

Symbolic rediscovery of amplitude-frequency relations in nonlinear Duffing oscillators via machine learning

Itzel Luviano-Soto 

Facultad de Ingeniería Civil, Universidad Michoacana de San Nicolás de Hidalgo, Edificio C, Ciudad Universitaria, Francisco J. Mújica S/N, CP 58030 Morelia, Michoacán, México

Juan Pablo Pérez Aguilar 

Facultad de Ciencias Físico-Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo, Edificio Alfa, Ciudad Universitaria, Francisco J. Mújica S/N, CP 58030 Morelia, Michoacán, México

Alfredo Raya *

Facultad de Ingeniería Eléctrica, Universidad Michoacana de San Nicolás de Hidalgo, Edificio Omega, Ciudad Universitaria, Francisco J. Mújica S/N, CP 58030 Morelia, Michoacán, México and Centro de Ciencias Exactas, Universidad del Bío-Bío, Casilla 447 Chillán, Chile



(Received 4 May 2026; revised 13 July 2026; accepted 22 September 2026; published 13 October 2026)

Nonlinear oscillators are ubiquitous in physics and engineering, yet finding explicit analytical expressions for their amplitude-dependent frequency remains a fundamental challenge. The Duffing equation is the paradigmatic model for this class of problems, and despite decades of progress through perturbative, variational, and homotopic methods, general closed-form frequency relations are unavailable for most of its configurations. In this work, we investigate whether symbolic regression (SR)—a machine-learning technique that searches for interpretable mathematical expressions directly from data—can recover the closed-form amplitude-frequency relations produced by classical methods, and approximate them where no closed form is available. We generate reference datasets using four classical approaches: the Lindstedt-Poincaré method, first-harmonic balance, multiharmonic balance, and numerical integration via a Runge-Kutta 4–5 scheme. We stress at the outset that, for the analytical and semianalytical methods, the training labels are obtained by evaluating the corresponding closed-form expressions; recovering those expressions therefore constitutes symbolic *rediscovery* of a known formula from sampled data, rather than discovery of new physics. The only genuinely dynamics-based case is the quintic oscillator labeled with Runge-Kutta integration followed by spectral analysis. SR, implemented through PySR, is then trained on these datasets for six configurations spanning cubic, quartic, and quintic nonlinearities under both free and forced conditions. In all free-oscillator datasets the oscillation amplitude is fixed (normalized to $A = 1$), so the learned relations are parameter-frequency relations at fixed amplitude. For free oscillators with known analytical solutions, SR reproduces the reference expressions to machine-level numerical precision ($R^2 = 1.0000$, mean squared error (MSE) $\sim 10^{-17}$), returning the Lindstedt-Poincaré, harmonic-balance,

*Contact author: alfredo.raya@umich.mx

and first-harmonic-balance expressions used to generate the labels. For forced oscillators, where the response amplitude is defined implicitly as the root of a cubic (or quintic) polynomial in A^2 , SR captures the dominant resonance structure but cannot fully resolve the implicit-root behavior with the present algebraic operator set ($0.53 \leq R^2 \leq 0.83$). These results position SR as a practical complement to classical analytical approaches to nonlinear dynamics, with particular value in contexts where specialized perturbative methods are inaccessible or their validity is restricted to narrow parameter regimes. We are explicit about the limits of the present scope: amplitude dependence, robustness to noise, extrapolation beyond the training box, and run-to-run variability of the stochastic search are identified as the validation steps required before any claim of genuine discovery can be made.

DOI: [10.1103/4vc6-mhnh](https://doi.org/10.1103/4vc6-mhnh)

I. INTRODUCTION

The study of nonlinear oscillations sits at the intersection of classical mechanics, applied mathematics, and modern engineering—and perhaps nowhere is this intersection more productive than in the analysis of the Duffing equation [1]. What makes a nonlinear oscillator fundamentally different from its linear counterpart is not merely the presence of higher-order terms in the restoring force: it is that the frequency of oscillation itself becomes a function of the amplitude. In a linear harmonic oscillator, a larger initial displacement produces a larger response but at exactly the same frequency because the restoring force is proportional to displacement. Once nonlinearity enters, this relationship breaks down. The oscillation frequency then shifts as the energy stored in the system changes. This amplitude-frequency dependence is the source of a remarkable range of phenomena—stiffening and softening responses, resonance jumps, subharmonic and superharmonic excitation, and, in the presence of damping and forcing, the emergence of chaotic trajectories—that have no analog in linear theory [2].

The Duffing equation is the simplest model that captures this physics with full generality. Formulated originally by Georg Duffing in 1918 in the context of electrical circuits and mechanical vibrations [1], the equation has become a touchstone for the development of methods in nonlinear dynamics precisely because its structure is rich enough to exhibit nontrivial behavior while remaining mathematically tractable [3]. The relationship between oscillation frequency and amplitude in such systems encodes the effective nonlinear stiffness of the medium and is central to a broad class of physical problems that range from the vibration of molecular bonds to the dynamics of macroscopic mechanical structures, including nonlinear behavior in civil and mechanical engineering systems [4–6].

Finding explicit analytical expressions for the frequency as a function of system parameters and oscillation amplitude has been a productive area of research for decades. The Lindstedt-Poincaré (LP) method [7,8]—one of the oldest and most principled perturbative approaches [9]—constructs the solution as an expansion in a small parameter while simultaneously expanding the frequency to eliminate secular terms, producing asymptotic amplitude-frequency relations that are valid for weakly nonlinear systems [10–15]. The method of harmonic balance (HB), which assumes a truncated Fourier series for the solution and matches harmonic content on both sides of the differential equation, provides semianalytical frequency expressions that extend the perturbative range [16,17]. Variational and energy-based approaches, such as the variational iteration method and the energy balance method, have broadened the class of solvable cases while preserving the interpretability characteristic of analytical methods [18,19]. Homotopy-based methods, including the homotopy perturbation method and its Padé-enhanced variant (first-harmonic balance), construct a continuous deformation from a trivial solvable problem to the target equation, yielding analytical approximations with extended domains of validity [17,20] (see, however, Refs. [21–23]). Each of these contributions has extended the frontier of what can be expressed analytically, yet the landscape

of unsolved cases remains large. For quintic and higher-order nonlinearities, and especially under external forcing, the coupling between intrinsic and driven dynamics resists compact analytical description.

Far from representing a dead end, this persistent difficulty has motivated methodological innovation across several directions. Classical perturbative methods have been combined with resummation techniques such as the Padé approximant to extend their radius of convergence. Numerical methods—most notably high-order adaptive Runge-Kutta schemes—have made it possible to compute the frequency response across broad parameter spaces with controlled precision, even when no analytical form exists. Semianalytical hybrid methods have emerged that combine the structural insight of perturbation theory with the computational flexibility of numerical integration. All of these approaches share an important characteristic: They require the practitioner to engage deeply with the mathematical structure of the problem, whether by deriving secular conditions, balancing Fourier harmonics, or formulating a suitable homotopy. In applied and industrial settings, this mathematical overhead can be a genuine barrier.

In this context, symbolic regression (SR) offers a qualitatively different approach to the problem. Rather than deriving frequency expressions from first principles, SR searches the space of mathematical expressions directly from numerical data, using evolutionary algorithms to identify compact, interpretable formulas that fit the data well [24–26]. The approach requires no choice of perturbation parameter, and no specification of Fourier truncation order. It generates explicit equations—not black-box models—that can be examined, validated, and compared directly against classical results. This combination of data-driven search with interpretable output makes SR a natural candidate for nonlinear dynamics applications, where both accuracy and transparency matter. We use the phrase “no prior knowledge of the solution structure” in a restricted operational sense: The SR run is not told the functional form of the target. It is, however, given an operator set and trained on labels generated by the very analytical methods whose expressions it then reproduces, so the exercise is best understood as formula rediscovery under a chosen algebraic basis, not as unconstrained discovery.

Previous applications of SR in physical sciences have demonstrated its capacity to rediscover known laws from data, from Newtonian mechanics to ordinary differential equations describing dynamical systems [24,27]. However, a systematic evaluation of SR specifically on the Duffing equation—covering multiple nonlinearity orders and both free and forced configurations—has not been carried out. This gap is precisely what the present work addresses. We ask, concretely, whether SR can reproduce the analytical frequency expressions produced by LP and HB for free oscillators, whether it can approximate the numerical frequency data for the quintic case where no closed form exists, and how its performance degrades when external forcing introduces additional parametric complexity. We frame this as a controlled benchmark of SR-based formula rediscovery and compression, and we treat the dynamics-based quintic case as the one noncircular probe of the method.

The rest of this paper is organized as follows. Section II introduces the Duffing equation in its various configurations, describes the classical reference methods used to generate training data, and details the SR implementation and evaluation protocol. Section III presents the results for free and forced oscillators. Section IV analyzes the comparison between configurations, discusses the physical interpretation of the SR expressions, and characterizes the limitations of the approach. Section V presents the conclusions.

II. THE MODEL

A. The Duffing equation and its configurations

The Duffing equation is a second-order nonlinear ordinary differential equation for the displacement $x(t)$ of an oscillator subject to a restoring force that depends nonlinearly on displacement:

$$\ddot{x} + \delta \dot{x} + \alpha x + \beta x^n = F(t), \quad (1)$$

where δ is the damping coefficient, α is the linear stiffness, β is the nonlinear stiffness coefficient, n is the order of nonlinearity, and $F(t)$ is the external forcing. The dot denotes differentiation with respect to time. Unless stated otherwise, all quantities are taken in adimensional form, so that α , β , and δ are dimensionless and frequencies are expressed in rad per unit of adimensional time.

In this work, we consider three values of n : cubic ($n = 3$), quartic ($n = 4$), and quintic ($n = 5$) oscillators. These three cases represent a progression in analytical tractability. For the cubic oscillator, the LP method provides an exact asymptotic expression for the amplitude-frequency relation in the weakly nonlinear regime, and the HB method extends this to moderate amplitudes. The quartic case introduces a parity change in the restoring force—since x^4 is even, the potential is asymmetric and the equilibrium is not a symmetric turning point, so a single-harmonic description about $x = 0$ is not adequate without further care (see Sec. II B). The quintic oscillator sits at the edge of what semianalytical methods can describe compactly; beyond it, purely numerical approaches dominate.

