Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative Study between Numerical Solver, Analytical Perturbation Theory and Physics-Informed Neural Networks in solving a quartic Anharmonic Oscillator

Mayank Goswami

mayankfph@iitr.ac.in

Indian Institute of Technology Roorkee

Harshwardhan Navadkar

Indian Institute of Technology Roorkee

Research Article

Keywords: Perturbation theory, Rayleigh–Schrödinger series, anharmonic oscillator, Schrödinger equation, Hamiltonian, Physics-Informed Neural Networks (PINNs), Rayleigh quotient, energy eigenvalues, wavefunctions, numerical diagonalization, quantum mechanics.

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11685613/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Comparative Study between Numerical Solver, Analytical Perturbation Theory and Physics-Informed Neural Networks in solving a quartic Anharmonic Oscillator

Harshwardhan Navadkar[1] and Mayank Goswami,[1,2]

[1]Department of Physics, IIT Roorkee
[2]MF School of AI and DS, IIT Roorkee
mayank.goswami@ph.iitr.ac.in

Abstract*—Perturbation theory is a common analytical technique used for estimating the shift in the energy levels of atoms when a perturbation is introduced in the system. However, as will be evident later on, the resulting Rayleigh-Schrodinger series is divergent for every non-zero perturbation λ. In this paper, we build and compare three solvers for the same perturbed Hamiltonian ($H = -\frac{1}{2}\partial_x^2 + \frac{1}{2}x^2 + \lambda x^4$) problem: an exact numerical solver (finite-difference discretization of the time-independent Schrodinger equation followed by Hamiltonian diagonalization, cross-validated against an independent Fock-basis spectral solver), the analytical perturbation series (derived from ladder operators) and a Physics-Informed Neural Network (PINN). Their results are compared as the perturbation is increased from 0 to 1, which lets us locate precisely where perturbation theory stops being useful and also observe how accurate PINNs can be. Third-order perturbation theory is exact at $\lambda = 0$ but is wrong by 18.6 energy units at $\lambda = 1$, an error more than twenty times larger than the quantity it is estimating. Over 80 cold-start trainings (five seeds at eight values of λ) the dedicated PINN reaches mean energy errors of 2.87×10^{-5} (ground state) and 3.62×10^{-5} (first excited state). Further reading the energy from the trained wavefunction through its Rayleigh quotient reduces these to 6.92×10^{-6} and 2.54×10^{-6}. At the points tabulated by Brevi et al. the relative energy error is $28.8 - 1.08 \times 10^3$ times lower for states $n \leq 2$, although sequential training fails for $n \geq 3$. We further construct a parametric PINN in which a single network takes (x, λ) jointly and learns $\psi(x; \lambda)$ and $E(\lambda)$ over $\lambda \in [0, 1]$ in one training run, realizing the lambda-as-input extension proposed as future work by Brevi et al. Its energy network is accurate to 2.46×10^{-3} for $\lambda > 0$, but its wavefunction yields Rayleigh-quotient energies accurate to 2.59×10^{-5}, so the PINN generalizes over λ but gives lesser accurate energy read-out mainly.

I. Introduction

The quantum harmonic oscillator is one of the most widely used systems of quantum mechanics because it is exactly solvable. Its energy spectrum, eigenfunctions and algebraic structure are all known in closed form. But no realistic system is actually fully harmonic. Most natural systems like molecular vibrations and lattice dynamics require corrections to the harmonic quadratic potential, and the simplest and most studied of these is the quartic anharmonic oscillator,

$$H = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 + \lambda x^4, \tag{1}$$

where $\hbar = m = \omega = 1$. The energy of the n^{th} energy level is given by: $E_n = (n + \frac{1}{2}), n \in \mathbb{W}$. For the perturbation strength $\lambda > 0$, we have no such closed-form solution for the energy.

When a perturbation of this kind is switched on, the standard procedure is to estimate the shifted levels with time-independent perturbation theory. The anharmonic oscillator is the textbook application of that method and, (funnily) simultaneously, its textbook counterexample: the Rayleigh-Schrodinger series for this potential has zero radius of convergence and diverges for every $\lambda \neq 0$ [1], [2]. That helps us because the approximation we are trying to compare has a known and quantifiable failure limit.

This report builds three independent solvers for the same Hamiltonian family and compares them as the perturbation strength is increased from zero to the non-perturbative regime: exact numerical solver(s), the analytical perturbation series generated programmatically from ladder operators, and a Physics-Informed Neural Network (PINN) that solves the equation directly by minimizing its residual (using given Physics equations) rather than learning from labelled data. A parametric extension of the network, trained once over a continuous range of perturbative strengths and evaluated at arbitrary unseen values, is developed as an extension. The contributions of this work are: (i) the construction and validation of reliable numerical reference energies, against which the accuracy of the PINN predictions can be assessed (Section VI-A); (ii) a minimal dedicated PINN (three-term loss, 8.5×10^3 parameters, CPU training in 16.18 s per network) whose accuracy is proven over 80 cold-start trainings and whose relative energy error at the points tabulated by Brevi et. al.

Corresponding author: mayank.goswami@ph.iitr.ac.in

[3] is 28.8–1.08×10^3 times lower for $n \leq 2$ (Section VI-H); (iii) a parametric PINN with the perturbative strength λ as an input, including the first excited state, together with a quantified accuracy v/s cost trade-off (Section VI-I); and (iv) the observation that the Rayleigh quotient of the trained network is a more accurate and certifiable energy estimate than the trainable eigenvalue, for both PINN variants (Section III-E).

The remainder of the report states the objectives (Section II), explains the required theory (Section III), reviews previous work (Section IV), describes the method (Section V), presents the results (Section VI), concludes (Section VII) and states the limitations (Section VIII).

II. Problem Objectives and Requirement

A. Objectives

This project tries to answer the quantitative question of where exactly perturbation theory stops being useful and in what perturbative regime do PINNs remain useful, quantifying the errors in prediction of energy in both dedicated and parametric networks. This is achieved through four clear objectives of the project:

1) **Establish a trustworthy ground truth.** Implement two unrelated numerical exact solvers, a real-space finite-difference grid and a Fock-basis spectral diagonalization, and require them to agree before any later comparison between energies and other observables is accepted.
2) **Derive the perturbation series.** Generate the Rayleigh-Schrodinger coefficients programmatically from ladder operators to arbitrary order and for arbitrary state index, so that the divergence can be exhibited directly.
3) **Locate the breakdown.** Map the perturbative error against the exact result across the range of perturbations being studied and identify the optimal truncation order at each perturbative strength.
4) **Build and assess PINN solvers.** Train a dedicated (one-network-per-perturbation) PINN for the ground and first excited states and a parametric PINN that covers the whole perturbative range in a single training run, and report the accuracies of both the PINNs.

B. Need for solving this problem

1) Many times perturbation theory is applied without reporting the region over which the theory would be applicable, beyond which it would become a meaningless estimate. For the quartic anharmonic oscillator, where the series provably diverges, that boundary can be located exactly. This gives us a concrete answer to a question that is usually left to judgement.
2) Strong anharmonicity is seen in low-barrier molecular vibrations and in tunnelling problems. Methods that work only where a small parameter exists are of no use there, so a solver whose accuracy is independent of perturbation strength would have an actual application there.
3) PINNs are being increasingly proposed for problems where no reference solution exists, which makes their claimed accuracy hard to check. The anharmonic oscillator is good test case because: the exact answer is computable to quite some significant figures, the analytical method has a known breakdown and a continuous physical parameter is involved, which is worth investigating.

III. Theory

A. Anharmonic Oscillator Hamiltonian and Conventions

Two different normalizations of the anharmonic Hamiltonian appear in literature, and confusing them leads to a λ-dependent discrepancy that could be mistaken for a numerical bug. Both conventions are explained:

1) Bender-Wu convention: The published eigenvalue tables used for external validation use this convention.

$$H\psi(x) = -\psi''(x) + (x^2 + \lambda x^4)\psi(x) = E\psi(x), \quad (2)$$

for which the unperturbed levels are the odd integers $E_n = 2n + 1$.

2) Natural-units convention: The perturbation theory and PINN parts of this project use Eq. (1), with unperturbed levels $E_n = n + \frac{1}{2}$. Here $\hbar = m = \omega = 1$. All results in Section VI other than Figs. 2- 3 are in this convention.

B. Rayleigh-Schrodinger Perturbation Theory

Writing $V = x^4$ and expanding both the energy and the state in terms of the powers of the perturbative strength, we get:

$$E = \sum_{k\geq0} E^{(k)}\lambda^k, \qquad |\psi\rangle = \sum_{k\geq0} \lambda^k|\psi^{(k)}\rangle, \quad (3)$$

with $|\psi^{(0)}\rangle = |n_0\rangle$ and $E^{(0)} = n_0 + \frac{1}{2}$, the standard non-degenerate recursion relation is:

$$|\psi^{(k)}\rangle = \hat{R}\Big(V|\psi^{(k-1)}\rangle - \sum_{j=1}^{k-1} E^{(j)}|\psi^{(k-j)}\rangle\Big), \quad (4)$$

$$E^{(k)} = \langle n_0|V|\psi^{(k-1)}\rangle, \quad (5)$$

where the reduced resolvent $\hat{R}$ is diagonal in the unperturbed basis is:

$$R_{mm} = \frac{1 - \delta_{mn_0}}{E^{(0)}_{n_0} - E^{(0)}_m}. \quad (6)$$

The exclusion of the $m = n_0$ subspace is what keeps non-degenerate perturbation theory well defined. The perturbation itself is built algebraically using the ladder operators, using $x = (a+a^\dagger)/\sqrt{2}$ where $a_{n-1,n} = \sqrt{n}$, so that x^4 is obtained as a matrix power in the harmonic-oscillator basis without any spatial discretization. The Eqs. (3)- (6) are from [4] with some changes in notation.

C. Divergence of series in quartic Oscillator

Bender and Wu showed that the coefficients $E^{(k)}$ of the quartic oscillator grow factorially with order [1], [2], so that:

$$\limsup_{k\to\infty} |E^{(k)}|^{1/k} = \infty \quad \Longrightarrow \quad R_{\text{conv}} = 0. \tag{7}$$

The series is therefore asymptotic rather than convergent: for a given λ, partial sums improve up to an optimal truncation order $k^*(\lambda)$ and then deteriorate beyond it, with k^* decreasing as λ grows.

D. Physics-Informed Neural Networks for Eigenvalue Problems

A PINN [5], [6] replaces the supervised loss seen in most neural networks with the residual of a differential equation, evaluated at different sampled collocation points and differentiated by automatic differentiation. For Eq. (1) the point-wise residual of a trial function ψ_θ with trial eigenvalue E_θ is:

$$r(x) = -\frac{1}{2}\psi_\theta''(x) + V(x)\psi_\theta(x) - E_\theta\,\psi_\theta(x), \tag{8}$$

with $V(x) = \frac{1}{2}x^2 + \lambda x^4$. Nothing is learned from examples: the governing equation is the training signal, and no reference solution enters the loss at any point.

