JOURNAL OF THEORETICAL AND APPLIED MECHANICS
62, 2, pp. 435-455, Warsaw 2024
https://doi.org/10.15632/jtam-pl/186718

ERROR MEASURES AND SOLUTION ARTIFACTS OF THE HARMONIC BALANCE METHOD ON THE EXAMPLE OF THE SOFTENING DUFFING OSCILLATOR[1]

Hannes Dänschel
Technische Universität Berlin, Institut für Mechanik, Berlin, Germany
corresponding author Hannes Dänschel, e-mail: hannes.daenschel@tu-berlin.de

Lukas Lentz
Hochschule Trier, Institut für Betriebs- und Technologiemanagement, Trier, Germany

Utz von Wagner
Technische Universität Berlin, Institut für Mechanik, Berlin, Germany

The Harmonic Balance Method (HBM) is one of the most often applied semi-analytic approximation methods in nonlinear dynamics. In earlier publications, the two coauthors already observed for the softening Duffing oscillator and other systems that especially low ansatz order HBM solutions may contain larger errors for some solution branches, and called this artifacts. In the present work, this problem is studied systematically with a new implementation of the method and applied again to the example of the softening Duffing oscillator. In conjunction with a mathematical definition for HBM artifacts we discuss and present possible *a posteriori* and *a priori* HBM error measures.

Keywords: Harmonic Balance Method (HBM), artifact solutions, softening Duffing oscillator

1. Introduction

In 1918, German engineer Georg Duffing published his seminal work (Duffing, 1918) exploring the dynamics of forced nonlinear oscillations. The considered system, nowadays called the Duffing oscillator, is represented by the second-order differential equation

$$x''(t) + \delta x'(t) + \alpha x(t) + \beta x^2(t) + \gamma x^3(t) = \widehat{u}\cos(\Omega t) \tag{1.1}$$

This equation describes the displacement x of a system subjected to a linear damping force (parameter δ) as well as linear (α), quadratic (β) and cubic (γ) restoring forces, along with a harmonic driving force defined by its amplitude $\widehat{u}$ and frequency Ω. The first and second order time derivatives of x are denoted by x' and x'', respectively. Thereby and in the following, all parameters and variables, including time, are considered to be dimensionless. Whilst Eq. (1.1) may be interpreted as a nonlinear extension of the standard harmonic oscillator (recovered when $\beta = 0$ and $\gamma = 0$), its dynamics is vastly more complex (Ueda, 1991). In the book by Kovacic and Brennan (2011), many details about history, applications, solution methods and phenomena of the Duffing equation can be found. Due to its simple form and the ability to display a plethora of nonlinear phenomena like multiple coexisting solutions, subharmonic and superharmonic components in the system response, bifurcations (Holmes and Rand, 1976) as well as chaotic behavior for certain parameter choices (Novak and Frenlich, 1982), the Duffing

[1] Paper presented during PCM-CMM 2023, Gliwice, Poland

oscillator rapidly became a model of significant theoretical and practical interest and was covered in introductory textbooks on nonlinear dynamics, e.g. Nayfeh and Mook (1979) or Strogatz (1994).

In the context of analyzing the Duffing oscillator, it is important to note that closed-form solutions in general are not available. Therefore, it becomes necessary to apply alternative methods, such as numerical integration or approximate analytical techniques, to explore the system behavior. While applying numerical integration, one starts from a given set of initial conditions and may end up in an asymptotically stable stationary periodic solution, quasiperiodic, chaotic or in general irregular behavior, or drifting to $\pm\infty$. Compared to this, semi-analytic approximation methods like Lindstedt-Poincaré perturbation analysis or the Harmonic Balance Method (HBM) applied here, calculate stationary solutions without transients, while other methods like Multiple (Time) Scales are also able to calculate transient behavior (Hagedorn, 1981). Nonlinear systems may have multiple stationary periodic solutions (some of them being stable and some unstable) for one excitation frequency, which is easily recognized, when methods are directly applied to calculate them. To get the variety of multiple stationary solutions while applying numerical integration, initial conditions have to be varied.

In scenarios where the focus is solely on stationary periodic solutions, the HBM can be applied with great benefit. It was introduced in Urabe (1965) as an application of the Galerkin method with harmonic ansatz functions, and nowadays is widely known under the name HBM. In the HBM, a finite Fourier series representation is used to approximate the exact solution of the nonlinear system under consideration.