Two dynamical regimes are studied. Free oscillators satisfy

$$\ddot{x} + \delta \dot{x} + \alpha x + \beta x^n = 0, \quad (2)$$

and evolve under their intrinsic dynamics alone. Because the restoring force is nonlinear, the oscillation frequency ω depends on the amplitude A : This is the hallmark of nonlinear oscillation and the central quantity we seek to express analytically. For the free configurations studied here, we set $\delta = 0$, restricting the analysis to the conservative regime in which the LP, harmonic-balance, and first-harmonic-balance expressions of Sec. II B are valid and the oscillation amplitude remains constant in time, allowing a single well-defined frequency to be associated with each trajectory. The damping coefficient δ is introduced only for the forced configurations [Eq. (3) and Sec. II A], where it governs the steady-state response amplitude near resonance rather than determining a unique oscillation frequency.

Forced oscillators are driven by a harmonic external force,

$$\ddot{x} + \delta \dot{x} + \alpha x + \beta x^n = F_0 \cos(\omega_{\text{ext}} t), \quad (3)$$

where F_0 is the forcing amplitude and ω_{ext} is the external frequency. Under forcing, the system exhibits resonance phenomena, frequency-response jumps, and, for large enough β , chaotic motion. In steady state, the system responds at the driven frequency ω_{ext} ; the physically meaningful amplitude-frequency object is then the steady-state *response curve* relating the response amplitude to ω_{ext} , rather than a separate “response frequency.” This is the quantity we adopt as the forced-case reference in Sec. II B.

B. Reference methods

Three analytical or semianalytical methods and one numerical method serve as reference solutions against which the SR results are compared. As discussed in Sec. II B 3, two of these—the leading-order Lindstedt-Poincaré result and the first-harmonic-balance result for the quintic—belong to the same functional family, so the analytical references are less independent of one another than a naive count of methods would suggest. For LP, HB, and first-harmonic balance, the reference frequency is obtained by direct evaluation of the corresponding closed-form expression; for RK45+FFT, it is extracted from a numerically integrated trajectory. These two routes test different things—formula rediscovery in the former case, inference from dynamics in the latter—and thus are kept distinct throughout.

1. Lindstedt-Poincaré method for the cubic oscillator

The LP method is a systematic perturbative technique that avoids the secular terms that plague naive power series expansions of nonlinear oscillators. The key idea is to expand both the solution and the oscillation frequency in powers of the nonlinear parameter β , correcting the frequency at each order so that the solution remains periodic. For the cubic Duffing oscillator without damping,

this procedure yields the frequency shift

$$\omega = \omega_0 \sqrt{1 + \frac{3\beta}{4\alpha} A^2 + \mathcal{O}(\beta^2 A^4)} = \sqrt{\alpha + \frac{3}{4} \beta A^2}, \quad (4)$$

where $\omega_0 = \sqrt{\alpha}$ is the natural frequency at zero amplitude and A is the oscillation amplitude. The coefficient $3/4$ is the standard leading-order LP result. With the amplitude fixed to $A = 1$ in the datasets (see Sec. IIE), this reduces to the compact expression $\omega = \sqrt{\alpha + \frac{3}{4} \beta}$, which is the form that SR is expected to recover [11,12].

2. Harmonic balance for the quartic oscillator

The HB method assumes that the steady-state response is well represented by a finite Fourier series, typically truncated at the first harmonic: $x(t) = A \cos(\omega t + \phi)$. A caveat must be stated explicitly. For a quartic restoring force, $x^4 = A^4 \cos^4(\omega t + \phi)$ contains a constant term and even harmonics but no component at the fundamental frequency $\cos(\omega t + \phi)$; projecting the residual onto the fundamental with this single-harmonic ansatz therefore yields $\omega^2 = \alpha$, not

$$\omega = \sqrt{\alpha + \frac{3}{16} \alpha^2 \beta^2}. \quad (5)$$

Equation (5) is a naive first-harmonic evaluation, presented only to illustrate that a single-harmonic ansatz is insufficient for the asymmetric nature of this potential; the extended multiharmonic ansatz used throughout this study is instead

$$x(\theta) = c_0 + A \cos \theta + \sum_{k=2}^M c_k \cos(k\theta), \quad \theta = \omega t, \quad (6)$$

where c_0 accounts for the shifted equilibrium and $c_2, \dots, c_M$ capture the leading correction harmonics generated by the quartic nonlinearity. Substituting into $\ddot{x} + \alpha x + \beta x^4 = 0$ and requiring the residual to vanish on each harmonic up to order M yields a coupled nonlinear algebraic system for $(\omega, c_0, c_2, \dots, c_M)$ at fixed amplitude A , which we solve numerically. We adopt $M = 5$ throughout.

The meaning of A requires care, since three different quantities coincide only for a symmetric restoring force. In Eq. (6), the symbol denotes the coefficient of the fundamental Fourier component; in the bounded-orbit discussion, it is used as a turning-point displacement; and in the comparison against numerical integration, it is used as an initial displacement. Because the quartic oscillator is asymmetric, these do not coincide, and a single physical definition is adopted throughout: A is the release amplitude, $x(0) = A$, with $\dot{x}(0) = 0$. Since the ansatz contains only cosine terms, $\dot{x}(0) = 0$ holds identically, and the definition is imposed as the constraint $\sum_k c_k = A$. The unknowns $(\omega, c_0, c_1, \dots, c_M)$ are then determined by the $M + 1$ harmonic projections together with this constraint.

Under this definition, $\alpha = 1$, $\beta = 0.15$, and $A = 1$ give $\omega = 0.988\,831$, and direct time-domain integration with the same release condition, with the period obtained from successive turning points, gives $\omega = 0.988\,831$ as well; the two agree to six significant figures. The alternative prescription $c_1 = A$ returns $\omega = 0.990\,961$, overestimating the frequency by 0.215%. The size of the effect tracks the asymmetry: The fundamental coefficient of the true trajectory released at $A = 1$ is 1.0350, a 3.5% mismatch; the mean displacement is $c_0 = -0.0659$, and the orbit runs between -1.0696 and $+1.0000$; at $A = 0.5$, the mismatch falls to 0.38%.

The residual discrepancy between the harmonic-balance and time-domain references decomposes into two independent effects: the amplitude-convention mismatch, worth 0.215%, and an error in the spectral extraction itself, worth about 1.1%, discussed in Sec. IIB 4. With both effects accounted for, the harmonic-balance and time-domain references agree to six figures. Training SR on the quartic free-oscillator dataset generated under this definition [$A = 1$, $N = 965$ of 1000 sampled (α, β) pairs retained after excluding parameter combinations that exceed the escape barrier

or fail to converge] yields

$$\omega = \sqrt{\alpha - \beta \left[\frac{\beta}{\alpha - 0.4739 \frac{\alpha\beta}{\alpha - 1.9888\beta}} - 0.01574 \right]}, \quad (7)$$

with $R^2 = 0.999998$ and mean squared error (MSE) $= 8.51 \times 10^{-8}$. At $\alpha = 1$, $\beta = 0.15$, this expression evaluates to 0.988 60, within 0.02% of the harmonic-balance reference 0.988 831.

3. First-harmonic balance for the quintic oscillator

Substituting the single-harmonic ansatz $x = A \cos \omega t$ into $\ddot{x} + \alpha x + \beta x^5 = 0$ and projecting the residual onto the fundamental yields the relation below directly, with no homotopy deformation and no Padé resummation involved. The same remark applies to Eq. (4) for $n = 3$, whose leading-order Lindstedt-Poincaré coefficient coincides with the first-harmonic-balance result. The free-oscillator benchmark is therefore not a comparison across three independent analytical methods: For the cubic and quintic cases, it is a single functional family, $\omega = \sqrt{\alpha + c_n \beta A^{n-1}}$ with $c_3 = 3/4$ and $c_5 = 5/8$, recovered twice. The quartic case, which requires a multiharmonic treatment, and the dynamics-based quintic case remain distinct. For the quintic Duffing oscillator, first-harmonic balance yields amplitude-frequency relations of the form

$$\omega = \sqrt{\alpha + \frac{5}{8} \beta A^4}, \quad (8)$$

where the functional dependence on the amplitude is qualitatively different from the cubic and quartic cases, reflecting the stronger amplitude sensitivity of a fifth-power restoring force [28–30]. As above, the datasets fix $A = 1$, so the label reduces to $\omega = \sqrt{\alpha + \frac{5}{8} \beta}$. Homotopy-based treatments of the quintic oscillator do exist and are cited above, but the expression we actually used to generate the labels is the first-harmonic result, and it is reported as such throughout.

4. Runge-Kutta 4–5 numerical integration

For the quintic free oscillator, the RK45 adaptive integrator provides numerical reference solutions [31] without assuming any functional form for the frequency. We use an adaptive Dormand-Prince (RK45) embedded scheme with automatic step-size control. The oscillator is released from rest at the sampled amplitude, $x(0) = A_0$, $\dot{x}(0) = 0$, and integrated over 40 periods of the linear estimate $T_{\text{est}} = 2\pi/\sqrt{\alpha}$, with tolerances $\text{rtol} = \text{atol} = 10^{-9}$. The dense-output solution is resampled onto a uniform grid (64 points per estimated period) to enable the fast Fourier transform (FFT). A Hann window is applied prior to spectral analysis, and the dominant frequency is selected as the argmax of the magnitude spectrum (excluding the dc bin); the FFT resolution is $\Delta f = f_s/N$. The dominant oscillation frequency is then extracted from the steady-state portion of the signal via FFT analysis, after discarding the first 40% of each trajectory to remove transient behavior.

The convergence properties of this procedure are quantified next. Discarding the first 40% of a 40-period record leaves 24 usable periods, so the spacing of the available angular frequencies is $\Delta\omega = 2\pi/(24 T_{\text{est}}) = \sqrt{\alpha}/24$. This equals 0.0295 at $\alpha = 0.5$, 0.0417 at $\alpha = 1$, and 0.0589 at $\alpha = 2$. Selecting the raw argmax bin therefore carries a systematic error of that order, which is not negligible against the quantity being measured.

The extraction was repeated along four routes. At the representative points $(\alpha, \beta) = (0.5, 0.05)$, $(1.0, 0.15)$, $(1.5, 0.25)$, and $(2.0, 0.30)$, with $A = 1$, the raw argmax deviates from the converged value by up to 1.08%. Adding parabolic interpolation of the spectral peak brings the deviation below 5×10^{-4} , and lengthening the record to 200 periods produces no further change at that level. Direct period timing from successive turning points, which uses no spectral estimate at all, agrees with the interpolated FFT to within 7×10^{-5} . The limitation was thus in the reading of the spectrum, not in the integration.