Two features distinguish an eigenvalue problem from a forward PDE solver. First, the eigenvalue is unknown, so E_θ must itself be a trainable parameter. The network then searches simultaneously for a function and for the scalar that makes that function a solution [3], [7]. Second, $r \equiv 0$ is satisfied exactly by $\psi \equiv 0$ for any E_θ, so the residual must also be supported by a normalization constraint. Excited states lead to a third requirement: the residual plus normalization is minimized equally well by any eigenstate, so an orthogonality constraint against the lower states is what selects the next eigenvalue up rather than a return to the ground state.

Importantly, for the present comparison, the residual of Eq. (8) is enforced at the actual value of λ, with the quartic term having the same importance as the quadratic one. The formulation contains no expansion in terms of the perturbative constant, hence there is no small parameter to violate and no series to diverge.

E. Guarantee of the Residual

A trained network can give two energy estimates: the trainable eigenvalue E_θ and the Rayleigh quotient of the learned wavefunction,

$$E_R = \frac{\langle\psi_\theta, \hat{H}\psi_\theta\rangle}{\langle\psi_\theta, \psi_\theta\rangle}. \tag{9}$$

Since $(\hat{H} - E_R)\psi_\theta$ is orthogonal to ψ_θ, the residual of Eq. (8) splits into two orthogonal parts and,

$$\|r\|^2 = \|(\hat{H} - E_R)\psi_\theta\|^2 + (E_R - E_\theta)^2\,\|\psi_\theta\|^2. \tag{10}$$

Of the loss terms, only the second depends on E_θ. For a uniform collocation grid over $[-6, 6]$, the PDE loss is approximately $\mathcal{L}_{\text{PDE}} \simeq |r|^2/12$. Consequently, near convergence, the dependence of the loss on $E\theta$ is reduced to the term $(E_\theta - E_R)^2/12$. This means that a small PDE residual provides only a linear bound on the energy difference: $|E_\theta - E_R| \le |r|$. In contrast, the Rayleigh quotient satisfies the Kato-Temple inequality, which provides a more informative bound on the energy error. [8],

$$|E_R - E_n| \le \frac{\sigma^2}{\delta}, \qquad \sigma^2 = \frac{\|(\hat{H} - E_R)\psi_\theta\|^2}{\|\psi_\theta\|^2}, \tag{11}$$

where δ is the distance from E_R to the nearest other eigenvalue. The error of E_R is therefore quadratic in the residual. Thus, a small residual provides a strong accuracy guarantee for the Rayleigh quotient E_R, but does not, by itself, guarantee that the learned energy E_θ is accurate.

IV. Previous Work

A. The Anharmonic Oscillator

The quartic anharmonic oscillator has been studied continuously since the late 1960s. Bender and Wu's analysis of the behaviour of the perturbation series [1], [2] discovered the vanishing radius of convergence and is the reference used for the divergence being exploited here. Resummation and multi-instanton techniques later could get usable results from the divergent series [9], at the cost of the calculations being heavier than the original expansion.

B. Neural Solvers for the Schrodinger Equation

Neural-network solvers for differential equations were initially explored by Lagaris et al. [6], who trained feed-forward networks to satisfy ordinary and partial differential equations and used an envelope to enforce boundary conditions by construction. The modern framework, including the automatic-differentiation residual loss used here, was formalized by Raissi et al. [5].

For the anharmonic oscillator specifically, Mutuk [7] calculated energy levels with a feed forward network and tabulated the results over a wide range of perturbative strength λ. Those tables serve as an external validation for our exact solver. Jin, Mattheakis and Protopapas [10] introduced the normalization and orthogonality losses used for PINN eigenvalue problems, including Eqs. (17)–(18) here. Brevi, Mandarino and Prati [3] applied PINNs to the same Hamiltonian convention as Eq. 1 for $\lambda = 0.005 \cdot 2^n$ up to $\lambda \approx 10$ and states up to $n = 5$, using a 7×256 wavefunction network and a separate 4×256 energy network ($\sim 6 \times 10^5$ parameters) trained with a seven-term weighted loss. Each λ is trained separately by transfer learning from a neighbouring value, and the reported relative energy errors against Lagrange-mesh results reach 2.3×10^{-3} (ground state: 2.7–5.8×10^{-4}). They identify a network taking (x, λ) as input as future work. The dedicated PINN here differs in using a three-term loss with the boundary condition built into the trial function, a two-stage Adam/L-BFGS optimiser and a network roughly 70 times smaller. The parametric PINN implements the

extension they propose. Brevi et al. do not report run-to-run variability; Section VI-H compares the two approaches point by point at their tabulated (n, λ).

C. Operator Learning

A parallel research track works on the related problem of learning solution maps between function spaces rather than individual solutions: DeepONet [11] and the Fourier Neural Operator [12] both learn operators that generalize across input functions. Our parametric PINN shares this goal of achieving many solves after only one training run, but differs in that it is trained purely on the physics residual (PINN way) and requires no dataset of solved instances. The closest precedent for this parametric architecture is the work of Mattheakis et al. [13], who made the interatomic distance an input of a PINN for H_2^+ and represented the energy with a separate network depending only on that parameter. We adopt the same separation for $E_\phi(\lambda)$, but apply it to an anharmonic perturbation across the breakdown of perturbation theory and extend it to an excited state. In our system it is this separate energy network, rather than the wavefunction network, that limits accuracy (Section VI-F).

D. Gaps Addressed Here

This project is motivated by four gaps in the previous research. First, most PINN studies of this system report accuracy against a single numerical reference without an independent cross-check or a convergence study of that reference. This is important because even published eigenvalue tables disagree with their own quoted cross-references at the 10^{-3} level for excited states, so a comparison based on one such table cannot resolve errors at that scale. Second, the PINN literature on this system trains a separate network per λ value and has not reported what a single network covering the whole range costs in accuracy, nor validated such a network's excited state by node count, fidelity and orthogonality. Third, previous PINN studies of this system do not report run-to-run variability across random seeds or computational cost against a conventional solver. Fourth, the trainable eigenvalue is reported as the PINN energy without a residual-based error bound.

V. Method

A. Exact Numerical Solvers

1) Finite-difference grid diagonalization: The spatial coordinate is discretized on a uniform grid $x_i \in [-L, L]$ with N points and spacing Δx. The second derivative is approximated by the three-point stencil

$$\psi''(x_i) \simeq \frac{\psi_{i+1} - 2\psi_i + \psi_{i-1}}{\Delta x^2}, \tag{12}$$

which turns Eq. (1) into a symmetric tridiagonal matrix eigenvalue problem $\mathbf{H}\boldsymbol{\psi} = E\boldsymbol{\psi}$ with

$$H_{ii} = \frac{1}{\Delta x^2} + \frac{1}{2}x_i^2 + \lambda x_i^4, \quad H_{i,i\pm1} = -\frac{1}{2\Delta x^2}. \tag{13}$$

We solve it with SciPy's `eigh_tridiagonal` function, outputting only the lowest few eigenpairs using its `select_range` argument rather than diagonalizing the full spectrum. Eigenvectors are rescaled by $1/\sqrt{\Delta x}$ to convert their discrete normalization $\sum_i |\psi_i|^2 = 1$ into a continuum normalization $\int |\psi(x)|^2 dx = 1$, and the overall sign returned by the eigensolver is fixed by requiring $\psi > 0$ at its largest-magnitude point. Unless stated otherwise we use $L = 6$ and $N = 2000$. The Bender-Wu validation script uses the same tridiagonal solver (it previously used a dense `eigh`; see Section VI-I).

2) Fock-basis spectral diagonalization: The second method doesn't discretize space. Using the ladder construction of Section III, x^4 is built as a matrix in the truncated harmonic-oscillator basis $\{|0\rangle, \ldots, |N_b - 1\rangle\}$, added to the diagonal H_0 with entries $n + \frac{1}{2}$ and $H = H_0 + \lambda x^4$ is diagonalized directly with $N_b = 200$. To avoid a truncation error in the last rows of the matrix, x^4 is built in an $(N_b + 4)$-dimensional basis and then truncated to $N_b \times N_b$. With $N_b = 200$ the six lowest levels are converged below 10^{-8} up to $\lambda = 10.24$ (Section VI-A), and these Fock-basis energies are the reference against which both PINNs are scored.

3) Cross-validation protocol: The two methods are very different in everything, so their agreement acts as a correctness check. We compare the three lowest eigenvalues in $\lambda \in \{0, 0.05, 0.1, 0.3, 0.5, 1.0\}$ and require $\max_n |E_n^{\text{grid}} - E_n^{\text{Fock}}| < 10^{-3}$ before any result obtained later on is trusted. Because the three-point stencil has the leading error $-\frac{\Delta x^2}{24}\langle p^4 \rangle$, which always lies below the true energy, the grid energies are also Richardson-extrapolated from grids with N and $2N$ points to remove the $O(\Delta x^2)$ term. We additionally require the extrapolated energies to agree with the Fock basis to 10^{-7} and check the raw grid error against the predicted $-\frac{\Delta x^2}{24}\langle p^4 \rangle$. The grid solver, run in the Bender-Wu convention, is additionally validated (raw and Richardson-extrapolated) against the published tables of Mutuk [7] at $\lambda \in \{0.025, 0.05, 0.1, 0.5, 1.0, 4.0\}$ for the ground state and at $\lambda = 0.1$ for the first four states. Where that reference's own tabulated column and the independent cross-references it reproduces disagree, we validate against the mutually consistent independent values.

4) Convergence protocol: Convergence of the reference is tested separately in each discretization parameter at $\lambda \in \{0, 0.05, 0.3, 1, 2, 5, 10.24\}$ for the states $n = 0, \ldots, 5$: the Fock-basis size $N_b \in [20, 300]$ against $N_b = 400$; the number of grid points $N \in [250, 8000]$ at $L = 6$ against $N_b = 400$, from which the observed order of convergence is extracted; and the domain half-width $L \in [3, 10]$, Richardson-extrapolating at each L so that only the domain-truncation error remains, again against $N_b = 400$. The mapping $2E_{\text{nat}}(\lambda) = E_{\text{BW}}(2\lambda)$ between the two conventions is checked numerically.

B. Perturbation Series Implementation

The recursion relation of Eqs. (4)- (6) is implemented directly, with the ladder algebra truncated at $N_b = 250$

basis states, far above the orders and state indices used here. Because the coefficients are generated, the same code path produces any order and any state index, and the implementation is validated by checking that it reproduces the known ground-state coefficients $\{\frac{1}{2}, \frac{3}{4}, -\frac{21}{8}, \frac{333}{16}\}$ exactly. Third-order truncation is used as the analytical baseline throughout.