In general, in the HBM, the accuracy of a solution can be improved by increasing the truncation order of the series. Below certain truncation orders, the solution behavior may differ largely from the real solutions. Some examples of such anomalous solutions produced by the HBM at lower approximation orders are e.g. documented in previous articles of the coauthors (von Wagner and Lentz, 2016, 2018, 2019) or in the book of Krack and Gross (2019). Corresponding error estimates were first derived by Urabe (1965) and the associated error bounds were later improved upon by García-Saldaña and Gasull (2013), Kogelbauer Brennan (2021) and Woiwode and Krack (2023). For further considerations, we refer to the book by Krack and Gross (2019).

As considered e.g. in von Wagner and Lentz (2016, 2018), applying the HBM to the softening Duffing oscillator results in the occurence of high amplitude solutions for low excitation frequencies with the shape of a "nose" with large residua for small ansatz orders. The increasing of the ansatz order results in a significant change of the shape and frequency range of occurence of these solutions. This was called *artifact behavior* by the authors but a comprehensive investigation of a rigorous definition and possible critera of their *a priori* or *a posteriori* detection is yet missing. Therefore, in the present work this problem is studied systematically by discussing error measures for the error in the HBM. Hereby, the object of study is again the softening Duffing oscillator where we restrict the problem to one with solutions with a vanishing mean value with the consequence of a vanishing constant, and even terms in the HBM ansatz. As shown in von Wagner and Lentz (2016, 2018) solutions with the non-vanishing mean value exist for the softening Duffing oscillator but are inconspicuous with respect to artifacts. Instead, we consider mainly the already mentioned "nose" shaped solution branch.

The paper is structured as follows. In Section 2, we provide a description of the HBM and its implementation performed by the first author. A test case is considered showing the problems of HBM as discussed in the following. In Section 3, a definition for the artifact solution is provided and then error measures based on numerical and geometrical considerations are introduced, and in Section 4 applied to the considered case of the softening Duffing oscillator. The paper ends with corresponding conclusions in Section 5.

2. Harmonic Balance Method (HBM)

In this Section, we present the theoretical basics required to obtain the frequency response of the softening Duffing system (1.1) by means of the HBM and numerical continuation methods. Let a range of excitation frequencies $\varOmega \in \mathbb{F} := [\underline{\varOmega}, \overline{\varOmega}]$, the system parameters $\delta, \alpha, \beta, \gamma$ and the harmonic excitation $\widehat{u}\cos(\varOmega t)$ be given. We denote the system frequency response as the set

$$\varGamma(\mathbb{F}) := \Big\{(\varOmega, \|x\|) \in \mathbb{R}^2 \mid x(t) = x(t+T) \text{ solves (1.1)},\ \ \varOmega \in \mathbb{F}\Big\} \tag{2.1}$$

with the period $T = 2\pi/\varOmega$ and a later to be specified solution amplitude $\|x\|$. Computing an approximation of (2.1) requires to find approximations to x by means of the HBM over samples of the frequency range $\mathbb{F}$.

2.1. Fundamentals

The HBM is a mean weighted residual method that comprises of two approximation steps. As preliminaries, consider a time domain $\mathbb{T} := [0, T]$, an *ansatz* or *approximation order* $n \in \mathbb{N}$ as well as a real Fourier space

$$\mathcal{F}_n(\mathbb{T}, \varOmega) := \Big\{x_n \ :\ \mathbb{T} \to \mathbb{R} \mid x_n(t) = c_0\phi_0(t) + \sum_{j=1}^{n} \big(c_{2j-1}\phi_{2j-1}(t) + c_{2j}\phi_{2j}(t)\big)\Big\} \tag{2.2}$$

with the basis functions $\phi_0(t) = 1$, $\phi_{2j-1}(t) = \cos(j\varOmega t)$ and $\phi_{2j}(t) = \sin(j\varOmega t)$, $j = 1, \ldots, n$. For convenience, we define the vector of Fourier coefficients $\mathbf{c}_n := [c_j]_{j=0}^{2n} \in \mathbb{R}^{2n+1}$ which also allows to identify $x_n \in \mathcal{F}_n$ with $\mathbf{c}_n \in \mathbb{R}^{2n+1}$. Finally, consider the residual function of the Duffing system (1.1)

$$r(t, x) = x''(t) + \delta x'(t) + \alpha x(t) + \beta x^2(t) + \gamma x^3(t) - \widehat{u}\cos(\varOmega t) = 0 \tag{2.3}$$