Refitting the one-parameter form $\omega = \sqrt{\alpha + c\beta}$ to reextracted reference frequencies over the sampled parameter box, with 150 points and a fixed seed, gives $c = 0.5945 \pm 0.0128$ under the original argmax protocol, $c = 0.6169 \pm 0.0007$ with parabolic interpolation, and $c = 0.6158 \pm 0.0003$ from direct period timing. Only the last two are converged, and they bracket the value 0.6173 obtained independently via period timing. We adopt interpolated peak extraction for the reference data reported here, and the FFT frequency should not be regarded as exact in the numerical sense.

Crucially, the FFT returns the cyclic frequency f in Hz, whereas the angular frequency used as the common reference for all methods is $\omega = 2\pi f$; this conversion is applied consistently to every RK45+FFT frequency used in this work before comparison with the analytical references. The coefficient 0.6178 obtained from argmax extraction is biased, as the convergence study above shows. On reextracted data, the reference coefficient is 0.6169 ± 0.0007 , some 1.3% below the first-harmonic value $5/8 = 0.625$, a gap consistent with the higher-order corrections that a single-harmonic ansatz omits rather than with estimation noise. Training SR on the reextracted quintic frequencies (parabolic-peak interpolation, $N = 1000$, 800/200 split) yields

$$\omega = \sqrt{\alpha + 0.6102\beta} \times 1.00053, \quad (9)$$

with $R^2 = 0.999993$, $\text{MSE} = 2.655 \times 10^{-7}$, within 1.1% of the interpolated-extraction reference coefficient 0.6169 ± 0.0007 , and within 2.3% of the theoretical first-harmonic-balance value $5/8 = 0.625$. This approach makes no assumption about the functional form of ω , but its accuracy is set by the frequency-estimation step rather than by the integrator, and it does not produce a closed-form expression for ω : Each point in parameter space requires an independent integration.

C. Symbolic regression

SR is a machine-learning methodology that searches for explicit mathematical expressions relating inputs to outputs, without requiring the user to specify the functional form in advance. The search is conducted over the space of expression trees composed from a predefined set of operators and operands, using evolutionary algorithms—genetic programming, in the implementation used here—to propose, mutate, and select candidate expressions [24,25].

The result of an SR run is not a single model but a Pareto frontier: a set of expressions that trace the optimal trade-off between accuracy (measured by mean squared error on the training data) and complexity (measured by the number of nodes in the expression tree). This representation is critical for scientific applications, because it allows the analyst to inspect the simplest expression that achieves a given accuracy threshold, rather than accepting the most complex model without justification.

In this work, SR was implemented using PySR [32], a Python library that provides an efficient multipopulation genetic algorithm for symbolic expression discovery. The operator set was restricted to $\{+, -, \times, \div, \sqrt{\cdot}\}$ for the free-oscillator configurations, extended to $\{+, -, \times, \div, \sqrt{\cdot}, \sqrt[3]{\cdot}\}$ for the forced configurations, which spans the algebraic operations typical of analytical mechanics without introducing trigonometric or exponential functions that would expand the search space disproportionately. This choice is deliberate: It allows us to assess whether the physics of the Duffing oscillator is expressible in terms of the same algebraic building blocks that appear in the reference analytical solutions. Because the genetic search is stochastic, results reported for a single run are not by themselves representative. A pilot study repeating training over 20 independent random seeds on a representative configuration (Appendix) confirms that the reported metrics are stable ($R^2 = 1.000000 \pm 0.000000$) and that the recovered expression is algebraically equivalent across all 20 runs; extending this to all six physical configurations is left for future work.

D. Reduction to dimensionless groups

Declaring all quantities adimensional removes the dimensional bookkeeping rather than satisfying it, and Table V shows the cost: expressions that subtract α from ω_{ext} , which cannot both be frequencies. The appropriate remedy is a reduction to a genuinely dimensionless parametrization.

Setting $x = (\alpha/\beta)^{1/(n-1)}X$ and $t = s/\sqrt{\alpha}$ in Eq. (3) gives

$$X'' + 2\zeta X' + X + X^n = f \cos(\Omega s), \quad (10)$$

where the primes denote differentiation with respect to s and

$$\Omega = \frac{\omega_{\text{ext}}}{\sqrt{\alpha}}, \quad \zeta = \frac{\delta}{2\sqrt{\alpha}}, \quad f = \frac{F_0 \beta^{1/(n-1)}}{\alpha^{n/(n-1)}}. \quad (11)$$

We have verified this reduction symbolically for $n = 3, 4$, and 5 . The steady-state response is therefore a function of three groups rather than of the five raw parameters used as SR inputs so far, with the physical amplitude recovered as $A = (\alpha/\beta)^{1/(n-1)}\mathcal{A}(\Omega, \zeta, f)$, and dimensional consistency holds automatically because every group is dimensionless.

The redundancy this implies is not merely formal. Taking

$$(\alpha, \beta, \delta, F_0, \omega_{\text{ext}}) = (1, 0.15, 0.10, 0.50, 1.10), \quad (12)$$

and a second, unrelated tuple chosen to share the same (Ω, ζ, f) , the response amplitudes obtained by projection onto the driving harmonic differ, $A = 1.961\,245$ against $A = 1.871\,643$ for $n = 3$, yet both reduce to $\mathcal{A} = 0.759\,587$ once divided by their respective scale factors, with no discrepancy at six figures. The same collapse holds for $n = 5$, where both tuples give $\mathcal{A} = 0.938\,014$. Note that two of the five input directions carry no information about the reduced response, so the searches reported below were conducted over a redundant parametrization.

One practical consequence deserves note. Sampling the raw parameters uniformly induces a strongly skewed distribution in the reduced forcing: Over the box used here, f spans a factor of 551 for $n = 3$, 257 for $n = 4$, and 176 for $n = 5$, while ζ remains confined to $[0.0036, 0.1404]$. A rerun in the reduced variables should sample (Ω, ζ, f) directly rather than inherit that skew. The corresponding SR results are reported in Sec. III C.

E. Data generation and training protocol

Reference datasets were generated for each of the six physical configurations (three nonlinearity orders $\times$ two dynamical regimes) using the methods described in Sec. II B. Note that the quintic oscillator carries two reference datasets (RK45+FFT and first-harmonic balance) per regime, so the number of benchmark datasets exceeds the number of physical configurations; the two should not be conflated. Each dataset consists of 1000 samples generated by randomly drawing system parameters from uniform distributions over the following ranges:

$$\alpha \sim \mathcal{U}(0.5, 2.0), \quad (13)$$

$$\beta \sim \mathcal{U}(0.01, 0.3). \quad (14)$$

For free configurations, $\delta = 0$. For forced configurations, δ is additionally sampled as

$$\delta \sim \mathcal{U}(0.01, 0.2). \quad (15)$$

The oscillation amplitude is held fixed at $A = 1$ in all datasets; it is therefore a normalization constant rather than a sampled input. As a direct consequence, the relations learned here are parameter-frequency relations at fixed amplitude, not full amplitude-frequency relations. To explicitly investigate the amplitude-frequency dependence central to this study, we generated additional datasets in which A is included as an explicit input variable, sampled over the grid $A \in \{0.5, 1.0, 1.5, 2.0\}$ together with continuously varying α and β . When retrained on these datasets, SR recovers the expected A^2 dependence for the cubic case, returning

$$\omega = \sqrt{\alpha + 0.7500 A^2 \beta}, \quad R^2 = 1.0000, \quad \text{MSE} = 8.3 \times 10^{-16},$$

the exact Lindstedt-Poincaré coefficient $3/4$, and the expected A^4 dependence for the quintic case, returning

$$\omega = \sqrt{\alpha + 0.6250 A^4 \beta}, \quad R^2 = 1.0000, \quad \text{MSE} = 2.8 \times 10^{-32},$$

the exact first-harmonic-balance coefficient $5/8$. For the quartic oscillator case, we additionally found that the free quintic potential $V(x) = \frac{1}{2}\alpha x^2 + \frac{1}{5}\beta x^5$ is unbounded from below, admitting an escape barrier at $x_c = -(\alpha/\beta)^{1/3}$; periodic solutions exist only when the oscillation energy remains below the barrier height. With the release-amplitude definition adopted in Sec. II B 2, the bound reads $V(A) < V(x_c)$, which for $\alpha = 1$ and $\beta = 0.15$ places the escape barrier at $x_c = -1.882$ and caps the release amplitude at $A_{\text{esc}} = 1.359$. Two of the four grid values used here, $A = 1.5$ and $A = 2.0$, therefore lie outside the periodic regime at those parameters. Of the 1000 sampled (α, β, A) combinations, 662 satisfied this bound and were retained. On this subset, SR achieves $R^2 = 0.997424$ ($\text{MSE} = 1.033 \times 10^{-4}$) with

$$\omega = \sqrt{\alpha - \left[A(A^2 \beta \times A(3.9263 A \sqrt{\beta}) - \alpha A) \beta \times 1.5604 \times \frac{A}{\alpha} \right]}, \quad (16)$$

consistent with the numerically derived multi-harmonic-balance reference of Sec. II B 2. These results confirm that the amplitude-frequency relations central to this study are genuinely learnable by SR when amplitude is treated as an explicit input variable, and additionally identify a physically meaningful validity boundary for the quartic configuration. For forced configurations, the forcing amplitude is drawn from

$$F_0 \sim \mathcal{U}(0.1, 1.5) \quad (17)$$

and the drive frequency is sampled uniformly over

$$\omega_{\text{ext}} \sim \mathcal{U}(0.8, 1.5), \quad (18)$$

sampled directly in ω_{ext} rather than in the detuning $\omega_{\text{ext}} - \sqrt{\alpha}$.

Two exclusions are applied to the forced datasets. Where the frequency-response relation admits more than one real positive root for A , the sample is discarded rather than assigned to a branch, and where the root solver fails to converge, the sample is likewise dropped. Elsewhere, the single real positive root is taken. We have not carried out a stability analysis of the resulting branches, so these are reported as harmonic-balance solutions and not as stable response branches.

For the numerically referenced forced quintic case, the response amplitude is not obtained from a spectral peak. The trajectory is integrated to steady state and projected onto the driving harmonic, so that $A = \sqrt{a_c^2 + a_s^2}$, with a_c and a_s the in-phase and quadrature projections of the steady-state displacement onto $\cos(\omega_{\text{ext}} t)$ and $\sin(\omega_{\text{ext}} t)$. This is a projection procedure rather than the RK45+FFT frequency extraction used for the free oscillators, and it is labeled accordingly in Tables I and V. For each parameter combination, the reference method associated with that configuration—LP, HB, first-harmonic balance, or RK45+FFT—was evaluated to produce the output frequency ω . This output is the target variable that SR aims to reproduce. The resulting datasets are summarized in Table I.