C. Dedicated PINN: One Network per Perturbative strength

1) Architecture: Each eigenstate is represented by a fully connected network with three hidden layers of 64 units and tanh activation functions, which map $x \mapsto \psi(x)$. Two decisions concerning the structure of the PINN have been taken:

- **Decay of the wavefunction is hard-coded:** The network output is multiplied by a Gaussian envelope,

$$\psi_\theta(x) = e^{-x^2/2}\, f_\theta(x), \tag{14}$$

  which builds the boundary condition $\psi \to 0$ as $x \to \pm\infty$ into the function space instead of adding a boundary loss term and relying on the network to learn it. The inner layers are free to make the output decay faster than the envelope where the true solution decays faster, as it does at large λ values.
- **Trainable eigenvalue:** The energy is a scalar `nn.Parameter` optimized along with the weights, which is what makes the model an eigenvalue solver rather than a PDE solver at a guessed E.
- **Activation function selection:** tanh is used rather than ReLU because the loss involves ψ'', and the second derivative of a ReLU network vanishes almost everywhere, which would make the residual meaningless. The network has 8514 trainable parameters including the energy.

2) Loss function: With $N_c = 1000$ collocation points uniformly spaced on $[-6, 6]$, the total loss for n^{th} state is calculated as:

$$\mathcal{L} = \mathcal{L}_{\text{PDE}} + w_{\text{norm}}\mathcal{L}_{\text{norm}} + w_\perp \mathcal{L}_\perp, \tag{15}$$

with $w_{\text{norm}} = 1$ (varied in Section VI-E5). Writing the discrete inner product as:

$$\langle f, g\rangle = \sum_i f(x_i) g(x_i) \Delta x,$$

and using the residual $r(x)$ of Eq. (8), the three terms are

$$\mathcal{L}_{\text{PDE}} = \frac{1}{N_c}\sum_i r(x_i)^2, \tag{16}$$

$$\mathcal{L}_{\text{norm}} = \big(\langle\psi_\theta, \psi_\theta\rangle - 1\big)^2, \tag{17}$$

$$\mathcal{L}_\perp = \sum_{m<n} \big\langle \psi_\theta, \psi^{(m)}\big\rangle^2, \tag{18}$$

where ψ'' is obtained by applying `torch.autograd.grad` twice with `create_graph=True`. The normalization term is what prevents the trivial solution $\psi \equiv 0$, which satisfies the residual exactly.

3) Excited states via orthogonality: The first excited state is found by training a second network using the ground-state solution as a reference and setting its energy parameter initially to $E = 1.5$. A penalty with $w_\perp = 5$ is added to discourage the new solution from overlapping with the ground state. Without this penalty, the network could simply return to the ground state again. The penalty helps it find the next higher energy state instead.

The penalty does not single out ψ_1. Every eigenstate ψ_k with $k \geq 1$ is orthogonal to ψ_0 and is therefore a zero-loss minimum of Eq. (15), and the optimizer reaches the minimum whose basin contains the initialization. Nor does it exclude a possibility of collapse onto ψ_0 for every weight. Along the path $\psi = \sqrt{c}\,\psi_0 + \sqrt{1-c}\,\psi_1$ with E_θ following E_R, Eq. (10) gives $\mathcal{L}(c) = c(1-c)(E_1 - E_0)^2/12 + w_\perp c$, so the collapsed state $c = 1$ is a local minimum whenever $w_\perp < (E_1 - E_0)^2/12$. This threshold grows with λ and reaches 0.31 at $\lambda = 1$. $w_\perp = 5$ is the smallest round value that exceeds it by more than an order of magnitude at every $\lambda \in [0, 1]$ and larger values stiffen the optimization without any benefits (Section VI-E5). Each excited state is checked by its node count, its fidelity against the exact ψ_0, ψ_1 and ψ_2, and by $|E_\theta - E_R|$.

4) Higher states: For the comparison with Brevi et al. (Section VI-H), states $0, \ldots, n$ are trained one after another from cold starts, each with the orthogonality penalty against all lower networks. For excited states with $n \geq 2$, the initial energy estimate is obtained from the WKB (Bohr–Sommerfeld) quantization condition, $\oint p\, dx = 2\pi(n + \frac{1}{2})$ and the number of Adam optimization steps is doubled.

5) Two-stage optimization: Training the PINN happens in two phases. The Adam optimiser (learning rate 2×10^{-3}, 1500 epochs) is robust to the large, noisy gradients of an untrained network but eventually plateaus rather than converging sharply. L-BFGS with a strong Wolfe line search (`max_iter`= 500, `tolerance_grad`= 10^{-6}) then converges to the solution using curvature information. Adam alone would need far more iterations to match its precision. Across the λ sweep each network is warm-started from the previous λ's weights. Warm starting is not required for convergence (Section VI-E1) and, with a fixed Adam budget, does not reduce wall time (13.59 s against 16.18 s per network for warm and cold starts, Section VI-I).

D. Parametric PINN: One Network for all Perturbative strengths

1) Architecture: The parametric model replaces $\psi_\theta(x)$ with $\psi_\theta(x, \lambda)$ and the scalar E_θ with a function $E_\phi(\lambda)$, learned by two sub-networks:

$$\psi_\theta(x, \lambda) = e^{-x^2/2}\, f_\theta(x, \lambda), \qquad f_\theta : \mathbb{R}^2 \to \mathbb{R}, \tag{19}$$

$$E_\phi(\lambda) = g_\phi(\lambda), \qquad g_\phi : \mathbb{R} \to \mathbb{R}. \tag{20}$$

Here f_θ has three hidden layers of 64 units and g_ϕ has two hidden layers of 32 units, giving 9730 parameters in total. The energy is kept in a separate network that never

sees x. This is to prevent E from depending on x during training. Splitting the networks makes that impossible by construction rather than preventing by penalizing it.

2) Block-structured batching: The normalization integral and the residual must both be evaluated per λ, so training cannot use scattered random (x, λ) pairs. Each step uses $N_\lambda = 20$ perturbative strengths and evaluates the full $N_x = 250$-point spatial grid for each, i.e. an $(N_\lambda \times N_x)$ block, so that

$$\mathcal{L}_{\text{norm}} = \frac{1}{N_\lambda} \sum_j \big(\langle \psi_\theta(\cdot, \lambda_j), \psi_\theta(\cdot, \lambda_j)\rangle - 1\big)^2 \tag{21}$$

is well defined for every λ_j in the batch.

3) Training schedule: During the Adam phase (5000 epochs, learning rate 2×10^{-3}) a fresh batch of perturbative strengths is drawn uniformly from $[0, 1]$ at every epoch. This randomization forces generalization across the continuous range instead of memorization of a fixed set and is what makes evaluation at an unseen λ meaningful. The subsequent L-BFGS phase (`max_iter`= 400, `tolerance_grad`= 10^{-8}) switches to a fixed set of 20 equally spaced perturbative strengths. This is because a line search needs the loss function to stay consistent each time it is evaluated. Using a different randomly selected batch of data each time would change the loss, making it difficult to determine the best step to take. The excited-state network uses the same orthogonality penalty as before, now imposed per λ. The loss has the same three terms and weights as Eq. (15), evaluated per λ_j and averaged over the block. Besides $E_\phi(\lambda)$, the energy is also read out as the Rayleigh quotient of $\psi_\theta(\cdot, \lambda)$ (Section III-E).

E. Training and Test Data

Training data (`data/train/collocation_points.npz`) contains the inputs that actually enter the loss: the spatial collocation grid (1000 points on $[-6, 6]$), the eight fixed perturbative strengths used by the dedicated solver, and the $[0, 1]$ range sampled by the parametric solver. No targets are stored, because none exist. The loss is actually the physics residual at those points.

Testing data (`data/test/exact_benchmark_energies.npz` and the same table as `.csv`) hold the ground and first excited state energies from the converged Fock-basis solver ($N_b = 200$), with the grid values alongside for comparison, at the same eight perturbative strengths. These values never enter any training loss and are used only afterwards for scoring, making them a PINN equivalent of testing dataset.

F. Evaluation Metrics

1) Energy error: The primary metric is the absolute eigenvalue error $|E_n^{\text{PINN}} - E_n^{\text{exact}}|$ at each lambda, reported alongside the corresponding third-order perturbative error. It is reported for both the trainable eigenvalue E_θ and the Rayleigh quotient E_R of Eq. (9). The Rayleigh quotient is evaluated with automatic differentiation on a 2000-point grid that differs from the training grid, together with its Kato-Temple bound σ^2/δ (Eq. (11)).

2) Wavefunction fidelity: Energy agreement alone is a weak test, since a model can reach close to a eigenvalue with a visibly wrong state during unstable training. We therefore also compute the fidelity of the eigenstate:

$$F_n = \big|\langle \psi_n^{\text{PINN}}, \psi_n^{\text{exact}}\rangle\big|^2, \tag{22}$$

after renormalizing the network output and fixing its sign at the largest-magnitude point. The fidelity is computed against the exact ψ_0, ψ_1 and ψ_2, and the index of the largest fidelity identifies which eigenstate was reached.

3) Excited-state checks: The number of nodes is counted as the number of sign changes of ψ where $|\psi| > 10^{-3} \max|\psi|$. The overlap $\langle \psi_0^{\text{PINN}}, \psi_1^{\text{PINN}}\rangle$ is reported both on the training quadrature, as the penalty sees it, and on the evaluation grid.

4) Seed robustness: Every stochastic step is seeded. To distinguish a reliable result from a lucky initialization, both networks are retrained from scratch (no warm start) over five seeds at all eight values of λ (three seeds for the parametric PINN, which is trained once per seed), and we report the mean, standard deviation and maximum of the energy error.

G. Computational Environment

All reported runs used Intel(R) Core(TM) Ultra 5 225H (14 physical cores, 14 PyTorch threads), 15.4 GB RAM, Windows 11, Python 3.14.3, PyTorch 2.14.0+cpu (device: cpu, float32 networks), NumPy 2.4.4 and SciPy 1.17.1. The three scripts were run one after another with no other heavy load. Timings are wall-clock medians with interquartile ranges from repeated measurements, and memory is the peak resident set size of the process.

VI. Results

A. Validation of the Exact Solvers

Table I reports the agreement between the two independent exact methods over the three lowest states. The observed discrepancy, 1.5×10^{-5} to 9.2×10^{-5}, is more than an order of magnitude better than the 10^{-3} acceptance threshold. The discrepancy grows with λ because it is dominated by the finite-difference truncation error of the grid method, which grows with λ as the wavefunction is squeezed and the discretization has to resolve sharper curvature. Richardson extrapolation removes it. The extrapolated grid energies agree with the Fock basis to 1.42×10^{-10} (last column of Table I), and the raw discrepancy matches the predicted $-\frac{\Delta x^2}{24}\langle p^4\rangle$ to within 6.06×10^{-6} relative.