The first step of the HBM is inserting the ansatz $x \approx x_n \in \mathcal{F}_n$ into the Duffing residual from which we obtain $r(t, x) \approx r(t, x_n) = r(t, \mathbf{c}_n)$. Since residual (2.3) is a third-degree polynomial in the trigonometric polynomial x_n, and the excitation $u \in \mathcal{F}_1$ is a simple harmonic, the convolution theorem yields that $r(t, \mathbf{c}_n)$ can be expressed as a truncated Fourier series of the order $3n$, i.e.

$$r(t, \mathbf{c}_n) = R_n(t, \mathbf{c}_n) := R_0(\mathbf{c}_n) + \sum_{j=1}^{3n} \big(R_{2j-1}(\mathbf{c}_n)\cos(j\varOmega t) + R_{2j}(\mathbf{c}_n)\sin(j\varOmega t)\big) \tag{2.4}$$

The second approximation step of the HBM requires that the first $2n + 1$ Fourier coefficients of (2.4) vanish, i.e. $R_i(\mathbf{c}_n) = 0$. This is imposed by the $2n + 1$ conditions

$$\langle R_n(\cdot, \mathbf{c}_n)\phi_i\rangle_{\mathcal{F}_n} = \frac{1}{T}\int_0^T R_n(t, \mathbf{c}_n)\phi_i(t)\, dt = 0 \qquad \forall i = 0, 1, \ldots, 2n \tag{2.5}$$

In fact, by the definition of Fourier coefficients the equality $R_i(\mathbf{c}_n) = \langle R_n(\cdot, \mathbf{c}_n)\phi_i\rangle_{\mathcal{F}_n} = 0$ holds for all $i = 0, 1, \ldots, 2n$ (Herman 2016). The conditions (2.5) basically ensure that the error introduced in the residual only comprises of higher order harmonics since for all $t \in \mathbb{T}$ it holds that

$$\begin{aligned} R_n(t, \mathbf{c}_n) = {} & \underbrace{R_0(\mathbf{c}_n) + \sum_{j=1}^{n} \big(R_{2j-1}(\mathbf{c}_n)\cos(j\varOmega t) + R_{2j}(\mathbf{c}_n)\sin(j\varOmega t)\big)}_{=0} \\ & + \sum_{j=n+1}^{3n} \big(R_{2j-1}(\mathbf{c}_n)\cos(j\varOmega t) + R_{2j}(\mathbf{c}_n)\sin(j\varOmega t)\big) \end{aligned} \tag{2.6}$$

The $2n+1$ equations (2.5) define the algebraic equation system

$$F_n(\mathbf{c}_n, \mathit{\Omega}) := [R_j(\mathbf{c}_n, \mathit{\Omega})]_{j=0}^{2n} = 0 \tag{2.7}$$

that can be solved for $\mathbf{c}_n$. In order to compute the frequency response $\mathit{\Gamma}(\mathbb{F})$, we explicitly include $\mathit{\Omega}$ as a parameter in (2.7). Finally, if $\mathbf{c}_n$ solves (2.7) we refer to $\mathbf{c}_n$, x_n and R_n as *HBM coefficient vector*, *HBM solution* and *HBM residual*, respectively.

2.2. Solvers

In the following, we discuss how to compute the frequency response $\mathit{\Gamma}(\mathbb{F})$ by solving the parameter-dependent algebraic system (2.7).

2.2.1. Determining Fourier coefficients

Solving the algebraic system (2.7) in order to obtain the Fourier coefficients $\mathbf{c}_n$ one requires to evaluate F_n for which the integrals (2.5) must be determined. Since the Duffing nonlinearities are polynomials in x, the evaluation can be done by obtaining the integral closed form as well as via the discrete convolution or discrete Fourier transform (Krack and Gross, 2019; Woiwode *et al.*, 2020). However, since we also want to investigate the influence of the algebraic structure on the solution artifacts we opted for an equivalent approach of obtaining an algebraic expression of the truncated Fourier series of the residual (2.4) in the Fourier coefficients $\mathbf{c}_n$. The algorithmic implementation used in this work is a pure Python implementation without the use of computer algebra tools. A publication about the associated theoretical details as well as the source code is planned.

