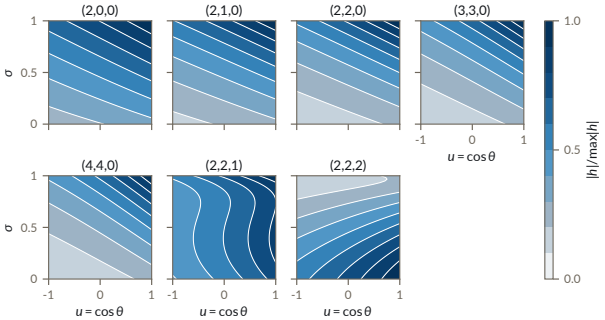
Each outer step: 1. polish the field at fixed ω (Levenberg–Marquardt on the weights); 2. update ω by a Gauss–Newton step on the projected residual. Stop when both components change by less than 10^{-4}.



Step 0: end of training. The four points with the largest starting errors, 4×84; over all 112 points the worst error falls from 6.6 % to 0.23 %. Dotted: 1 %.

Learned fields

The learned $|h|$ is **regular on the whole square** and nearly separable, although no product form is imposed.

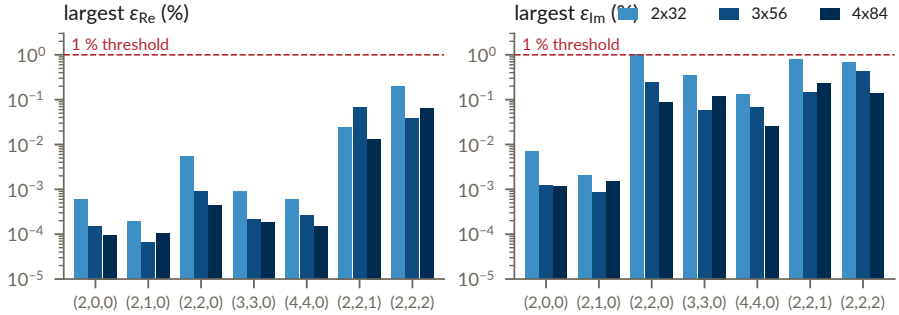


- **No oscillations:** the hyperboloidal height function absorbs the outgoing and ingoing phases, so h is smooth up to $\sigma = 0$ and $\sigma = 1$.
- **Near product form:** Kerr separates, so $h \approx f(\sigma)\, g(u)$, with $s_2/s_1 \sim 10^{-5}$: learned, not imposed.
- **Overtones:** higher eigenstates of the same radial problem, damped about 3 and 5 times faster; the damping enters through $\lambda = -i\, r_+ \omega$, and, loosely as for excited states, the field gains radial variation: (2,2,2) has one radial sign change and peaks at null infinity.

$a/M = 0.8$, 4×84, each panel scaled to its maximum. s_2/s_1: second over first singular value of h.

Network-size comparison

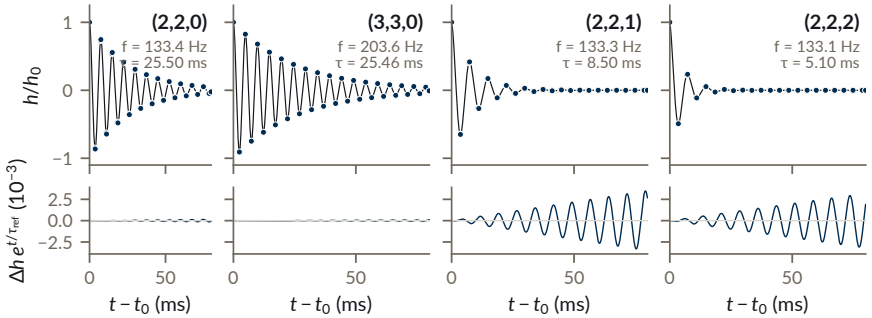
All three network sizes stay below **1 %**; the smallest only just, at 0.997 %.



Largest error over the 16 spins, per branch and network. Worst component: 0.997 % (2×32), 0.427 % (3×56), 0.229 % (4×84).

Ringdown waveforms

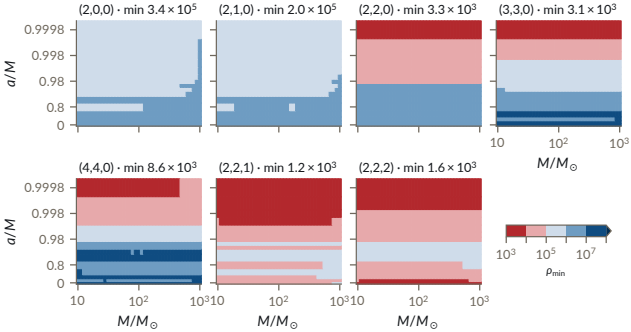
At $M = 200\,M_\odot$, $a/M = 0.98$, PINN and reference ringdowns **coincide by eye**.