Fock basis: Table II gives the smallest N_b for which all of $E_0, \ldots, E_5$ are within 10^{-8} of $N_b = 400$. The required basis grows with λ and reaches 200 at $\lambda = 10.24$. The plateau of about 10^{-10} beyond that is floating-point round-off, which grows with $\|H\|$.

Grid resolution: Over $N \in [250, 8000]$ the observed order of convergence is 1.999–2.006 at every λ up to 10.24 (Fig. 1, centre), and Richardson extrapolation agrees with

TABLE I
CROSS-VALIDATION OF THE TWO INDEPENDENT EXACT SOLVERS USING MAXIMUM ABSOLUTE EIGENVALUE DIFFERENCE OVER THE THREE LOWEST STATES. GRID ($L = 6$, $N = 2000$) AND RICHARDSON-EXTRAPOLATED GRID (N, $2N$) VERSUS FOCK BASIS (F, $N_b = 200$).

λ	$\max_n \lvert E_n^{\text{grid}} - E_n^{\text{F}} \rvert$	$\max_n \lvert E_n^{\text{Rich.}} - E_n^{\text{F}} \rvert$	Status
0.00	1.46×10^{-5}	2.87×10^{-11}	OK
0.05	2.25×10^{-5}	3.16×10^{-11}	OK
0.10	2.86×10^{-5}	2.31×10^{-11}	OK
0.30	4.72×10^{-5}	5.75×10^{-11}	OK
0.50	6.19×10^{-5}	8.72×10^{-11}	OK
1.00	9.21×10^{-5}	1.42×10^{-10}	OK

TABLE II
SMALLEST FOCK-BASIS SIZE N_b FOR WHICH $|E_n(N_b) - E_n(400)| < 10^{-8}$ FOR ALL $n = 0, \ldots, 5$.

λ	0	0.05	0.3	1	2	5	10.24
N_b	20	30	50	75	100	150	200

the Fock basis to 1.42×10^{-10} for $\lambda \leq 1$ and 1.13×10^{-9} at $\lambda = 10.24$.

Domain size: After Richardson extrapolation at each L the remaining error is the domain truncation (Table III). At $\lambda = 0$, $L = 4$ still gives errors of 4.83×10^{-7} (E_0) and 1.44×10^{-5} (E_1), while $L = 5$ gives 3.57×10^{-9}; for $\lambda \geq 1$ even $L = 3$ is converged. The domain therefore matters at small λ. At large λ the wavefunction narrows, the grid points per wavefunction width fall from 117.8 to 59.84 (ground state, $\lambda = 0$ to 10.24), and the $N = 2000$ grid error of E_0 grows from 1.13×10^{-6} to 1.66×10^{-5}: resolution, not domain size, limits the grid there.

External benchmarks and conventions. The Richardson-extrapolated Bender-Wu ground-state energies agree with the published values [7] to 8.23×10^{-11} (raw grid: 3.38×10^{-5}), and $2E_{\text{nat}}(\lambda) = E_{\text{BW}}(2\lambda)$ holds to 3.90×10^{-11}.

B. Reference Spectrum

Table IV lists the reference energies used throughout. Figure 2 shows the first eight levels as functions of the perturbative strength in the Bender-Wu convention. Every level increases monotonically, with the spacing between neighbouring levels increasing as λ grows, because of greater confinement of the particle in the potential due to the quartic term. Figure 3 shows the three lowest eigenfunctions at $\lambda = 1$, vertically shifted by their corresponding energy eigenvalues. The nodal structure is preserved, with the n^{th} excited state having n nodes, while the wavefunctions become more localized than their harmonic oscillator counterparts due to the stronger confinement because of the quartic term.

C. Perturbation Theory and Its Divergence

The ladder-operator method reproduces the known ground-state coefficients exactly. It returns $\{0.5, 0.75, -2.625, 20.8125\}$ for orders 0 to 3, i.e. $\{\frac{1}{2}, \frac{3}{4}, -\frac{21}{8}, \frac{333}{16}\}$, matching the closed-form result to better than 10^{-6}. Extending it beyond the textbook range exposes the factorial growth directly (Table V). By sixth order the coefficient becomes -6.4×10^4 for the ground state and -2.0×10^6 for the first excited state.

Table VI shows the asymptotic character by showing partial sums against the exact ground-state energy. At $\lambda = 0.05$ the series is still well behaved: every added order improves the estimate, down to 4.9×10^{-4} at sixth order. At $\lambda = 0.1$ the optimal truncation order has already dropped to roughly three, after which further terms degrade the answer. At $\lambda = 0.3$ the best estimate is first order, and at $\lambda = 0.5$ the sixth-order partial sum is -900, an error of nearly three orders of magnitude. This is the meaning of zero radius of convergence [1], [2].

TABLE III
DOMAIN-SIZE CONVERGENCE: $\max(|\Delta E_0|, |\Delta E_1|)$ OF THE RICHARDSON-EXTRAPOLATED GRID AT HALF-WIDTH L AGAINST THE FOCK BASIS ($N_b = 400$).

L	$\lambda = 0$	$\lambda = 0.3$	$\lambda = 1$	$\lambda = 10.24$
3	6.02×10^{-3}	2.00×10^{-6}	4.41×10^{-11}	5.52×10^{-10}
4	1.44×10^{-5}	2.63×10^{-11}	1.37×10^{-11}	5.62×10^{-10}
5	3.57×10^{-9}	1.50×10^{-11}	9.76×10^{-12}	5.60×10^{-10}
6	6.96×10^{-12}	1.74×10^{-11}	2.97×10^{-11}	5.50×10^{-10}
8	1.97×10^{-11}	5.45×10^{-12}	2.80×10^{-11}	5.72×10^{-10}

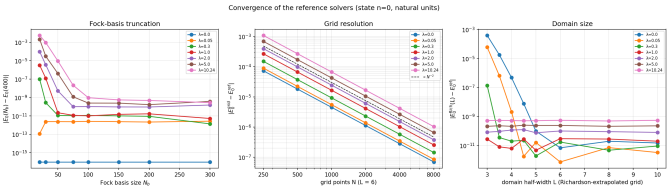


Fig. 1. Convergence of the reference solvers for E_0: Fock-basis truncation (left), grid resolution with the N^{-2} slope (centre), and domain size after Richardson extrapolation (right), for λ from 0 to 10.24.

D. Where Perturbation Theory Fails

Table VII shows the failure for the third-order truncation used as the baseline. At $\lambda = 0$ the series is exact, as it must be. The error then grows by roughly an order of magnitude for every increment of 0.1 in the perturbative strength over the early range: 9.0×10^{-4} at $\lambda = 0.05$, 1.0×10^{-2} at $\lambda = 0.1$, 0.41 at $\lambda = 0.3$. By $\lambda = 1$ the third-order estimate is 19.44 against an exact value of 0.80, the approximation exceeds the quantity it approximates by a factor of 24. Third-order perturbation theory usually retains three-digit accuracy only for $\lambda \lesssim 0.05$ and loses all meaning by $\lambda \approx 0.3$.

E. Dedicated PINN

The dedicated solver trains one network per λ and per state. Table VIII reports its energy errors for both read-outs, the Kato-Temple bounds and the fidelities of

TABLE IV
CONVERGED REFERENCE ENERGIES (FOCK BASIS, $N_b = 200$) IN THE CONVENTION OF EQ. (1), AT THE EIGHT PERTURBATIVE STRENGTHS FORMING THE TESTING DATASET.

λ	E_0	E_1	E_2
0.00	0.500000	1.500000	2.500000
0.05	0.532643	1.653436	2.873980
0.10	0.559146	1.769503	3.138624
0.20	0.602405	1.950544	3.536299
0.30	0.637992	2.094642	3.844783
0.50	0.696176	2.324406	4.327525
0.75	0.754708	2.550725	4.795769
1.00	0.803771	2.737892	5.179292

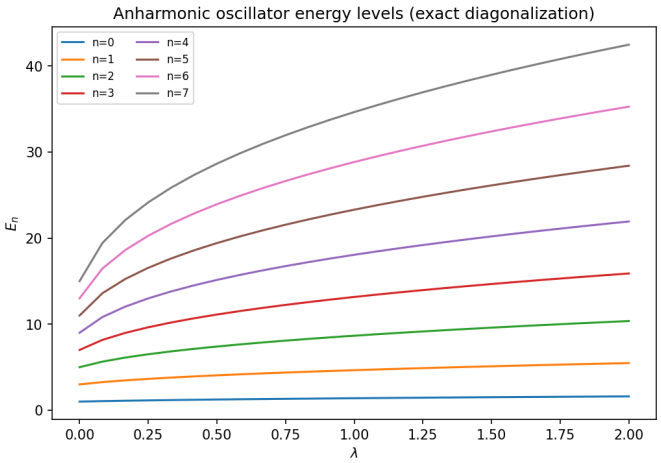


Fig. 2. Energy levels $E_n(\lambda)$ for $n = 0, \ldots, 7$ from exact diagonalization in the Bender-Wu convention of Eq. (2), where $E_n(0) = 2n+1$. Higher states are more strongly affected by the quartic term, since they sample larger $|x|$ where x^4 dominates x^2.

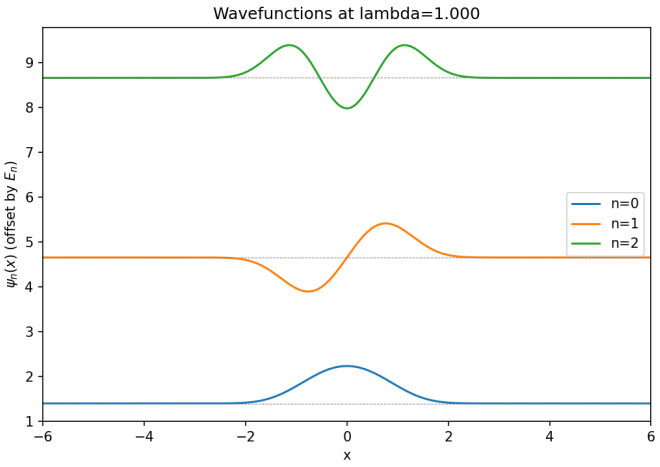


Fig. 3. The three lowest eigenfunctions at $\lambda = 1$, each offset vertically by its own energy (dashed lines). The n^{th} state retains exactly n nodes, and all three decay rapidly outside $|x| \lesssim 2$.

TABLE V
RAYLEIGH-SCHRODINGER COEFFICIENTS $E^{(k)}$ OF THE SERIES $E = \sum_k E^{(k)}\lambda^k$, GENERATED FROM THE LADDER-OPERATOR RECURSION RELATION OF EQS. (4)-(5). THE FACTORIAL GROWTH INDICATES A DIVERGENT ASYMPTOTIC SERIES.