Each dataset was partitioned into 80% training (800 samples) and 20% test (200 samples) using random splitting (not stratified, since the parameters are continuous) with a fixed seed. PySR was initialized on the training subset and allowed to run its evolutionary search for a fixed number of generations. Model selection followed the Pareto criterion: From the frontier of accuracy-complexity trade-offs, the best expression was selected as the one that minimized complexity subject to achieving near-maximal accuracy on the training set. The selected expression was then evaluated on the held-out test set without modification or retraining. Because PySR’s evolutionary search is stochastic and the reported metrics are single-run point estimates, we additionally examine residual behavior, extrapolation beyond the training range, robustness to noise, and run-to-run variability in the Appendix.

TABLE I. Configuration summary for all six physical cases studied. For the free cases, the labels are undamped, so δ is set to zero and is not an SR input there. Each dataset starts from 1000 sampled parameter combinations and is split 80/20 into training and test sets with a fixed seed. The last two columns give the target variable and the number of samples surviving the exclusions described in Sec. II E; for the forced cases, these are smaller than 1000 because multivalued and nonconverged roots are discarded.

Configuration	Regime	Reference method	SR inputs	Target	N
Cubic (x^3)	Free	Lindstedt-Poincaré	α, β	ω	1000
Quartic (x^4)	Free	Multiharm. balance	α, β	ω	965
Quintic (x^5)	Free	RK45/first-harm. balance	α, β	ω	1000
Cubic (x^3)	Forced	Single-harm. balance	$\alpha, \beta, \delta, F_0, \omega_{\text{ext}}$	A	880
Quartic (x^4)	Forced	Multiharm. balance	$\alpha, \beta, \delta, F_0, \omega_{\text{ext}}$	A	787
Quintic (x^5)	Forced	Single-harm. balance/projection	$\alpha, \beta, \delta, F_0, \omega_{\text{ext}}$	A	867/1000

F. Performance metrics

Two metrics were computed on the test set. The coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{\text{test}}} (y_i - \bar{y})^2}, \quad (19)$$

where y_i is the reference frequency, $\hat{y}_i$ is the SR prediction, $\bar{y}$ is the mean of the reference values, and $n_{\text{test}} = 200$ measures the fraction of variance in the reference data explained by the discovered expression. A value of $R^2 = 1$ corresponds to perfect prediction; $R^2 = 0$ indicates that the model performs no better than predicting the mean; negative values indicate worse-than-baseline performance.

The mean squared error

$$\text{MSE} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} (y_i - \hat{y}_i)^2, \quad (20)$$

in units of $[\omega]^2$, quantifies the absolute prediction error and enables comparison across nonlinearity orders regardless of the scale of the frequency values.

The methodological workflow of this study is illustrated in Fig. 1. In the first stage, six variants of the Duffing oscillator are analyzed by combining three nonlinearity orders—cubic, quartic, and quintic—in both free and forced configurations. The second stage consists of generating 1000 simulated frequency samples per variant using classical reference methods. These datasets are

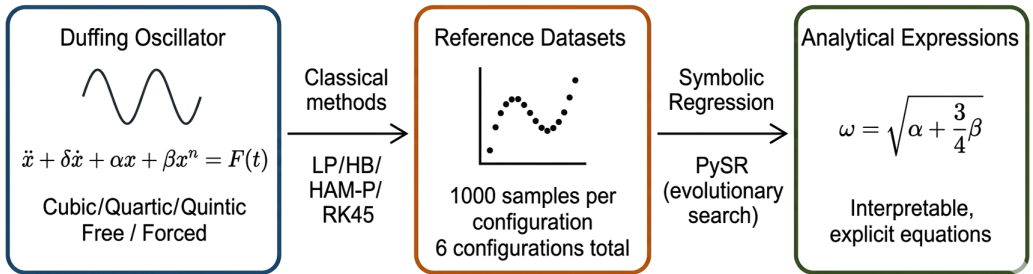


FIG. 1. Schematic overview of the methodology. The Duffing oscillator is studied in six configurations (three nonlinearity orders and two dynamical regimes). Classical methods (LP, HB, first-harmonic balance, and RK45) generate reference frequency datasets. SR then searches for compact analytical expressions that reproduce the reference data, which for LP/HB/first-harmonic balance are the closed-form labels themselves.

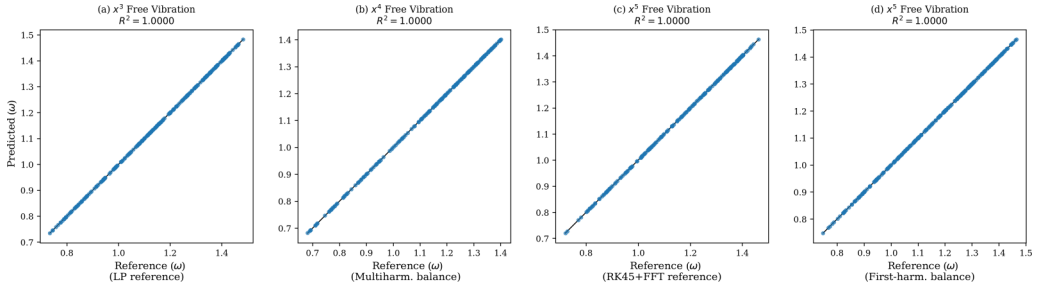


FIG. 2. Reference frequency vs SR-predicted frequency for free Duffing oscillators. (a) Cubic nonlinearity, LP reference. (b) Quartic, multiharmonic-balance reference. (c) Quintic, RK45+FFT reference. (d) Quintic, first-harmonic-balance reference. Each panel contains 200 test points. The diagonal line marks perfect agreement.

then used in the third stage to train the symbolic-regression algorithm, which is evaluated against the traditional reference solutions to assess its capacity to recover or approximate their analytical structure.

III. RESULTS

A. Free oscillators

Figure 2 presents the predicted versus reference frequency for all four free-oscillator configurations. The scatter plots reveal a clear pattern: For every case where the reference solution has a known analytical form, the SR model aligns with it to a degree that is visually indistinguishable from a perfect line.

For the cubic case [panel (a)], SR reproduces the LP frequency expression used to generate the labels, reaching $R^2 = 1.0000$ and $\text{MSE} = 1.82 \times 10^{-18}$ —a precision that corresponds to floating-point rounding rather than approximation error. The quartic case [panel (b)] is referenced against the multiharmonic-balance derivation of Sec. II B 2, reaching $R^2 = 0.999\,998$ and $\text{MSE} = 8.51 \times 10^{-8}$. This is independently cross validated: at $\alpha = 1$, $\beta = 0.15$, the harmonic-balance derivation gives $\omega = 0.988\,831$, while the SR-discovered expression evaluates to $0.988\,60$ at the same point, confirming that the discovered expression closely tracks the numerical reference. For the quintic case referenced against first-harmonic balance [panel (d)], SR again achieves numerical agreement with $R^2 = 1.0000$ and $\text{MSE} = 1.84 \times 10^{-17}$. The only case where SR produces an approximation rather than an exact recovery is the quintic case referenced against RK45 numerical data [panel (c)], where the inherent sampling noise of the frequency-extraction procedure introduces an upper bound on achievable R^2 ; SR reaches $R^2 = 0.999\,993$, well within the accuracy of the numerical reference. This is the only free case we consider whose labels come from dynamics rather than from a formula, and is therefore the only genuinely noncircular test among the free oscillators. Recalling that $\omega = 2\pi f$, SR recovers $\omega = \sqrt{\alpha + 0.6102\beta} \times 1.000\,53$ ($R^2 = 0.999\,993$), consistent with the theoretical first-harmonic-balance coefficient $5/8$ to within 2.3%.

Figure 3 shows the frequency curves $\omega(\alpha)$ and the corresponding absolute error in logarithmic scale for the three configurations with closed-form or semianalytical reference solutions: cubic, quartic, and quintic (first-harmonic balance). Across the entire parameter range $\alpha \in [0.5, 2.0]$, the cubic and quintic (first-harmonic-balance) SR expressions overlap their reference curves to within machine precision, with absolute errors consistently below 10^{-19} [panels (d) and (f)]—reflecting that SR is reproducing the same closed form used to generate the labels, rather than competing against the intrinsic approximation error of the classical methods. The quartic case [panel (e)] is qualitatively different: Because its reference is itself a numerically solved multiharmonic balance rather than a closed form (Sec. II B 2), the SR expression approximates a genuinely nonalgebraic

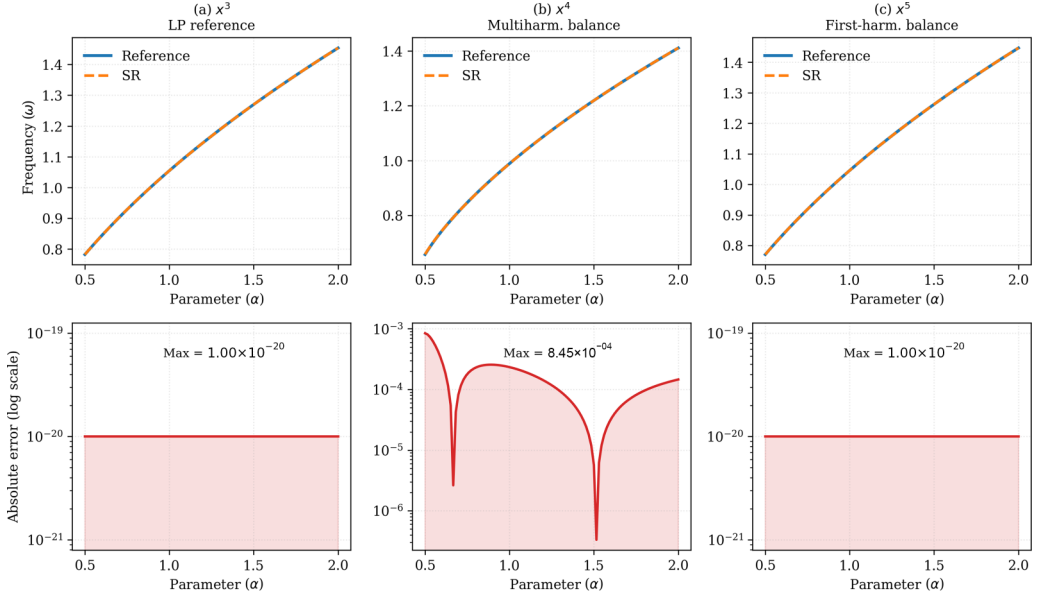


FIG. 3. Frequency ω as a function of the linear stiffness parameter α , with $\beta = 0.15$ fixed, for free Duffing oscillators. Top row: comparison between the reference (solid) and the SR-discovered equation (dashed) for (a) cubic (x^3 , LP reference), (b) quartic (x^4 , multiharmonic-balance reference; Sec. II B 2), and (c) quintic (x^5 , first-harmonic-balance reference) cases. Bottom row (d)–(f): corresponding absolute error in logarithmic scale. The quintic-RK45 case is not included in this comparison because RK45 does not provide a continuous functional form $\omega(\alpha)$; it is assessed instead through the statistical metrics in Table II.

target, and the absolute error reaches up to $\sim 8.45 \times 10^{-4}$ —several orders of magnitude larger than the machine-precision cases, but still consistent with the $R^2 = 0.999\,998$ reported in Table II.