Strips: difference over the reference envelope $e^{-t/\tau_{\rm ref}}$. After 80 ms it reaches 3.5×10^{-3} for (2,2,1) and stays below 1.6×10^{-4} for the fundamentals.

Waveform distinguishability

Telling PINN and reference ringdowns apart would need an SNR of at least $\mathbf{1.2 \times 10^3}$.



$\geq \mathbf{1.2 \times 10^3}$

SNR needed: the lowest $\rho_{\min}$, (2,2,1) at $a/M = 0.9998$

$\approx \mathbf{80}$

Combined SNR of GW250114, all detectors, whole signal: 15 times below

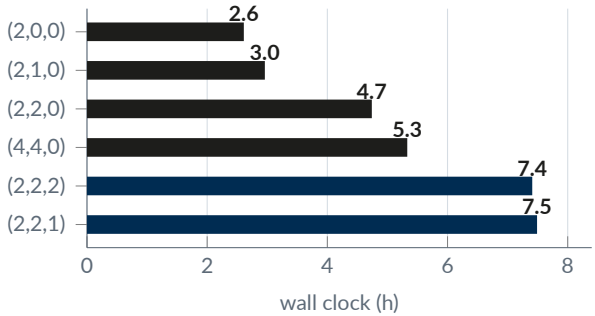
$\approx \mathbf{300}$

Loudest ringdown SNR per year expected for the Einstein Telescope: 4 times below

$\rho_{\min} = \sqrt{D/[2(1-\mathcal{F})]}$, $\mathcal{F}$ the noise-weighted match, aLIGO design noise, $D = 4$. [LVK 2025; Branchesi et al. 2023; Lindblom et al. 2008]

Computational cost

One 16-spin branch takes **2.6 to 7.5 h** on one H100; spins run in sequence, branches in parallel.



Wall clock in hours, 4×84, NVIDIA H100 PCIe; overtones in blue.

151 ms
one training step, 4×84, 4 096 points, Apple M5 CPU

4.3 / 5.5 / 7.4 h
(2,2,2) on one H100 for the three sizes

13.0 h
(3,3,0), on a slower Quadro RTX 6000

Part 5

Conclusions

1 Introduction

2 State of the art

3 Methods

4 Results

5 Conclusions

Summary

A non-spectral solver of the **two-dimensional** Teukolsky eigenproblem works on Kerr up to $a/M = 0.9998$.

0.23 %

Accuracy. Largest of 112 errors; below 1 % on all three sizes.

0.9998

Near extremality. Damping ratios $1:3:5$ of (2,2,0), (2,2,1), (2,2,2) recovered.

$> 10^3$

Waveforms. SNR needed to tell PINN from reference.

Key ingredients

Spin continuation, a **frequency solve** on the polished residual, and the **Cauchy-horizon-fixing gauge** near extremality.

Limitations

The significance is methodological: on smooth Kerr, spectral solvers are **more precise and faster**.

Precision and cost: against spectral	Error estimate: without references	Kerr-specific aids: to adapt
Errors up to 0.23 %.	Only network-size comparisons.	η, the Cauchy gauge and the zero-damping predictors use Kerr's horizons.
Hours of GPU per branch.	Developed on these branches.	
One branch per run.		

Why PINNs

The advantage is **flexibility** where a fixed basis and a linear eigensolve stop being natural.

grid data

Numerical backgrounds.
Coefficients enter point by point.

any ω

No linearisation.
Non-polynomial ω dependence is allowed.

one code

Parameters, inverse problems.
$\omega(a, \alpha)$ from one training.

In short

Not more precise, not faster on Kerr: **fewer analytical assumptions**, which is what beyond-Kerr problems need.

Future work

Next: backgrounds that **do not separate**, in this order.

1 · Non-separable tests: first

Kerr–Newman.
Numerical rotating sGB and EdGB backgrounds.

2 · Networks: second

Spin and coupling as inputs.
Couplings fitted to ringdowns.

3 · Trust: throughout

Error estimates without references.
Branch diagnostics for crowded spectra.