Order k	$E^{(k)}$, $n = 0$	$E^{(k)}$, $n = 1$
0	0.5	1.5
1	0.75	3.75
2	-2.625	-20.625
3	20.8125	244.6875
4	-241.289	-4066.29
5	3580.98	83220.6
6	-63982.8	-1.979×10^6

both states against the converged reference. With the Rayleigh quotient the error is at most 5.90×10^{-6} (E_0) and 3.31×10^{-6} (E_1) over the eight values of λ; the trainable eigenvalue is less reliable, reaching 1.75×10^{-5} (Figure 4, Section VI-E5). The infidelity $1-F$ stays below 1.77×10^{-6} for both states, confirming that the correct energy is reached with the correct state. The first excited state, obtained purely through the orthogonality penalty, has one node at every λ, $1-F_1 \leq 5.65\times10^{-7}$, and an overlap with the ground-state network of at most 7.00×10^{-6} on both the training quadrature and the evaluation grid (Table IX, Fig. 5).

1) Seed robustness: Retraining both networks from scratch over five seeds at all eight values of λ, with no warm start, gives the statistics of Table X (80 networks). 80 of the 80 networks reached the intended state with the correct node count. The spread across seeds at a fixed λ is comparable to the spread across λ (Fig. 6), so seed variability dominates the scatter, not λ . The reported accuracy is therefore a property of the method and warm starting is not needed for accuracy.

2) Training dynamics: Figure 7 shows the ground-state loss history of the dedicated network at each of the eight lambdas, with $\mathcal{L}_{\text{PDE}}$, $w_{\text{norm}}\mathcal{L}_{\text{norm}}$ and $w_\perp\mathcal{L}_\perp$ shown separately, and Fig. 8 the same for the first excited state. Every curve has two segments showing the two-stage optimizer. The first 1500 points are Adam epochs and the remaining are L-BFGS closure evaluations. The horizontal axis counts forward passes, not optimizer iterations, since each L-BFGS step may evaluate the closure several times during its line search. The L-BFGS segment is not of fixed length: it stops at `max_iter` = 500 or when the change in loss between iterations falls below its tolerance, so its length varies between lambdas.

Unlike the parametric case, the collocation grid here is fixed and the whole grid is used at every step. The loss is therefore a deterministic function of the weights and any residual oscillation near the end of the Adam segment is Adam's own step-size behaviour at learning rate 2×10^{-3}.

Each perturbative strength is warm-started from the previous one, but the Adam optimizer state is reset at each perturbative strength, so a short re-adaptation transient is expected at the start of every panel. At convergence the loss is dominated by $\mathcal{L}_{\text{PDE}}$ (5.34×10^{-7} to 1.07×10^{-5}), with $\mathcal{L}_{\text{norm}} \leq 1.93\times10^{-9}$ and $w_\perp\mathcal{L}_\perp \leq 2.44\times10^{-10}$. Lowering the L-BFGS gradient tolerance from 10^{-6} to 10^{-9} changes the energy errors by at most 0 (Section VI-E5), so L-BFGS terminates on its loss-change criterion rather

TABLE VI

PARTIAL SUMS OF THE GROUND-STATE PERTURBATION SERIES AGAINST THE EXACT ENERGY, SHOWING THE OPTIMAL TRUNCATION ORDER (BOLDED) BEYOND WHICH ADDITIONAL TERMS MAKE THE ESTIMATE WORSE.

Order	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.3$	$\lambda = 0.5$
0	0.500000	0.500000	0.500000	0.500000
1	0.537500	0.575000	**0.725000**	**0.875000**
2	0.530937	0.548750	0.488750	0.218750
3	0.533539	**0.569562**	1.050687	2.820312
4	0.532031	0.545434	−0.903754	−12.26025
5	0.533150	0.581243	7.798029	99.645386
6	**0.532150**	0.517261	−38.84544	−900.0861
Exact	0.532643	0.559146	0.637992	0.696176

TABLE VII

THIRD-ORDER PERTURBATION THEORY AGAINST THE EXACT GROUND-STATE ENERGY. THE FINAL COLUMN IS THE ABSOLUTE ERROR, PLOTTED ON A LOGARITHMIC SCALE IN FIG. 11.

λ	E_0^{PT3}	E_0^{exact}	Abs. error
0.00	0.50000	0.500000	0
0.05	0.53354	0.532643	8.96×10^{-4}
0.10	0.56956	0.559146	1.04×10^{-2}
0.20	0.71150	0.602405	1.09×10^{-1}
0.30	1.05069	0.637992	4.13×10^{-1}
0.50	2.82031	0.696176	2.12
0.75	8.36621	0.754708	7.61
1.00	19.43750	0.803771	1.86×10^{1}

than on the gradient tolerance.

3) Sensitivity to architecture and collocation density: Table XI varies the width (32, 64, 128), the depth (2, 3, 4 hidden layers) and the collocation density ($N_c = 250$, 500, 1000, 2000) at $\lambda = 0.1$ and $\lambda = 1$, with three cold-start seeds per configuration (2.2×10^3 to 3.3×10^4 parameters; Fig. 9, left). The mean error stays between 2.98×10^{-6} and 2.93×10^{-4}, with $1 - F \leq 1.13 \times 10^{-6}$ in every configuration. No configuration is systematically better than the 3×64, $N_c = 1000$ baseline; the clearest degradation is $N_c = 500$ at $\lambda = 0.1$ (2.93×10^{-4}).

4) Local minima of the orthogonality loss: Three tests probe the landscape described in Section V-C3. (i) Below the collapse threshold the excited-state network can fall onto ψ_0: at $\lambda = 1$, where $(E_1 - E_0)^2/12 = 0.31$, $w_\perp = 0$ and 0.1 collapse onto ψ_0 ($1 - F_1 \approx 1$), while every $w_\perp \geq 0.3$ reaches ψ_1. At $\lambda = 0$, where the threshold is 0.083, every value of $w_\perp$, including 0, reaches ψ_1. (ii) The penalty also admits higher eigenstates: initializing E at the WKB estimate of E_2 at $\lambda = 1$ converges to $E = 5.1793$ with 2 nodes, i.e. to ψ_2 ($E_2 = 5.179292$, Table IV). (iii) For $n \geq 3$ in the comparison of Section VI-H the sequential construction fails differently: at $n = 3$, $\lambda = 1.28$ the network has $1 - F = 0.109358$ and 5 nodes, while its trainable eigenvalue is within 0.0268713 (relative) of the reference and its Rayleigh quotient within 0.0903762. The mismatch between E_θ and E_R flags such a failure without any reference data; for the $n \leq 1$ networks of the main

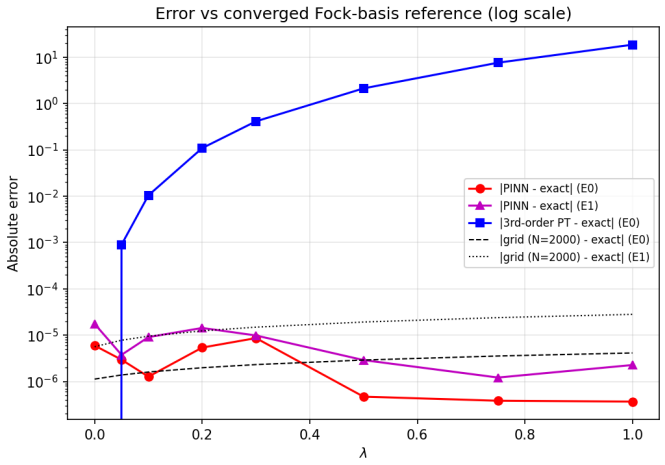


Fig. 4. Dedicated PINN absolute energy error (trainable eigenvalue, E_0 and E_1) against the converged Fock-basis reference on a logarithmic scale, with third-order perturbation theory and the $N = 2000$ grid error for comparison. The Perturbation theory error grows fast, by four orders of magnitude while the PINN error stays below 8.64×10^{-6}.

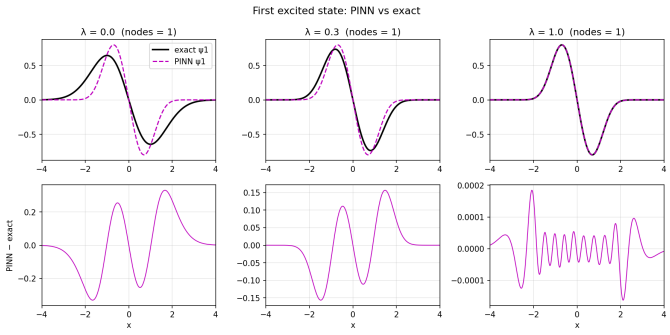


Fig. 5. Dedicated PINN against exact ψ_1 at $\lambda = 0$, 0.3 and 1 (top) and their difference (bottom). The network has one node at every λ.

sweep $|E_\theta - E_R| \leq 1.51787 \times 10^{-5}$ at every λ.

5) Loss components and weights: The final loss components differ by orders of magnitude (Section VI-E2): the normalization and orthogonality constraints are satisfied far below the residual, so $\mathcal{L}_{\mathrm{PDE}}$ alone sets the final accuracy. Table XII and Fig. 9 (right) vary $w_\perp$ for the excited state at $\lambda = 0$ and $\lambda = 1$. Above the collapse threshold the final $\mathcal{L}_{\mathrm{PDE}}$ at $\lambda = 1$ grows steadily with $w_\perp$ (2.89481×10^{-6} at $w_\perp = 0.3$, 4.40834×10^{-6} at 5, 2.58809×10^{-5} at 100), so larger weights stiffen the problem without improving the energy. Varying $w_{\mathrm{norm}} \in \{0.1, 1, 10, 100\}$ for the ground state (Table XIII) changes $|\Delta E_0|$ from 2.38419×10^{-7} to 7.85887×10^{-4} at $\lambda = 0$ and keeps it between 2.09627×10^{-6} and 2.86203×10^{-5} at $\lambda = 1$.

The Kato-Temple bound of Eq. (11) holds for E_R in 16 of the 16 cases of Table VIII, but the same bound is met by E_θ in only 14. This is not an optimizer-tolerance artifact: lowering `tolerance_grad` from 10^{-6} to 10^{-9} changes the number of L-BFGS closure evaluations by at most 0 and the energies by at most the amount given in Section VI-E2. It is the weak coupling of E_θ to the loss derived in

TABLE VIII

DEDICATED PINN (WARM-STARTED SWEEP) AGAINST THE CONVERGED FOCK-BASIS REFERENCE (TABLE IV). $|\Delta E_n|$ IS GIVEN FOR THE TRAINABLE EIGENVALUE E_θ AND FOR THE RAYLEIGH QUOTIENT E_R; KT_n IS THE KATO-TEMPLE BOUND σ^2/δ OF EQ. (11) ON $|E_R - E_n|$; $1 - F_n$ IS THE INFIDELITY OF STATE n.