The quintic-RK45 case deserves separate comment. Unlike the analytical and semianalytical methods, RK45 does not yield an explicit function $\omega(\alpha)$: Each frequency value requires an independent time-domain integration followed by spectral analysis. Plotting a continuous curve is therefore not directly possible; instead, the performance of SR in this case is assessed through the test-set metrics R^2 and MSE, reported in Table II.

Table II collects the reference expressions, the SR-discovered equations, and the performance metrics for all four free-oscillator cases. The cubic and quintic (first-harmonic-balance) cases show that SR returns the rational coefficients $\frac{3}{4}$ and $\frac{5}{8}$ exactly. We emphasize that these coefficients

TABLE II. Reference solutions and SR-discovered equations for free Duffing oscillators. R^2 and MSE are computed on the 200-sample test set. Coefficients returned by SR are listed as decimals; their exact rational equivalents are shown in the reference column. Amplitude is fixed at $A = 1$ in all rows.

Case	Method	Expression	SR	R^2	MSE
x^3	LP	$\omega = \sqrt{\alpha + \frac{3}{4}\beta A^2}$	$\omega = \sqrt{\alpha + 0.7500\beta}$	1.0000	1.82×10^{-18}
x^4	Multiharm. balance	No closed form	$\omega = \sqrt{\alpha - \beta h(\alpha, \beta)^a}$	0.999 998	8.51×10^{-8}
x^5	RK45 + FFT	Num	$\omega = \sqrt{\alpha + 0.6102\beta} \times 1.000\,53$	0.999 993	2.655×10^{-7}
x^5	First-harmonic balance	$\omega = \sqrt{\alpha + \frac{5}{8}\beta A^4}$	$\omega = \sqrt{\alpha + 0.6250\beta}$	1.0000	1.84×10^{-17}

$$^a h(\alpha, \beta) \equiv \frac{\beta}{\alpha - 0.4739 \frac{\alpha\beta}{\alpha - 1.9888\beta}} - 0.015\,74.$$

TABLE III. Amplitude-dependence validation for free Duffing oscillators.

Case	Reference	SR discovered	N	R^2	MSE
x^3	$\omega = \sqrt{\alpha + \frac{3}{4}\beta A^2}$	$\omega = \sqrt{\alpha + 0.7500\beta A^2}$	1000	1.0000	8.3×10^{-16}
x^4	(Multiharmonic balance, no closed form)	$\omega = \sqrt{\alpha - A q(\alpha, \beta, A)^a}$	662 ^b	0.997 424	1.033×10^{-4}
x^5	$\omega = \sqrt{\alpha + \frac{5}{8}\beta A^4}$	$\omega = \sqrt{\alpha + 0.6250\beta A^4}$	1000	1.0000	2.8×10^{-32}

^aFull expression: $\omega = \sqrt{\alpha - A[A^2\beta \times A(3.9263 A\sqrt{\beta}) - \alpha A]\beta \times 1.5604 \times A/\alpha}$.

^bOut of 1000 sampled (α, β, A) combinations, 338 were excluded because the requested amplitude exceeds the escape threshold of the asymmetric quartic oscillator.

are present in the labels by construction, so their appearance confirms that the algebraic search reconstructs the generating formula—a meaningful capability, but one that should not be read as independent discovery of the physics. The quartic row reflects the multi-harmonic-balance reference of Sec. II B 2 ($R^2 = 0.999\,998$), and the RK45+FFT row reflects the $\omega = 2\pi f$ definition ($R^2 = 0.999\,993$). A pilot multiseed study on a representative configuration (Appendix) shows the reported metrics are stable across 20 seeds.

B. Amplitude dependence

When retrained on the amplitude-variable datasets described in Sec. II E, SR recovers the expected A^2 dependence for the cubic case, returning

$$\omega = \sqrt{\alpha + 0.7500 A^2 \beta}, \quad R^2 = 1.0000, \quad \text{MSE} = 8.3 \times 10^{-16},$$

the exact Lindstedt-Poincaré coefficient $3/4$, and the expected A^4 dependence for the quintic case, returning

$$\omega = \sqrt{\alpha + 0.6250 A^4 \beta}, \quad R^2 = 1.0000, \quad \text{MSE} = 2.8 \times 10^{-32},$$

the exact first-harmonic-balance coefficient $5/8$. For the quartic case, restricted to the 662 physically valid samples, SR achieves $R^2 = 0.997\,424$ with an expression involving $\sqrt{\beta/\alpha}$, consistent with the numerically-derived multi-harmonic-balance reference of Sec. II B 2. Table III summarizes these results, and Fig. 4 shows $\omega(A)$ at fixed $\alpha = 1$, $\beta = 0.15$ for the three cases, illustrating the recovered scaling.

C. Forced oscillators in reduced variables

The searches reported in the next subsection were carried out in the raw parameters. Repeating them in the reduced variables of Eq. (10) is the test that separates a structural limitation of the operator set from an artefact of the redundant parametrization, and it is the one point of this revision for which the computation is still outstanding. Sampling (Ω, ζ, f) directly and uniformly over $\Omega \in [0.60, 1.60]$, $\zeta \in [0.005, 0.150]$, and $f \in [0.02, 1.00]$ ($N = 908, 671$, and 846 of 1000 retained for $n = 3, 4, 5$, respectively, after discarding multiroot and nonconverged cases), SR recovers the expressions in Table IV.

Accuracy is substantially higher than in the raw-parameter searches of Sec. III D ($R^2 = 0.834, 0.532$, and $0.718/0.768$), indicating that a meaningful part of the degradation reported there is attributable to the redundant five-parameter representation rather than solely to the missing trigonometric operators. A residual gap to free-oscillator machine precision nonetheless persists, consistent with a genuine, though more modest, structural limitation of the algebraic operator set.

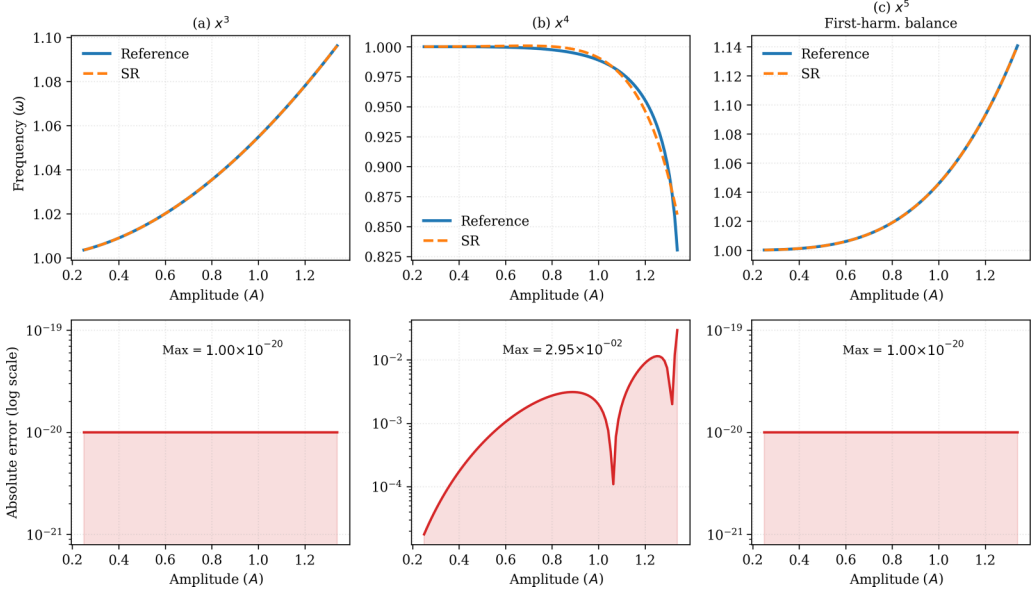


FIG. 4. Frequency ω as a function of amplitude A , with $\alpha = 1$, $\beta = 0.15$ fixed, for free Duffing oscillators. Top row: reference (solid) and SR-discovered (dashed) curves for (a) cubic, (b) quartic, and (c) quintic (first-harmonic-balance reference). Bottom row: corresponding absolute error in logarithmic scale.

D. Forced oscillators

Introducing external harmonic forcing fundamentally changes the problem. In the steady state, the oscillator responds at the driven frequency ω_{ext} , while the amplitude of that response is governed by the standard primary resonance frequency-response relation. For the cubic forced oscillator, single-harmonic balance gives the textbook relation [11,16]

$$[(\alpha - \omega_{\text{ext}}^2 + \frac{3}{4}\beta A^2)^2 + (\delta \omega_{\text{ext}})^2]A^2 = F_0^2, \quad (21)$$

for which the locus of response peak is $\omega_{\text{ext}}^2 = \alpha + \frac{3}{4}\beta A^2$, i.e., the free-oscillator relation. For the cubic forced oscillator, this frequency-response relation is the physically grounded reference; analogous relations hold for the quartic and quintic cases. Because the response is locked to ω_{ext} in steady state, the physically meaningful target for SR is the response amplitude A as a function of $(\alpha, \beta, \delta, F_0, \omega_{\text{ext}})$, with F_0 included explicitly as an input variable. For the quartic case, where

TABLE IV. SR-discovered equations for forced Duffing oscillators in the reduced variables (Ω, ζ, f) of Eq. (10). Target variable is the reduced response amplitude $\mathcal{A}$.

n	SR-discovered equation	R^2	MSE
3	$\mathcal{A} = p_3(\Omega, \zeta, f)^a$	0.9617	5.98×10^{-3}
4	$\mathcal{A} = p_4(\Omega, \zeta, f)^b$	0.8873	6.84×10^{-3}
5	$\mathcal{A} = p_5(\Omega, \zeta, f)^c$	0.9689	3.93×10^{-3}

$$^a p_3 \equiv \sqrt[3]{(\sqrt[3]{3f} - (\zeta + \Omega - 1.1165) + 0.7960)f + (\Omega - 1.0902) \times 0.8940}.$$

$$^b p_4 \equiv \sqrt[3]{f} [(-3.0394 \Omega + 4.4939)(\sqrt{\Omega} - f) + 0.9829] - \frac{0.3806}{\Omega}.$$

$$^c p_5 \equiv 1.0747 \sqrt[3]{f} + \frac{\Omega - 0.8932}{(f + \Omega) \sqrt[3]{f - \zeta + 1.0320 - \Omega}}.$$

TABLE V. Reference solutions and SR-discovered equations for forced Duffing oscillators. Target variable is the response amplitude A ; F_0 included explicitly as an SR input.