2.2.2. Newton's method

The next step is to solve (2.7). Let $\mathit{\Omega} \in \mathbb{F}$ be fixed, an approximation order $n \in \mathbb{N}$ be given and assume the Jacobian of F w.r.t. $\mathbf{c}_n$ is regular. Then, for any initial guess $\mathbf{c}_n^0 \in \mathbb{R}^{2n+1}$ "close enough" to $\mathbf{c}_n$, the algebraic system (2.7) can be solved iteratively by Newton's method for an approximated solution $\mathbf{c}_n \approx \mathbf{c}_n^k \in \mathbb{R}^{2n+1}$ subject to a prescribed iteration error tolerance $\varepsilon > 0$ s.t. for some $k \in \mathbb{N}$, we have $\|\mathbf{c}_n^k - \mathbf{c}_n^{k-1}\| \leqslant \varepsilon$ (Deuflhard, 2011).

2.2.3. Displaying the results

For error measures and visualization of the HBM results, the amplitude of $x_n \leftrightarrow \mathbf{c}_n$ must be measured. One option of computing the amplitude of x_n is the maximum-norm $\|x_n(\mathbf{c}_n)\|_\infty$. However, in order to compute this in a robust manner one needs to compute the roots of x_n'. Determination of the derivative x_n' is trivial. The roots of the trigonometric polynomial x_n' can be interpreted as the eigenvalues of the associated Frobenius companion matrix (Edelman and Murakami, 1995). However, obtaining these eigenvalues is accompanied by typical numerical challenges of this type of problem (De Terán *et al.*, 2013). Instead, we use the readily available Euclidean norm of the HBM coefficient vector $\|\mathbf{c}_n\|_2$ to measure the system's amplitude.

2.2.4. Numerical continuation

At this point, we want to discuss how to compute an approximation of the frequency response $\mathit{\Gamma}(\mathbb{F})$. In principal, in analogy to (2.1) we could formulate the problem of computing an approximation to $\mathit{\Gamma}(\mathbb{F})$ by the set

$$\Big\{(\mathit{\Omega}, \|x_n(\mathbf{c}_n)\|_\infty) \in \mathbb{R}^2 \mid x_n(\mathbf{c}_n, t) = x_n(\mathbf{c}_n, t+T) \text{ solves (1.1)},\ T > 0,\ \mathit{\Omega} \in \mathbb{F}\Big\}$$

However, the implication

$$\mathbf{c}_n \text{ solves (2.7)} \;\Rightarrow\; x_n(\mathbf{c}_n,t) = x_n(\mathbf{c}_n,t+T) \text{ solves (1.1) for } T>0$$

and the choice of $\|\mathbf{c}_n\|_2$ over $\|x_n(\mathbf{c}_n)\|_\infty$ as an amplitude measure suggests the alternative problem: Compute an approximation of $\varGamma(\mathbb{F})$ by an *approximated frequency response* that is the set

$$\varGamma_n(\mathbb{F}) := \Big\{(\varOmega, \|\mathbf{c}_n\|_2) \in \mathbb{R}^2 \mid \mathbf{c}_n \text{ solves (2.7)},\ \varOmega \in \mathbb{F}\Big\} \tag{2.8}$$

The nonlinearity of the Duffing system allows for multiple *solution branches* of the (approximated) frequency response $B_n^i \subset \varGamma_n(\mathbb{F})$, $i = 1,2,\ldots$, where $\varGamma_n(\mathbb{F}) = \{B_n^1, B_n^2, \ldots\}$. With this, computing $\varGamma_n(\mathbb{F})$ reduces to finding each solution branch B_n^i individually. In advanced implementations of the HBM this is typically done by *numerical continuation methods* (Krack and Gross, 2019; Woiwode *et al.*, 2020). The basic idea behind these methods is simple: Fix the parameter $\varOmega$, solve $F_n(\mathbf{c}_n, \varOmega) = 0$ for $\mathbf{c}_n$ via Newton's method by starting at $\mathbf{c}_n^0$, compute the increment $\varOmega \leftarrow \varOmega + \varDelta\varOmega$ for an "optimal" choice of $\varDelta\varOmega$, perform the update $\mathbf{c}_n \leftarrow \mathbf{c}_n^0$ and repeat. Here, determining an "optimal" choice of $\varDelta\varOmega$ depends on F as well as a prescribed error tolerance ε. Additionally, the solvability of (2.7) is only given if the Jacobian of F w.r.t. $\mathbf{c}_n$ is regular. Both topics are addressed by the specific algorithmic implementation of a numerical continuation method. A popular choice for these methods is the *pseudo-arclength method* since it can follow turning points of the solution branch and it has a robust implementation in the code `AUTO` (Deuflhard, 2011). However, it relies on empirically-based control of the stepsize $\varDelta\varOmega$ which happened to fail on several of the authors' examples. A noteworthy alternative is the *asymptotic numerical method* (ANM). Woiwode *et al.* (2020) provide a thorough comparison of the pseudo-arclength method and the ANM. However, in order to avoid the drawbacks of the pseudo-arclength method we opted to use the *global quasi-Gauss-Newton method* (GQGNM) as proposed by Deuflhard *et al.* (1987). Similar to the pseudo-arclength method, the GQGNM constitutes a predictor-corrector scheme in which, first, starting at the current point, a prediction step is made, scaled by a stepsize s, in the direction of the solution branch tangent. The thereby introduced error is then corrected via a quasi-Gauss-Newton iteration s.t. a prescribed iteration tolerance $\varepsilon > 0$ is fulfilled. The GQGNM employs an error estimate-based control of the stepsize s and can deal with turning points. To different capabilities it is implemented in the codes `ALCON1` and `ALCON2` (Deuflhard *et al.*, 1987; Deuflhard, 2011) in Fortran. As to the best knowledge of the authors, a Python version of said codes is not publicly available. However, since in this work the evaluation of F is implemented in a Python routine, we implemented our own version of `ALCON1` in Python of which the source code is planned to be published as well.