λ	E_0^{exact}	$\lvert\Delta E_0\rvert$, E_θ	$\lvert\Delta E_0\rvert$, E_R	KT_0	E_1^{exact}	$\lvert\Delta E_1\rvert$, E_θ	$\lvert\Delta E_1\rvert$, E_R	KT_1	$1 - F_0$	$1 - F_1$
0.00	0.500000	5.96046e-06	5.90384e-06	2.57584e-05	1.500000	1.75238e-05	2.34511e-06	1.14096e-05	1.76662e-06	5.64780e-07
0.05	0.532643	2.95327e-06	5.37966e-07	5.72399e-06	1.653436	3.75363e-06	3.31212e-06	4.26096e-05	5.53278e-08	3.82586e-07
0.10	0.559146	1.26918e-06	1.26923e-06	1.72782e-05	1.769503	9.06408e-06	2.36295e-06	3.31837e-05	8.93272e-08	3.44434e-07
0.20	0.602405	5.39120e-06	5.37988e-06	9.00241e-05	1.950544	1.43023e-05	9.53632e-07	2.31586e-05	2.81026e-07	5.86821e-08
0.30	0.637992	8.63984e-06	2.02783e-07	4.91374e-06	2.094642	9.83695e-06	7.89771e-07	1.99341e-05	6.82804e-09	2.40160e-08
0.50	0.696176	4.69741e-07	4.71608e-07	1.10544e-05	2.324406	2.89434e-06	1.47050e-06	4.25741e-05	1.31619e-08	3.29039e-08
0.75	0.754708	3.83852e-07	4.02879e-07	1.03203e-05	2.550725	1.21583e-06	1.32145e-06	4.44804e-05	9.71756e-09	2.57128e-08
1.00	0.803771	3.67746e-07	3.98289e-07	1.23541e-05	2.737892	2.26705e-06	1.69381e-06	6.65707e-05	7.77396e-09	2.78971e-08

TABLE IX

EXCITED-STATE VALIDATION OF THE DEDICATED PINN: NODE COUNTS OF (ψ_0, ψ_1) AND THE OVERLAP OF THE TWO NETWORKS ON THE EVALUATION GRID AND ON THE TRAINING QUADRATURE.

λ	Nodes	$\langle\psi_0\lvert\psi_1\rangle$ (eval.)	$\langle\psi_0\lvert\psi_1\rangle$ (train)
0.00	(0,1)	1.29243e-06	1.30297e-06
0.05	(0,1)	2.02737e-06	2.01803e-06
0.10	(0,1)	-1.65592e-07	-1.48915e-07
0.20	(0,1)	-6.99548e-06	-6.98610e-06
0.30	(0,1)	-4.37061e-07	-4.54289e-07
0.50	(0,1)	4.63861e-06	4.64312e-06
0.75	(0,1)	2.27236e-07	2.33228e-07
1.00	(0,1)	-3.61860e-06	-3.61758e-06

TABLE X

COLD-START SEED STUDY OF THE DEDICATED PINN: ABSOLUTE ENERGY ERROR OVER FIVE SEEDS AND EIGHT VALUES OF λ FOR THE TRAINABLE EIGENVALUE AND THE RAYLEIGH QUOTIENT.

State	Read-out	Mean	Std	Max
E_0	E_θ	2.86893e-05	4.38102e-05	2.17634e-04
E_0	E_R	6.92032e-06	1.32696e-05	7.77130e-05
E_1	E_θ	3.61587e-05	5.72695e-05	2.93671e-04
E_1	E_R	2.53554e-06	4.07462e-06	2.53420e-05

Section III-E.

F. Parametric PINN

The parametric network covers the entire range $\lambda \in [0, 1]$ after a single training run. Figure 10 shows its predictions for both states against the exact curves and against third-order perturbation theory. The PINN curves are almost indistinguishable from the exact ones across the whole range, while the perturbative curve diverges around $\lambda \approx 0.2$ and leaves the plotting area entirely by $\lambda \approx 0.9$.

Figure 11 is a quantitative comparison on a logarithmic scale. The PINN outperforms perturbation theory starting from $\lambda = 0.05$ itself. Before that, perturbation theory is more accurate, being exact at $\lambda = 0$. Above it, the perturbative error climbs by four orders of magnitude, while the error of the energy network stays between 2.25×10^{-4} and 3.48×10^{-3} for $\lambda > 0$ (0.008232 at $\lambda = 0$), with no systematic growth with λ. Over the dense sweep (Fig. 13, lower panel) this error varies smoothly with λ, the signature of a systematic fitting error of E_ϕ rather than noise. The Rayleigh quotient of the same $\psi_\theta(x, \lambda)$ gives 7.27×10^{-6} to 7.71×10^{-5} for $\lambda > 0$. Over three seeds the $\lambda > 0$ mean errors are 0.00246138 (E_ϕ) and 2.58749e-05 (E_R) for E_0, and 0.00484641 and 0.000102061 for E_1 (Table XIV).

24 of the 24 evaluations (3 seeds × 8 values of λ) reach the intended states with node counts $(0, 1)$. The excited state has $1 - F_1 \leq 1.19827 \times 10^{-4}$ for $\lambda > 0$ (3.7074×10^{-3} at $\lambda = 0$), and $|\langle\psi_0|\psi_1\rangle| \leq 0.00174738$.

1) Training dynamics: Figure 12 shows the loss history for both parametric networks. The persistent noise band

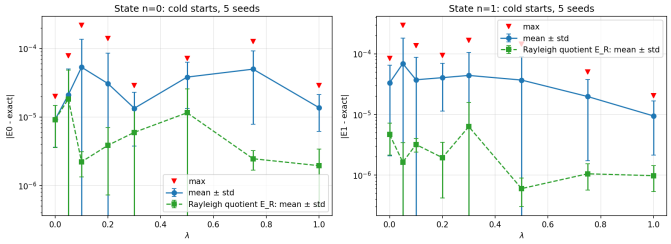


Fig. 6. Cold-start seed study: mean ± standard deviation and maximum of $|E_n - E_n^{\mathrm{exact}}|$ over five seeds at each λ, with the Rayleigh-quotient error for comparison; ground state (left) and first excited state (right).

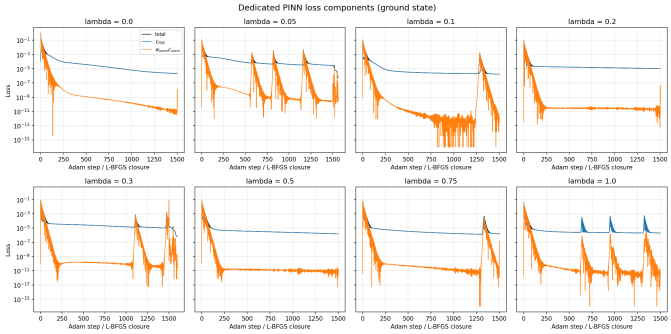


Fig. 7. Dedicated PINN ground-state training loss at each of the eight perturbative strengths (logarithmic scale), with the loss components $\mathcal{L}_{\mathrm{PDE}}$, $w_{\mathrm{norm}}\mathcal{L}_{\mathrm{norm}}$ and $w_\perp\mathcal{L}_\perp$ shown separately. The horizontal axis counts forward passes: the first 1500 are Adam epochs, and the remainder are L-BFGS closure evaluations. Each λ is warm-started from the previous one, with the Adam state reset at the start of each panel. The L-BFGS segment ends once the change in loss falls below the L-BFGS tolerance, so its length varies between lambdas.

TABLE XI
MEAN ABSOLUTE ENERGY ERROR OVER THREE COLD-START SEEDS FOR DIFFERENT WIDTHS, DEPTHS AND COLLOCATION DENSITIES N_c.

		$\lvert\Delta E_0\rvert$		$\lvert\Delta E_1\rvert$	
Configuration	Params	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 0.1$	$\lambda = 1$
3×64, $N_c = 1000$	8514	7.38162e-05	1.62424e-05	5.40827e-05	1.04528e-05
Width 32	2210	6.49452e-06	9.39429e-06	2.98162e-06	3.69618e-06
Width 128	33410	5.08568e-05	6.92426e-06	2.61144e-04	1.42674e-05
Depth 2	4354	7.52351e-05	5.29506e-06	1.57342e-05	4.80742e-06
Depth 4	12674	1.33024e-04	1.23284e-05	7.62522e-06	1.25190e-05
$N_c = 250$	8514	2.10364e-06	8.68473e-05	5.96799e-05	2.74572e-05
$N_c = 500$	8514	5.08387e-06	1.65866e-05	2.92777e-04	7.19298e-06
$N_c = 2000$	8514	4.28571e-06	1.04739e-05	1.23964e-05	4.09216e-06

TABLE XII
ORTHOGONALITY-WEIGHT SWEEP FOR THE EXCITED STATE (SEED 0): EIGENSTATE REACHED, ENERGY ERROR AND FINAL $\mathcal{L}_{\text{PDE}}$.

	$\lambda = 0$		$\lambda = 1$		
$w_\perp$	State reached	$\lvert\Delta E_1\rvert$	State reached	$\lvert\Delta E_1\rvert$	$\mathcal{L}_{\text{PDE}}$
0	1	2.62260e-06	0	1.93410	8.49236e-06
0.1	1	1.66893e-06	0	1.93411	3.28899e-05
0.3	1	1.43051e-06	1	1.15612e-05	2.89481e-06
1	1	7.39098e-06	1	2.32437e-05	6.77324e-06
5	1	1.75238e-05	1	9.41962e-06	4.40834e-06
20	1	2.29120e-04	1	9.89646e-06	6.38489e-06
100	1	2.24113e-05	1	2.44400e-05	2.58809e-05

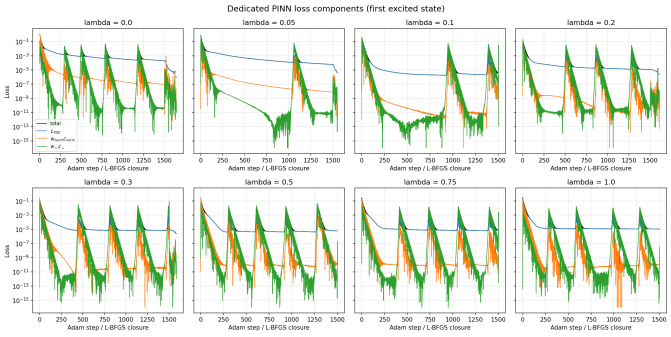


Fig. 8. Same as Fig. 7, for the first excited state.