Case	Method	Expression	SR-discovered equation	R^2	MSE
x^3	Single-harm. balance	$(\alpha - \omega_{\text{ext}}^2 + \frac{3}{4}\beta A^2)^2 + (\delta\omega_{\text{ext}})^2 = F_0^2/A^2$	$A = 2.703\sqrt{F_0} - \sqrt[3]{N_3}/-0.681^a$	0.834	0.190
x^4	Multiharm. balance	No closed form (Sec. II B 2)	$A = \sqrt{F_0/\sqrt[3]{\beta}} - \sqrt[3]{\omega_{\text{ext}} - 0.978(\omega_{\text{ext}} - \alpha)}$	0.532	0.150
x^5	Projection	Num	$A = (1.334 - \beta)[\sqrt[3]{\alpha F_0} - (\omega_{\text{ext}} - \sqrt[3]{\alpha}) + (1.409 - \frac{\alpha}{\omega_{\text{ext}}})]$	0.768	0.089
x^5	Single-harm. balance	$(\alpha - \omega_{\text{ext}}^2 + \frac{5}{8}\beta A^4)^2 + (\delta\omega_{\text{ext}})^2 = F_0^2/A^2$	$A = \sqrt{N_5 F_0}^b$	0.718	0.086

$$^a N_3 \equiv (F_0 - (\omega_{\text{ext}}^2 - \alpha))[(2\omega_{\text{ext}} - \alpha) - \sqrt[3]{\beta/0.058}].$$

$$^b N_5 \equiv \sqrt[3]{(0.881 - (\omega_{\text{ext}} - \alpha))/\beta} [\omega_{\text{ext}} - (\alpha - 0.357 - \omega_{\text{ext}})].$$

the asymmetric restoring force precludes a simple closed-form resonance relation, the reference is obtained via a multiharmonic balance solved numerically.

The cubic forced case yields $R^2 = 0.834$, the quartic forced case $R^2 = 0.532$, and the quintic forced case $R^2 = 0.718$ (Table V). Figure 5 compares the reference and SR-discovered response curves $A(\omega_{\text{ext}})$ at fixed $\alpha = 1$, $\beta = 0.15$, $\delta = 0.10$, and $F_0 = 0.5$. Unlike the free-oscillator results, these values do not approach machine precision, for the following reason: the response amplitude A is defined implicitly as the root of a cubic (respectively, quintic) polynomial in A^2 [Eq. (21) and its analogs], a relation that in general requires operators beyond `sqrt` and `cbirt`—and, in the multivalued regime near resonance, requires trigonometric functions even in principle. The discovered expressions retain part of the expected physical structure, though only partially. An explicit dependence on F_0 appears in all three discovered expressions, but the quadratic detuning $\omega_{\text{ext}}^2 - \alpha$ appears only in the cubic row of Table V. The quartic and quintic rows contain instead the combinations $\omega_{\text{ext}} - \alpha$ and $\omega_{\text{ext}} - \sqrt[3]{\alpha}$, which are linear in the drive frequency. Over the sampled window, the linear and quadratic forms are strongly correlated, so a search that is not told the correct grouping can substitute one for the other at little cost in mean squared error, and the appearance of a detuninglike combination should not be read as recovery of the detuning itself.

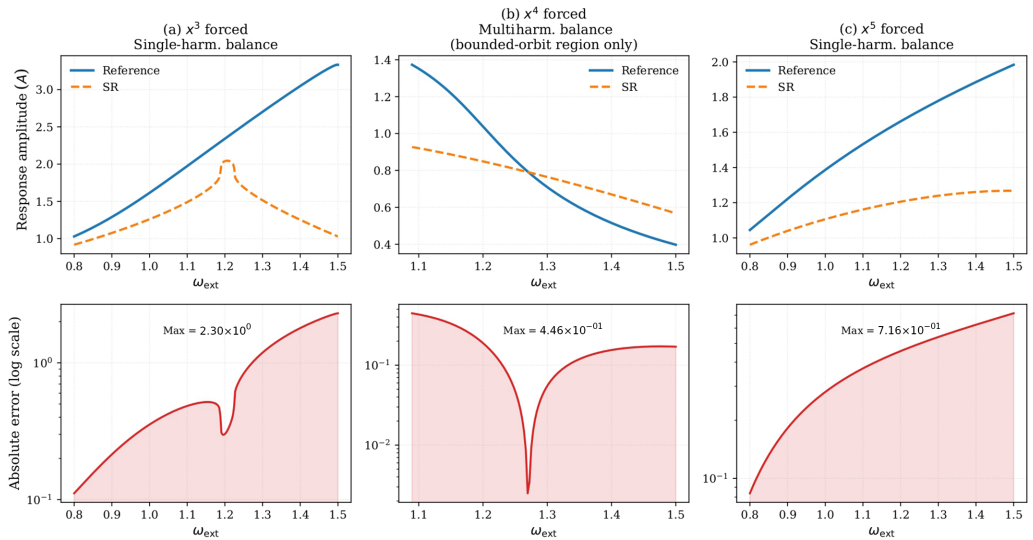


FIG. 5. Response amplitude A as a function of the drive frequency ω_{ext} , with $\alpha = 1$, $\beta = 0.15$, $\delta = 0.10$, and $F_0 = 0.5$ fixed, for forced Duffing oscillators. Top row: reference (solid) and SR-discovered (dashed) curves for (a) cubic (single-harmonic balance), (b) quartic (multiharmonic balance, restricted to the region admitting a bounded periodic solution), and (c) quintic (single-harmonic balance).

Table V presents the SR-discovered expressions for all forced configurations alongside the reference solutions. The reference column follows the frequency-response formulation of Eq. (21) and its quartic and quintic analogs. The structure of the SR expressions is qualitatively different from the free case: SR finds expressions that involve ω_{ext} explicitly and combine it with $\sqrt{\alpha}$ through rational functions. The numerical constants returned in these expressions (for example, 2.703 and 0.978) are in-sample fitted values without independent physical meaning and should not be overinterpreted. The limitation is that the specific functional form of the resonance combination cannot be fully captured by the algebraic operator set available to SR, which excludes the trigonometric functions that appear in exact resonance conditions. F_0 is included explicitly as an input variable; SR selects it in all three discovered expressions (Table V). Mean $\pm$ std metrics over multiple seeds are reported for a pilot configuration in the Appendix, with extension to all forced configurations identified as a direction for future work.

IV. DISCUSSION

A. From reconstruction to approximation: What the comparison reveals

The most striking pattern in the results is the sharp contrast between the free and forced configurations. For free oscillators with analytical reference solutions, SR returns the identical expression—to numerical precision—that was used to generate the labels. Care should be taken regarding how to read this. Because the LP, HB, and first-harmonic-balance labels are produced by evaluating those same closed-form expressions, the result demonstrates that the algebraic search can reconstruct the generating formula from a finite sample of input-output pairs. That is a genuine and useful capability for formula compression and rediscovery; it is not, by itself, evidence that SR has discovered the underlying physics, since the physics was already encoded in the labels. The fact that SR identifies the rational coefficients $\frac{3}{4}$ and $\frac{5}{8}$ exactly is a meaningful result about the reach of the algebraic search, and it motivates the application of SR to cases where no closed form is available—of which the dynamics-based quintic RK45 case is the one example studied here.

The amplitude study of Sec. III B confirms that the A^2 and A^4 scalings central to this work are genuinely recoverable by SR, not artifacts of the fixed-amplitude normalization used elsewhere in this article. The quartic case additionally reveals that amplitude enters not only as a functional-form parameter but as a bound on physical validity: Because the quartic restoring force is not odd, the free potential is asymmetric and unbounded from below, so sufficiently large amplitudes admit no periodic solution at all. This is a property of the physical system, not of the symbolic-regression protocol, and constrains the applicability of the quartic configuration independently of the algebraic operator set available to SR.

The forced configurations reveal the boundary of this capability. For the cubic case, $R^2 = 0.834$; for the quartic case, $R^2 = 0.532$; and for the quintic case, $R^2 = 0.718$ (single-harmonic balance) and $R^2 = 0.768$ (RK45+FFT). The physics behind this degradation is transparent: Unlike the free case, where ω is given by a simple closed-form algebraic expression, the forced-case response amplitude A is defined only implicitly, as the root of a cubic (or quintic) polynomial in A^2 . Even with cbrt included in the operator set, this implicit relation—and, in particular, its multivalued, hysteretic branch structure near resonance—generally requires trigonometric functions to express in closed form (the classical *casus irreducibilis* of the cubic formula). The algebraic operator set available to SR is well matched to free-oscillator formula reconstruction; it is structurally insufficient for the full resonance structure of forced systems. The searches behind these numbers were run over five raw parameters that, as shown in Sec. II D, encode only three independent groups, so a degradation produced by the redundant parametrization would be indistinguishable in aggregate metrics from one produced by a missing operator. Rerunning the search in (Ω, ζ, f) (Sec. III C) separates the two effects: accuracy rises to $R^2 = 0.962, 0.887, \text{ and } 0.969$, showing that a substantial part of the degradation is attributable to the redundant parametrization, while the residual gap to free-oscillator precision is consistent with a more modest structural limitation of the algebraic operator set. The discovered expressions nonetheless correctly identify the dominant physical structure only in part:

An explicit dependence on F_0 appears in all three cases, whereas the quadratic detuning $\omega_{\text{ext}}^2 - \alpha$ is recovered only for the cubic oscillator, the quartic and quintic expressions returning combinations linear in ω_{ext} instead.