2.2.5. Generating initial guesses

As it is often the case in nonlinear dynamics, the task of finding "good" initial guesses for Newton's method in order to find all system frequency responses can be challenging. Fortunately, the Duffing system (1.1) allows for a systematic generation of certain initial guesses $\mathbf{c}_n^0 \in \mathbb{R}^{2n+1}$ for arbitrary approximation orders n in order to compute certain branches of its frequency response. The required approach involves two steps:

- Let $n = 1$ and determine all solutions $x_{1,i}$, $i = 1,2,3$, analytically at $\varOmega = 0$. From this, obtain the associated HBM vectors $\mathbf{c}_{1,i}$. Then, starting at $\mathbf{c}_{1,i}$ for each $i = 1,2,3$ compute the associated branch $B_{1,i} \subset \varGamma_1(\mathbb{F})$ by the above introduced global quasi-Gauss-Newton method.
- Next, increase the ansatz order to $N = n + \varDelta n$. Then compute $\mathbf{c}_{N,i}$ at $\varOmega = 0$ by solving (2.7) via Newton's method and take $\mathbf{c}_{N,i}^0 = \big[\mathbf{c}_{n,i}^{\mathrm{T}}, 0^{\mathrm{T}}\big]^{\mathrm{T}} \in \mathbb{R}^{2N+1}$ as the initial guess for each

$i = 1, 2, 3$. Then compute the branches $B_{N,i} \subset \Gamma_N(\mathbb{F})$ accordingly. In fact, since two of the three solutions $\mathbf{c}_{n,i}$ are elements of the same branch, only two instead of three solution branches have to be computed.

- Repeat the previous step until each branch B_i is "sufficiently well" approximated by the approximated branch $B_{n,i}$ for some n.

Remark. The approach of computing the approximated frequency responses Γ_n as described above appears to be quite robust. In particular, the employed GQGNM required barely any tweaks of the user-adjustable parameters. However, the approach of generating initial values as described above does not yield *all* existing frequency responses of the Duffing system (1.1), cf, von Wagner and Lentz (2016). It only allows for a robust computation of the solution types already occurring for $n = 1$. In order to not over-complicate the problem, we did not endeavor to compute additonal solutions.