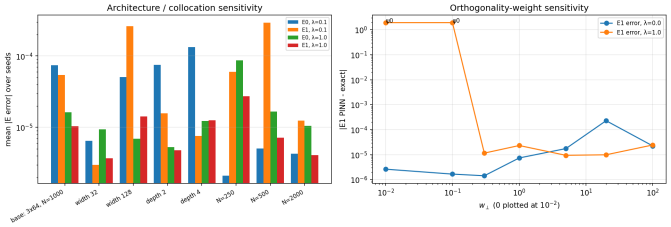


Fig. 9. Left: mean energy error over three seeds against architecture and collocation density. Right: excited-state energy error against the orthogonality weight $w_\perp$ at $\lambda = 0$ and $\lambda = 1$.

during the Adam phase is not instability but the stochastic resampling of perturbative strengths at every epoch. Each step sees a different 20-point subset of the continuous range, so the loss is a different random estimate each time. The excited-state curve stays consistently above the ground-state curve. The final L-BFGS phase, visible as the smooth descending tail after epoch 5000, reduces the loss by roughly an order of magnitude in a few hundred closure evaluations.

TABLE XIII
NORMALIZATION-WEIGHT SWEEP FOR THE GROUND STATE (SEED 0).

w_{norm}	$\lvert\Delta E_0\rvert$, $\lambda = 0$	$\lvert\Delta E_0\rvert$, $\lambda = 1$
0.1	2.38419e-07	2.09627e-06
1	5.96046e-06	2.86203e-05
10	8.24630e-05	2.49329e-06
100	7.85887e-04	2.51037e-05

TABLE XIV
PARAMETRIC PINN SEED STUDY: ABSOLUTE ENERGY ERROR OVER THREE SEEDS AND THE SEVEN VALUES $\lambda > 0$, FOR THE ENERGY NETWORK E_ϕ AND THE RAYLEIGH QUOTIENT E_R OF $\psi_\theta(x, \lambda)$.

State	Read-out	Mean	Std	Max
E_0	E_ϕ	0.00246138	0.00119189	0.00435713
E_0	E_R	0.0000258749	0.0000235881	0.0000770669
E_1	E_ϕ	0.00484641	0.00350618	0.0125001
E_1	E_R	0.000102061	0.000112439	0.000469976

2) Generalization to unseen perturbative strengths: Figure 13 shows the operational payoff. A 60-point sweep which would require 60 separate trainings with the dedicated approach is produced from the trained Parametric PINN network in 0.9961 ms of inference after a single 744.249 s training run. The curve tracks the exact result across the full range, with the visible deviations concentrated at the endpoints $\lambda = 0$ and $\lambda = 1$, the boundary

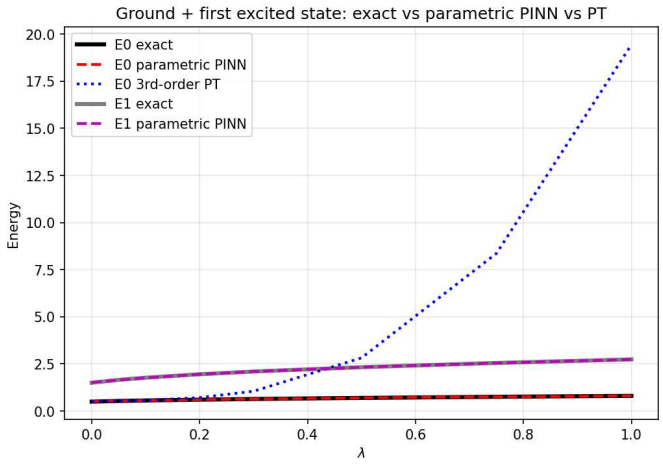


Fig. 10. Ground and first excited state energies: exact diagonalization, Parametric PINN, and third-order perturbation theory.

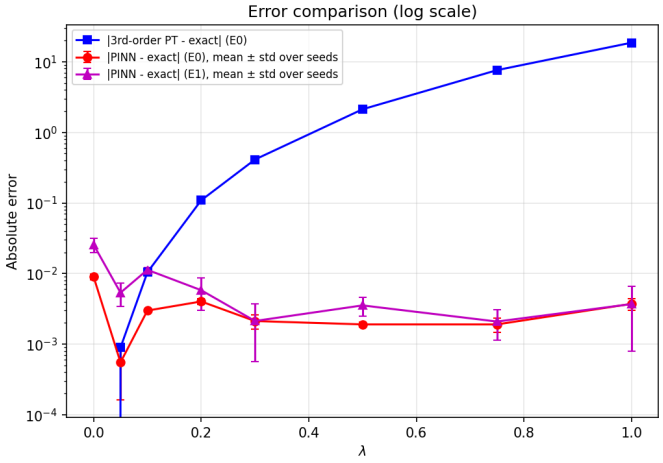


Fig. 11. Parametric PINN absolute energy error of the energy network E_ϕ (mean ± standard deviation over three seeds) for both states, with third-order perturbation theory, on a logarithmic scale. Perturbation theory error grows by four orders of magnitude while the PINN error stays between 2.25×10^{-4} and 3.48×10^{-3}.

of the sampling interval where the network has training signal on one side only. Mean and maximum absolute errors of E_ϕ over the dense sweep are 0.00195784 and 0.008232 for E_0, and 0.00340656 and 0.017974 for E_1.

G. Dedicated versus Parametric

The trainable eigenvalue of the dedicated PINN is roughly 85.8 times more accurate than the parametric energy network, but reading the energy from the wavefunction through the Rayleigh quotient narrows the gap to a factor of about 3.74 for $\lambda > 0$: the accuracy cost of generality lies mainly in the energy network, not in $\psi_\theta(x, \lambda)$. Table XV summarizes the comparison. The Parametric PINN answers a different and harder question. Representing a single eigenpair $(\psi(x), E)$ is a one-dimensional fit but representing the family $(\psi(x, \lambda), E(\lambda))$ is a two-dimensional fit constrained by the same residual, from a

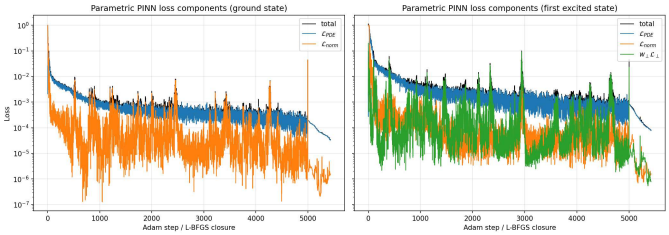


Fig. 12. Parametric PINN training loss for both states (logarithmic scale), with the components $\mathcal{L}_{\mathrm{PDE}}$, $w_{\mathrm{norm}}\mathcal{L}_{\mathrm{norm}}$ and $w_\perp \mathcal{L}_\perp$ shown separately. The stochastic resampling of lambdas each epoch produces the visible noise band during the Adam phase. The smooth tail after epoch 5000 is the L-BFGS phase on a fixed set of lambdas, which decreases the loss by a further order of magnitude.

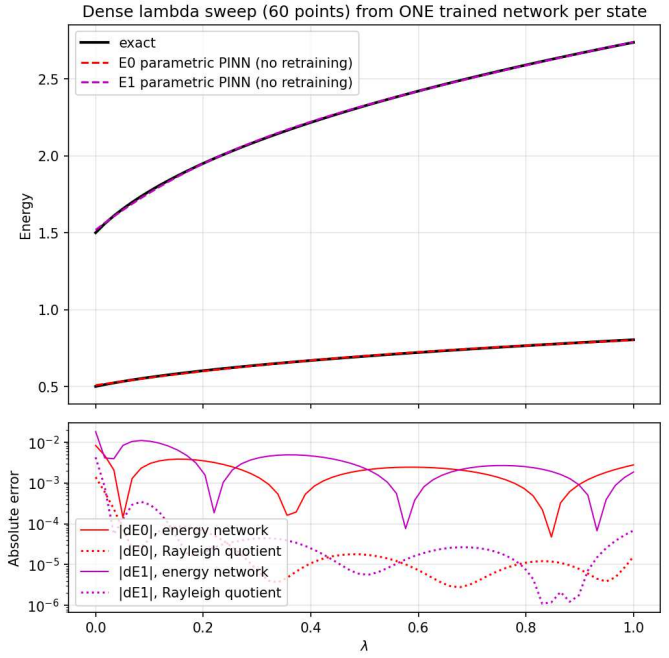


Fig. 13. A 60-point sweep in λ evaluated from the single trained network with no retraining. Top: energies of both states against the exact curves. Bottom: absolute error of the energy network E_ϕ (solid) and of the Rayleigh quotient of $\psi_\theta(x; \lambda)$ (dotted) for both states. The largest deviations occur at the endpoints of the training range.

comparable number of parameters, with the additional requirement that normalization and residual hold separately at every perturbative strength. The accuracy gap should be considered a result of the differences in the problem being solved.

H. Comparison with previous PINN results

Table XVI compares the two set-ups. For a point-by-point comparison, Table XVII retrains the dedicated PINN at the (n, λ) values tabulated in Table IV of [3], in the same convention, with states $0, \ldots, n$ trained sequentially from cold starts (Section V-C4) and scored against the Fock-basis energies. For $n \leq 2$ the relative energy error is 28.8–1.08×10^3 times lower than in [3] (192–1.78×10^5 times with E_R), and $1 - F$ is 1.30–45.5 times lower, with about 70 times fewer parameters and a median of 18.71 s

TABLE XV
THE TRADE-OFF BETWEEN THE TWO PINN DESIGNS.

	Dedicated	Parametric
Network input	x	(x, λ)
Parameters (per state)	8,514	9,730
Trainings for N lambdas	N	1
Cost at a new λ	full retrain	one forward pass
Time at a new λ	27.19 s	98.374 μs
Typical $\lvert\Delta E_0\rvert$, E_θ / E_ϕ	2.86893e-05	0.00246138 ($\lambda > 0$)
Typical $\lvert\Delta E_0\rvert$, E_R	6.92032e-06	2.58749e-05 ($\lambda > 0$)
Collocation points	1000 in x	250×20 in (x, λ)

of CPU training per state against 20–40 min on a GPU. For $n = 3$ and $n = 5$ the sequential construction fails: the networks have the wrong node counts (5, 4 and 8 instead of 3, 5 and 5) and $1 - F = 0.109$–0.930 (Section VI-E4), whereas [3] succeeds with a symmetry loss and a far larger network.

I. Computational Cost

Table XVIII lists the cost of every solver. The parametric network repays its training cost after $N^* = T_{\text{param}}/(t_{E_0} + t_{E_1}) \approx 27.4$ values of λ relative to the dedicated PINN. Against the exact solver: training costs as much as 239030 exact solves, even though one query of $E(\lambda)$ is 32.5 times faster than an exact solve. Its importance lies in being a differentiable approximation to $E(\lambda)$ and $\psi(x; \lambda)$, especially in problems where a conventional solver is computationally expensive.

VII. Conclusion

The question posed in Section II has been answered quantitatively and all four objectives were met. The ground truth was established by two numerically independent exact solvers, converged to 1.13×10^{-9} in basis size, grid resolution and domain size for λ up to 10.24 (Section VI-A), the perturbation series was derived programmatically and validated against the closed-form coefficients, the divergence of the series was observed and studied and both PINN variants were built and scored against the held-out exact energies (the testing dataset equivalents of PINN).