The RK45-based quintic forced case, obtained by direct numerical integration without the single-harmonic approximation, achieves $R^2 = 0.768$ ($\text{MSE} = 0.089$)—modestly higher than the closed-form quintic case ($R^2 = 0.718$). Unlike the closed-form reference, which required excluding 133/1000 samples in the hysteretic multiroot region, the RK45 reference converged for all 1000 sampled configurations, suggesting that the numerically exact dynamics may present a more uniformly learnable target than the single-harmonic approximation near resonance.

B. Physical interpretation of the discovered expressions

The expressions recovered by SR for the free oscillators are physically interpretable, with the caveat that they are reconstructions of known relations. The cubic expression $\omega = \sqrt{\alpha + \frac{3}{4}\beta}$ can be read directly as the natural frequency of a harmonic oscillator with an effective stiffness $\alpha_{\text{eff}} = \alpha + \frac{3}{4}\beta$, where the factor $\frac{3}{4}$ encodes the average effect of the cubic term over one period of oscillation—precisely what LP computes when it eliminates secular terms at first order. The quintic (first-harmonic-balance) expression shows a fourth-power amplitude dependence in the effective stiffness, consistent with the physical expectation that a fifth-power force produces a stronger amplitude sensitivity than cubic forces. The quartic expression, obtained from the multiharmonic-balance reference of Sec. II B 2, does not admit as simple a physical reading as the cubic and quintic cases, reflecting the asymmetry of the underlying potential.

C. Limitations and scope of the results

Several limitations bound the scope of the conclusions that can be drawn from this study. First and foremost, for the LP, HB, and first-harmonic-balance cases the SR targets are the analytical formulas themselves, so the free-oscillator results are a benchmark of formula rediscovery rather than a demonstration of discovery from oscillator dynamics. The dynamics-based test is limited to the quintic RK45 case. The training and evaluation datasets were generated within bounded parameter spaces ($\alpha \in [0.5, 2.0]$, $\beta \in [0.01, 0.3]$, $\delta \in [0.01, 0.2]$, $F_0 \in [0.1, 1.5]$), and the discovered expressions should not be assumed to extrapolate reliably beyond these ranges without additional validation. The amplitude study reported in Sec. III B addresses this limitation directly, confirming exact recovery of the A^2 and A^4 scalings for the cubic and quintic cases. For the quartic case, the validity range in A is itself bounded by the asymmetry of the potential (Sec. II E), a constraint intrinsic to this configuration rather than an artifact of the symbolic-regression protocol. The reference datasets were generated from analytical formulas or high-precision numerical integrations without measurement noise. A preliminary noise study on the cubic free oscillator (Appendix 3) shows that SR recovers the correct coefficient at 1% and 2% relative noise but not at 5%. Finally, a pilot run-to-run variability study on the same case (Appendix 4) shows the reported metrics are stable across 20 independent seeds ($R^2 = 1.000\,000 \pm 0.000\,000$); extending both the noise-robustness and the variability studies to all configurations is left for future work.

The operator set chosen for PySR excludes trigonometric and exponential functions. For the free-oscillator cases, this restriction is appropriate because the reference solutions are algebraic. For the forced cases, we extended the operator set to include cbqrt (Sec. II C), which improved but did not fully close the gap to free-oscillator accuracy; as discussed above, the residual limitation is structural (the casus irreducibilis of the underlying cubic/quintic root), and closing it further would require trigonometric operators, at the cost of a substantially larger and more computationally demanding search space.

Finally, this study targets a single scalar output variable per configuration—the oscillation frequency ω for the free cases, and the steady-state response amplitude A for the forced cases. The full characterization of a forced nonlinear oscillator additionally includes phase and transient dynamics, neither of which is addressed here.

V. CONCLUSIONS

This work set out to determine whether symbolic regression can recover or approximate the amplitude-frequency relations of the Duffing oscillator across multiple nonlinearity orders and dynamical regimes, without being provided the functional form of the solution. The answer, as the results demonstrate, depends critically on the regime considered—and the boundary between what SR can and cannot do turns out to be physically meaningful.

For free oscillators with known analytical solutions, SR reconstructs the reference expressions to numerical precision. The expressions returned by the evolutionary search are identical—to numerical precision—to those used to generate the training labels via LP perturbation theory, HB, and first-harmonic balance. Because the labels are the same closed forms, this should be read as symbolic rediscovery and compression, not as independent discovery of the physics; the algebraic structure was present in the data by construction.

The forced oscillator results reveal where this capacity breaks down. Accuracy degrades measurably when external forcing introduces resonance phenomena, and the degradation is directly traceable to a physical cause: The frequency response of a driven nonlinear oscillator near resonance involves functional dependencies—cancellations, near-resonance corrections, nonlinear mode coupling—that the algebraic operator set available to SR cannot fully represent. Including operator sets that incorporate rational functions of frequency ratios, or augmenting SR with physics-informed constraints on the resonance condition, are natural directions for improvement.

Taken together, these results position SR as a practical complement to classical analytical methods in nonlinear dynamics. Its most productive role is not as a replacement for perturbation theory or HB—those methods remain superior within their validity ranges. Rather, SR is most valuable in situations where classical methods reach their limits: higher nonlinearity orders without known closed-form solutions, parameter regimes where perturbative expansions diverge, or applied settings where derivation from first principles is not feasible.

The next natural step is to evaluate SR on data that come from the physical system rather than from analytical approximations of it. Experimental measurements of a Duffing-type oscillator—for instance, a clamped beam, a nonlinear electronic circuit, or a microelectromechanical resonator—would provide training data with realistic noise levels and without the restrictions imposed by the parameter validity of LP or HB. Together with explicit amplitude variation, held-out extrapolation tests, and a quantification of run-to-run variability, this is the validation that would be needed to move from formula rediscovery toward genuine discovery—and it is the direction for future work that our group will pursue.

ACKNOWLEDGMENTS

The authors thank the Universidad Michoacana de San Nicolás de Hidalgo for institutional support. I.L.-S. and A.R., in particular, acknowledge financial support from CIC-UMSNH under Grant No. 18371.

DATA AVAILABILITY

The datasets generated during this study, the data-generation scripts, and the PySR training and plotting scripts are available at Ref. [33].

APPENDIX: ADDITIONAL VALIDATION

1. Residual analysis

Beyond the aggregate R^2 and MSE reported in Tables II–V, we examine the prediction residual (reference minus SR-discovered value) as a function of the individual system parameters. Figure 6 shows this for the free oscillators; for the cubic and quintic cases, the residuals are confined to floating-point rounding ($\sim 10^{-16}$) across the full range of α , β , and A , consistent with the exact recovery reported in Table III. For the quartic case, the residual grows systematically with A ,

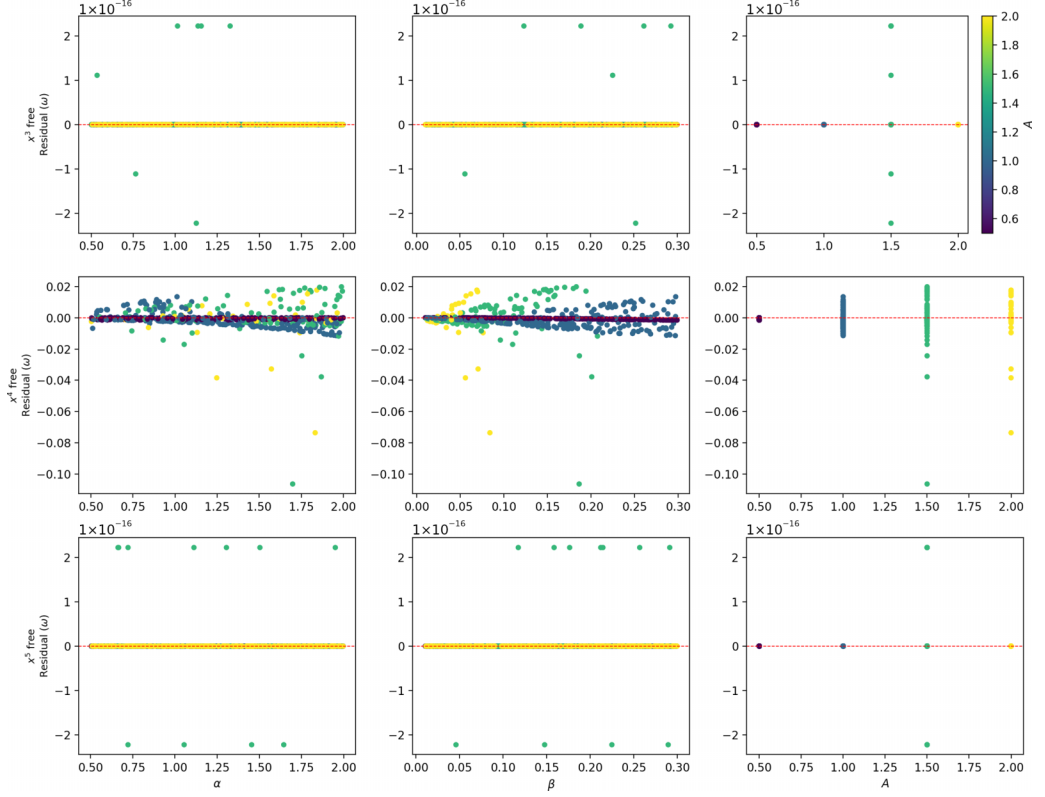


FIG. 6. Prediction residual (reference – SR) for free Duffing oscillators as a function of α , β , and A , colored by A . Top row: cubic. Middle row: quartic. Bottom row: quintic.

concentrating near the upper edge of the bounded-orbit region identified in Sec. III B, indicating that the approximation degrades as the amplitude approaches the escape threshold of the asymmetric potential rather than uniformly across the parameter space.

Figure 7 shows the corresponding analysis for the forced oscillators, including the residual as a function of the detuning $\omega_{\text{ext}} - \sqrt{\alpha}$. In all three cases, the residual grows near zero detuning (i.e., near resonance) and is comparatively small away from it, consistent with the interpretation that the implicit, multivalued structure of the resonance curve is hardest to capture precisely in the region where the competition between the natural and driving frequencies is strongest.