References: physics

Teukolsky 1973, ApJ 185, 635

Leaver 1985, Proc. R. Soc. A 402, 285

Chandrasekhar, Detweiler 1975, Proc. R. Soc. A 344, 441

Kokkotas, Schmidt 1999, Living Rev. Relativ. 2, 2

Berti, Cardoso, Starinets 2009, CQG 26, 163001

Cardoso et al. 2009, PRD 79, 064016

Zenginoğlu 2008, CQG 25, 145002

Ansorg, Panosso Macedo 2016, PRD 93, 124016

Panosso Macedo 2020, CQG 37, 065019

Assaad, Panosso Macedo 2025, CQG 42, 245008

Stein 2019, JOSS 4, 1683

Dias, Godazgar, Santos 2015, PRL 114, 151101

Chung, Wagle, Yunes 2024, PRD 109, 044072

Chung, Yunes 2024, PRL 133, 181401

Blázquez-Salcedo et al. 2024, PRD 109, 064028; 2025, PRD 111, L021505

Lindblom, Owen, Brown 2008, PRD 78, 124020

LIGO–Virgo–KAGRA 2025, PRL 135, 111403

Branchesi et al. 2023, JCAP, Einstein Telescope

References: machine learning and PINNs

Lagaris, Likas, Fotiadis 1998, IEEE TNN 9, 987

Raissi, Perdikaris, Karniadakis 2019, JCP 378, 686

Baydin et al. 2018, JMLR 18, 153

Rahaman et al. 2019, ICML

Wang, Teng, Perdikaris 2021, SIAM J. Sci. Comput. 43, A3055

Krishnapriyan et al. 2021, NeurIPS

Bu, Karpatne 2021, SDM

Yu et al. 2022, CMAME 393, 114823

Levenberg 1944; **Marquardt** 1963

Golub, Pereyra 1973, SIAM J. Numer. Anal. 10, 413

Luna et al. 2023, PRD 107, 064025; 2024, PRD 109, 124064

Pombo, Pizzuti 2026, JCAP 2026, 009

Ferrer-Sánchez et al. 2026, doi:10.5281/zenodo.23656788

Questions

Antonio Ferrer-Sánchez
Antonio.Ferrer-Sanchez@uv.es · www.uv.es
Preprint: arxiv.org/abs/2608.19774

VNIVERSITAT DE VALÈNCIA
Escola Tècnica Superior d'Enginyeria (ETSE-UV)

Backup: overtone loss terms

Both terms act only for $n > 0$ and stop an overtone sliding onto the smoother $(2, 2, 0)$.

$$\mathcal{L}_{\rm br} = \left(\frac{\mathrm{Re}\,\omega - \mathrm{Re}\,\omega_\star}{s_R}\right)^2 + \left(\frac{\mathrm{Im}\,\omega - \mathrm{Im}\,\omega_\star}{s_I}\right)^2, \qquad \mathcal{L}_{\rm sh} = \frac{1}{1 + E_\sigma/s_\sigma} + \frac{\eta_{\sigma\sigma}}{1 + E_{\sigma\sigma}/s_{\sigma\sigma}} + \frac{\eta_C}{1 + C_\sigma/s_C}$$

Branch

$\omega_\star$: extrapolation of the accepted same-branch frequencies; s_R, s_I: their scatter.

Shape

E_σ, $E_{\sigma\sigma}$: radial gradient and curvature energy of h; C_σ: inner–outer contrast.

Weights

$\eta_{\sigma\sigma} = \eta_C = 0.5$, $s_\sigma = s_{\sigma\sigma} = 1$, $s_C = 1/4$; ramped in after a first eigenpair.

Backup: relative-residual scale

$\mathcal{L}_{\rm rel} = \langle |\mathcal{R}|/(\mathcal{S}+\epsilon)\rangle$ compares the residual with the size of the terms that should cancel in it.

$$\mathcal{S} = |\alpha_2 h_{\sigma\sigma}| + |\alpha_1 h_\sigma| + |\alpha_0 h| + |\gamma_2 h_{uu}| + |\gamma_1 h_u| + |\lambda\beta_1 h_\sigma| + |\lambda\beta_0 h| + |w\lambda^2 h|$$

Terms

α_i, γ_i: coefficients of $\mathcal{L}_1$; β_i: of $\mathcal{L}_2$; w: of λ^2.

Why

α_2 vanishes at $\sigma = 0, 1$: a plain residual would ignore the horizon.

Details

$\mathcal{S}$ is detached from training; edge weight $\epsilon_E + \sigma(1-\sigma) + (1-u^2)$; $\epsilon = 10^{-6}$.

Backup: separability of the learned fields

Ratio of the second to the first singular value of $h(\sigma, u)$ at $a/M = 0.8$: **below** 2×10^{-5} for the fundamentals.

Branch	s_2/s_1
(2,0,0)	6.7×10^{-6}
(2,1,0)	2.1×10^{-6}
(2,2,0)	1.9×10^{-5}
(3,3,0)	1.3×10^{-5}
(4,4,0)	1.5×10^{-5}
(2,2,1)	2.2×10^{-4}
(2,2,2)	3.0×10^{-3}

81×81 grid, trained 4×84 networks. The exact Kerr field is separable; the coherence term of $\mathcal{L}_{\rm mode}$ favours it, so this is a consistency check.