Third-order perturbation theory is exact at no perturbation, is overtaken by the neural solver at $\lambda = 0.05$, has lost three significant figures by $\lambda = 0.1$, and by $\lambda = 1$ returns an estimate twenty-four times larger than the true ground-state energy. The divergence is the expected asymptotic behaviour of Perturbation theory. Partial sums at $\lambda = 0.5$ oscillate to -900 by sixth order and the optimal truncation order falls from six or more at $\lambda = 0.05$ to one at $\lambda = 0.5$. The PINN error, by contrast, shows no systematic growth with λ, because its loss enforces the Schrodinger equation at the actual perturbative strength and contains no expansion in λ to diverge. At the points tabulated in [3] the dedicated PINN's relative energy error is 28.8–1.08×10^3 times lower for $n \leq 2$, with a network about seventy times smaller, but sequential training fails for $n \geq 3$. Across 80 cold-start trainings the Rayleigh quotient of the trained network is more accurate than the trainable eigenvalue (6.92×10^{-6} against 2.87×10^{-5} mean error for E_0), and it is the read-out certified by the residual.

The parametric extension shows that the parameter dependence itself can be learned. One training run yields $\psi(x; \lambda)$ and $E(\lambda)$ over the whole range, and a 60-point sweep that would cost 60 trainings using dedicated PINN costs 0.9961 ms of inference after a single 744.249 s training run instead. That generality costs precision in the energy network (0.00246138 against 2.86893×10^{-5} mean E_0 error for $\lambda > 0$), but much less in the wavefunction, whose Rayleigh quotient gives 2.58749×10^{-5}. In one dimension the exact solver remains cheaper than either PINN (Section VI-I). The trade-off, rather than accuracy, should determine which design is appropriate for a given application.

The problem posed is therefore solved for the ground and first excited states of the one-dimensional quartic oscillator over $\lambda \in [0, 1]$.

VIII. Limitations

1) **Stationary states only:** The states computed here are solutions of the *time-independent Schrödinger equation*. To study how a wavefunction evolves with time, we would need to solve the time-dependent equation $i\,\partial_t\psi = H\psi$, starting from a specified initial state. This requires a different approach, since the wavefunction is complex and the network must predict both its real and imaginary parts. This could be explored in future works as a separate extension, as it differs quite significantly from the current architecture.
2) **Higher states:** Only $n = 0$ and $n = 1$ are solved across the full λ sweep. States up to $n = 2$ are reliable at the points of Section VI-H, but for $n \geq 3$ the sequential construction converges to functions with the wrong node count and low fidelity, consistent with a network that satisfies the residual only at the fixed collocation points (Section VI-E4). Resampling the collocation points every epoch, adding a parity constraint as in [3], or solving for all states simultaneously are the natural remedies.
3) **One parameter per spatial dimension:** The parametric PINN generalizes naturally to several perturbative strengths, $\psi(x; \lambda_1, \lambda_2, \dots)$, at the cost of one input dimension per parameter. Its effect on the PINN accuracy could be tested in future works. Extension to higher spatial dimensions also involves the usual collocation-sampling cost.
4) **Increasing accuracy of parametric PINN:** The main limitation is the energy network: reading E from ψ_θ through the Rayleigh quotient removes most of the gap to the dedicated PINN (Section VI-F). The error of approximately 0.00246138 of the energy

TABLE XVI
COMPARISON OF THE DEDICATED AND PARAMETRIC PINNS DEVELOPED IN THIS WORK WITH THE APPROACH OF BREVI ET AL. [3].

	Brevi et al. [3]	**This work**	
		Dedicated	**Parametric**
λ as network input	No (per-λ transfer learning)	No	Yes
States / λ range	$n \leq 5$ / 0.005–10.24	$n \leq 1$ / 0–1 ($n \leq 5$ / 0.005–10.24 in Table XVII)	$n \leq 1$ / 0–1
Loss terms	7 (weighted)	3	3 (per λ)
Parameters	$\approx 5.9 \times 10^5$ (estimated)	8,514	9,730
Optimiser	Adam	Adam + L-BFGS	Adam + L-BFGS
Max relative E error	2.3×10^{-3} (ground state 2.7–5.8×10^{-4})	1.35×10^{-5} (E_θ), 1.18×10^{-5} (E_R)	0.016464 (E_ϕ), 0.00282207 (E_R)
Training time (hardware)	NA	16.1773 s per network (CPU)	744.249 s per seed, both states (CPU)
Multi-seed study	No	Yes (5 seeds, all λ, both states)	Yes (3 seeds)

TABLE XVII
POINT-BY-POINT COMPARISON WITH BREVI ET AL. [3] (THEIR TABLE IV) AT THE SAME (n, λ): RELATIVE ENERGY ERROR OF THE TRAINABLE EIGENVALUE AND OF THE RAYLEIGH QUOTIENT, INFIDELITY AND NODE COUNT. ROWS WITH $n \geq 3$ ARE FAILURES OF THE SEQUENTIAL CONSTRUCTION (SECTION VI-E4).

		This work, |rel. err.|		Brevi et al.	$1 - F$		
n	λ	E_θ	E_R	|rel. err.|	This work	Brevi et al.	Nodes
0	0.005	2.14845e-06	2.16252e-06	5.78×10^{-4}	1.87949e-07	3.0×10^{-6}	0
0	10.24	9.50332e-06	1.42771e-06	2.74×10^{-4}	1.99962e-08	2.6×10^{-8}	0
2	0.16	2.11337e-06	4.05255e-07	2.29×10^{-3}	7.86536e-08	3.1×10^{-6}	2
2	0.32	1.04681e-06	5.10434e-07	6.28×10^{-4}	9.79559e-08	1.8×10^{-6}	2
2	0.64	2.78523e-06	4.62936e-10	8.25×10^{-5}	6.14928e-08	2.8×10^{-6}	2
3	1.28	2.68713e-02	9.03762e-02	3.10×10^{-4}	0.109358	7.9×10^{-7}	5
5	0.005	0.00858664	0.0810216	5.78×10^{-6}	0.930164	3.8×10^{-6}	4
5	10.24	0.00238768	0.188601	1.80×10^{-6}	0.846989	1.7×10^{-9}	8

TABLE XVIII
COMPUTATIONAL COST ON THE HARDWARE OF SECTION V-G: MEDIAN WALL TIME WITH INTERQUARTILE RANGE IN BRACKETS (RANGE OVER SEEDS FOR THE PARAMETRIC TRAINING). BATCHED $E(\lambda)$ IS PER QUERY; THE OLD METHOD IS THE DENSE `eigh` THE BENDER-WU SCRIPT USED BEFORE.

Task	Wall time
Grid, $N = 2000$, 2 states	1.80645 ms (0.4316)
Fock basis, $N_b = 200$, 2 states	2.3824 ms (1.16465)
Richardson grid, 2 states	3.09895 ms (0.53145)
Bender-Wu grid, 8 states	3.6716 ms (0.5609)
Dense `eigh`, $N = 2000$ (old method)	3.71357 s
Dedicated, warm, per network	13.5939 s (2.27785)
Dedicated, cold, per network	16.1773 s (11.4146)
Dedicated, $8\lambda \times 2$ states	252.838 s
Parametric, 2 states, per seed	744.249 s (690.162–765.232)
Parametric $E(\lambda)$, one query	98.3742 μs
Parametric $E(\lambda)$, batched	0.4573 μs
Parametric $\psi(x; \lambda)$, 2000 points	1058.52 μs
Peak memory, exact solver	353.9 MiB
Peak memory, dedicated PINN	413.3 MiB
Peak memory, parametric PINN	427.6 MiB

network may not be a limit of the method. Several improvements could be possible: wider networks could increase representational capacity, Fourier features in λ could help the network capture more complex variations of the solution with the perturbation parameter, adaptive collocation could place more points in regions where the residual is large and loss re-weighting could improve the balance between the Schrodinger residual and the normalization constraint. Testing these approaches could reduce the error further and bring the parametric PINN's accuracy closer to that of the dedicated numerical solver.

5) **Boundary effects in the lambda range:** The dense sweep shows its largest deviations at $\lambda = 0$ and $\lambda = 1$, the edges of the sampling interval where the network receives training signal from one side only. Training on a slightly wider interval than the one reported could mitigate this but has not been tested.
6) **Trainable eigenvalue:** E_θ is only weakly coupled to the loss near convergence, so only E_R carries a residual-based error bound (Section III-E). A loss that ties E_θ to the Rayleigh quotient could be tested

IX. VIDEO

https://youtu.be/SQAeLG8Srd0

References

[1] C. M. Bender and T. T. Wu, "Anharmonic oscillator," *Physical Review*, vol. 184, no. 5, p. 1231, 1969.

[2] C. M. Bender and T. T. Wu, "Large-order behavior of perturbation theory," *Physical Review Letters*, vol. 27, no. 7, p. 461, 1971.

[3] L. Brevi, A. Mandarino, and E. Prati, "Addressing the non-perturbative regime of the quantum anharmonic oscillator by physics-informed neural networks," *New Journal of Physics*, vol. 26, no. 10, p. 103015, 2024.

[4] J. C. Pommerening and D. P. DiVincenzo, "Explicit diagrammatic solution of normalised, nondegenerate rayleigh–schrödinger perturbation theory," in *Annales Henri Poincaré*, vol. 23, pp. 2745–2768, Springer, 2022.

[5] 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*, vol. 378, pp. 686–707, 2019.

[6] I. E. Lagaris, A. Likas, and D. I. Fotiadis, "Artificial neural networks for solving ordinary and partial differential equations," *IEEE transactions on neural networks*, vol. 9, no. 5, pp. 987–1000, 1998.

[7] H. Mutuk, "Energy levels of one-dimensional anharmonic oscillator via neural networks," *Modern Physics Letters A*, vol. 34, no. 12, p. 1950088, 2019.

[8] T. Kato, "On the upper and lower bounds of eigenvalues," *Journal of the Physical Society of Japan*, vol. 4, no. 4-6, pp. 334–339, 1949.

[9] J. Zinn-Justin and U. D. Jentschura, "Multi-instantons and exact results i: Conjectures, wkb expansions, and instanton interactions," *Annals of Physics*, vol. 313, no. 1, pp. 197–267, 2004.

[10] H. Jin, M. Mattheakis, and P. Protopapas, "Physics-informed neural networks for quantum eigenvalue problems," in *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2022.

[11] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via deeponet based on the universal approximation theorem of operators," *Nature machine intelligence*, vol. 3, no. 3, pp. 218–229, 2021.

[12] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," *arXiv preprint arXiv:2010.08895*, 2020.

[13] M. Mattheakis, G. R. Schleder, D. T. Larson, and E. Kaxiras, "First principles physics-informed neural network for quantum wavefunctions and eigenvalue surfaces," *arXiv preprint arXiv:2211.04607*, 2022.