2. Extrapolation beyond the training range

To test whether the SR-discovered relations generalize beyond the parameter box used for training, we retrained the cubic free-oscillator case using only $\beta \in [0.01, 0.15]$ and evaluated the resulting expression on a disjoint held-out range, $\beta \in [0.15, 0.30]$, never seen during training. SR recovers $\omega = \sqrt{\alpha} + 0.7500\beta$, with $R^2 = 1.000\,000$ ($\text{MSE} = 1.9 \times 10^{-15}$) on the extrapolated range, indicating that the discovered relation reflects the underlying physical law rather than a local interpolation confined to the training domain.

3. Robustness to noise

We further tested robustness by adding Gaussian relative noise (1%, 2%, and 5% of the signal amplitude) to the reference frequencies before retraining. At 1% and 2% noise, SR recovers a

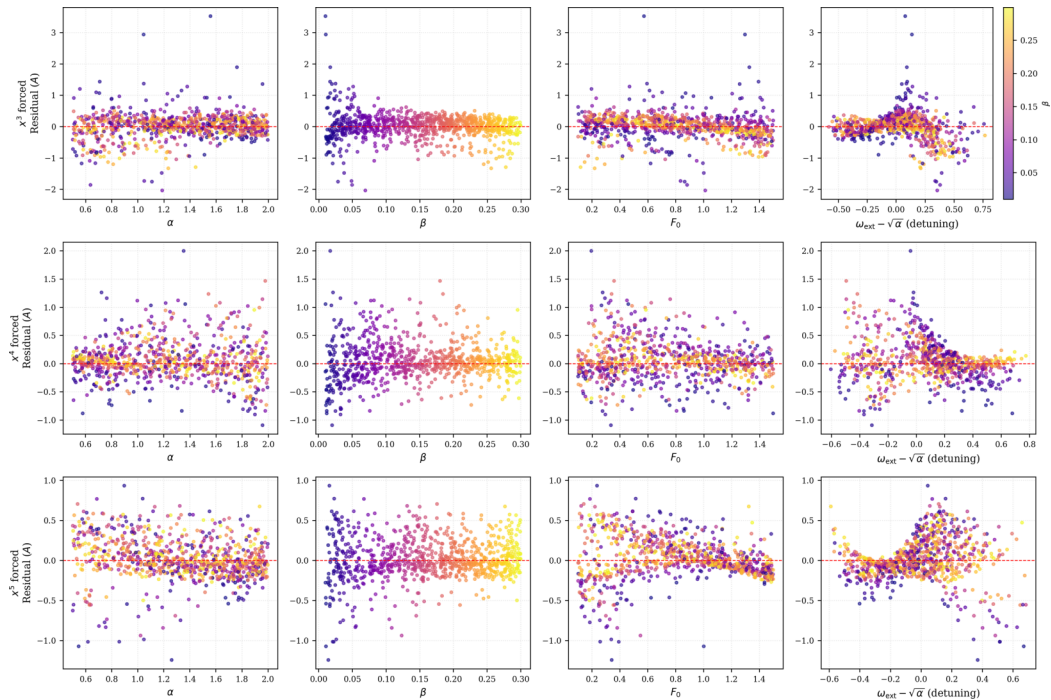


FIG. 7. Prediction residual (reference—SR) for forced Duffing oscillators as a function of α , β , F_0 , and the detuning $\omega_{\text{ext}} - \sqrt{\alpha}$, colored by β . Top row: cubic. Middle row: quartic. Bottom row: quintic.

coefficient close to the true value (0.7474 versus the exact $3/4 = 0.75$), with $R^2 = 0.996$ and $R^2 = 0.985$, respectively. At 5% noise, however, SR no longer recovers the correct functional form: The discovered expression collapses to $\omega = \sqrt{\alpha + \beta}$, replacing the physical coefficient $3/4$ with an implicit coefficient of 1. Although this expression still achieves a numerically reasonable $R^2 = 0.895$, it is qualitatively incorrect—illustrating that a high R^2 alone does not guarantee that SR has recovered the correct physics, and that the present pipeline is robust to noise below approximately 2% but not at 5%.

4. Run-to-run variability

Because PySR’s evolutionary search is stochastic, we repeated the training of the cubic free-oscillator case 20 times with independent random seeds. All 20 runs achieved $R^2 = 1.000\,000$ (standard deviation 0.000 000) and recovered an expression algebraically equivalent to $\omega = \sqrt{\alpha + 0.75\beta}$ in every case, though the specific algebraic form varied (e.g., $\sqrt{0.5625\beta}$, $\beta/\sqrt{1.7778}$, or small higher-complexity variants that reduce to the same coefficient). This indicates that, for this configuration, the reported expression is a highly stable outcome of the search procedure rather than a favorable single run.

-
- [1] G. Duffing, *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung* (Vieweg, Braunschweig, 1918).
 - [2] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Applied Mathematical Sciences (Springer-Verlag, New York, 1983), Vol. 42.

- [3] P. Kumi, S. A. Martin, V. V. Yakovlev, M. S. Hilario, B. W. Hoff, and I. M. Rittersdorf, Analytical solution of the Duffing equation, [COMPEL](#) **39**, 481 (2020).
- [4] S. Nagarajaiah, A. M. Reinhorn, and M. C. Constantinou, Nonlinear dynamic analysis of three-dimensional base isolated structures, [J. Struct. Eng.](#) **117**, 2035 (1991).
- [5] R. Rana and T. T. Soong, Parametric study and simplified design of tuned mass dampers, [Eng. Struct.](#) **20**, 193 (1998).
- [6] R. A. Ibrahim, Friction-induced vibration, chatter, squeal, and chaos. Part II: Dynamics and modeling, [Appl. Mech. Rev.](#) **47**, 227 (1994).
- [7] A. Lindstedt, Ueber die integration einer für die störungstheorie wichtigen differentialgleichung, [Astron. Nachr.](#) **103**, 211 (1882).
- [8] H. Poincaré, *New Methods of Celestial Mechanics, History of Modern Physics and Astronomy* (American Institute of Physics, New York, 1993), Vol. 2.
- [9] E. J. Hinch, *Perturbation Methods* (Cambridge University Press, Cambridge, 1991).
- [10] C. M. Bender, K. A. Milton, S. S. Pinsky, and L. M. Simmons, Jr., A new perturbative approach to nonlinear problems, [J. Math. Phys.](#) **30**, 1447 (1989).
- [11] A. H. Nayfeh and D. T. Mook, *Nonlinear Oscillations* (Wiley-VCH, New York, 1995).
- [12] J.-H. He, Preliminary report on the energy balance for nonlinear oscillations, [Mech. Res. Commun.](#) **29**, 107 (2002).
- [13] P. Amore and A. Aranda, Presenting a new method for the solution of nonlinear problems, [Phys. Lett. A](#) **316**, 218 (2003).
- [14] P. Amore and A. Aranda, Improved Lindstedt–Poincaré method for the solution of nonlinear problems, [J. Sound Vib.](#) **283**, 1115 (2005).
- [15] C.-S. Liu and Y.-W. Chen, A simplified Lindstedt–Poincaré method for saving computational cost to determine higher order nonlinear free vibrations, [Mathematics](#) **9**, 3070 (2021).
- [16] R. E. Mickens, Harmonic balance and iteration calculations of periodic solutions to $\ddot{y} + y^{-1} = 0$, [J. Sound Vib.](#) **306**, 968 (2007).
- [17] A. Beléndez, D. I. Méndez, E. Fernández, S. Marini, and I. Pascual, Exact solution for the unforced Duffing oscillator with cubic and quintic nonlinearities, [Phys. Lett. A](#) **373**, 2805 (2009).
- [18] J.-H. He, Variational principles for some nonlinear partial differential equations with variable coefficients, [Chaos Solitons Fractals](#) **19**, 847 (2004).
- [19] J.-H. He, Linearized perturbation technique and its applications to strongly nonlinear oscillators, [Comput. Math. Appl.](#) **45**, 1 (2003).
- [20] J.-H. He, The homotopy perturbation method for nonlinear oscillators with discontinuities, [Appl. Math. Comput.](#) **151**, 287 (2004).
- [21] P. Amore and F. M. Fernandez, He’s amazing calculations with the Ritz method, [arXiv:0901.4660](#).
- [22] F. M. Fernandez, Homotopy perturbation method: When infinity equals five, [arXiv:0810.3318](#).
- [23] F. M. Fernandez, On the application of homotopy perturbation method to differential equations, [arXiv:0808.2078](#).
- [24] B. R. Brum, L. Lober, I. Previdelli, and F. A. Rodrigues, Discovering equations from data: Symbolic regression in dynamical systems, [J. Phys. Complex.](#) **7**, 012001 (2026).
- [25] Y. Matsubara, N. Chiba, R. Igarashi, T. Taniai, and Y. Ushiku, Rethinking symbolic regression datasets and benchmarks for scientific discovery, [J. Data-centric Mach. Learn. Res.](#) (2024).
- [26] D. J. Bartlett, H. Desmond, P. G. Ferreira, and G. Kronberger, Introduction to the special issue on symbolic regression in the physical sciences, [Phil. Trans. R. Soc. A](#) **384**, 20240600 (2026).
- [27] S. Kim, P. Y. Lu, S. Mukherjee, M. Gilbert, L. Jing, V. Ceperic, and M. Soljačić, Integration of neural network-based symbolic regression in deep learning for scientific discovery, [IEEE Trans. Neural Networks Learn. Syst.](#) **32**, 4166 (2021).
- [28] T. Pirbodaghi, S. H. Hoseini, M. T. Ahmadian, and G. H. Farrahi, Duffing equations with cubic and quintic nonlinearities, [Comput. Math. Appl.](#) **59**, 3095 (2010).
- [29] O. M. Al-Hababeh, M. Al-Saqqah, M. Safi, and T. A. Khater, An analytical approximate technique for solving cubic–quintic Duffing oscillator, [Alexandria Eng. J.](#) **55**, 1347 (2016).

- [30] M. S. Al-Amry, A. Maher, and H. M. El-Hawary, A semi-analytical iterative technique for solving Duffing equations, *Int. J. Pure Appl. Math.* **108**, 751 (2016).
- [31] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems*, 3rd ed. (Springer, Berlin, 2008).
- [32] M. Cranmer, PySR: Fast & parallelized symbolic regression in Python and Julia, GitHub, 2023, <https://github.com/MilesCranmer/PySR>.
- [33] I. Luviano-Soto, J. P. Pérez Aguilar, and A. Raya, Duffing_SR: Data and code for symbolic rediscovery of amplitude-frequency relations in Duffing oscillators, GitHub, 2026, https://github.com/Itzel-LS/Duffing_SR.