2.3. Test case and reference solution

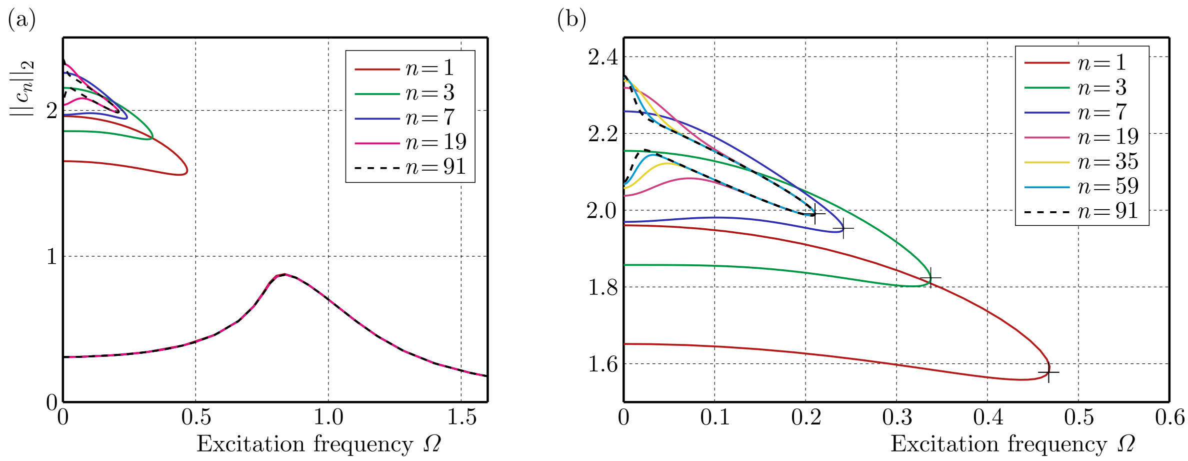

Fig. 1. Frequency response of the Duffing equation (1.1) with the test case parameters (2.9) for several approximation orders n as obtained by the solvers described in Section 2.2. The system exhibits two types of solution branches: A branch with small- respectively medium ("standard") and a branch with large-amplitude ("nose") responses. The nose solutions differ strongly in frequency range and amplitude for different ansatz orders n: (a) fequency response $\Gamma_n(\mathbb{F})$ for $\mathbb{F} = [0, 1.6]$, (b) "nose" branch of frequency response $\Gamma_n([0, 0.6])$

We consider the parameter set

$$\begin{aligned} &\delta = 2D\omega = 0.4 \qquad &&\alpha = \omega^2 = 1 \qquad &&\beta = 0 \\ &\gamma = -0.4 \qquad &&\hat{u} = 0.3 \qquad &&\Omega \in [0, 1.6] \end{aligned} \tag{2.9}$$

with $D = 0.2$ and $\omega = 1$ as the *test case* of the Duffing system (1.1) for this study. As already mentioned in the introduction, all parameters are considered to be dimensionless which also holds for the displacement x and time t. Solver-wise we used an iteration error tolerance of $\varepsilon = 10^{-14}$ throughout this study. Corresponding results are shown in Fig. 1 for different ansatz orders n. The system exhibits the — for the softening case — well known types of solution branches: One small-, respectively medium- and two large-amplitude responses which we refer to as "standard" and "nose" branches or responses. For the resonance peak of the standard response at $\Omega \approx \omega = 1$, there are for the parameter set (2.9) no multiple solutions due to moderate damping. The standard response covers the entire considered frequency range $\mathbb{F} = [0, 1.6]$ with amplitudes in the range of $\|\mathbf{c}_n\|_2 \in [0.19, 0.87]$. Our special focus in the following is on the nose responses, however. These nose responses cover only the frequency range from zero to the characteristic turning point marked by +, which differs largely for different ansatz orders n. This behavior has been denoted as "artifacts" by the second and third author of the present paper and investigated in several papers, e.g. von Wagner and Lentz (2016, 2018, 2019). In these prior publications certain solutions are considered as artifacts that exceed a maximum amplitude threshold w.r.t. the neglected higher order terms of the HBM method, i.e., in the Duffing equation, the terms with harmonics of order $n+1$ to $3n$.

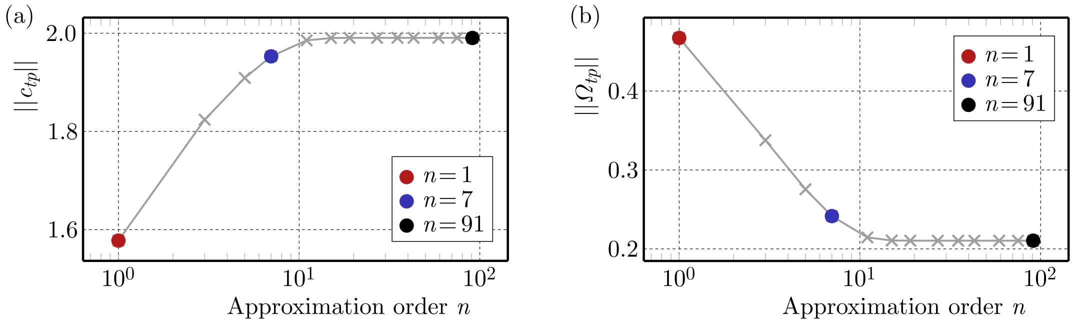

Fig. 2. Convergence of the amplitude $\|c_{tp}\|$ and excitation frequency Ω_{tp} of the turning point

In contrast, in the present paper we consider new suitable aspects for defining and identifying artifact solutions as presented in Sections 3 and 4. Thereby, several numerical and geometrical measures, e.g. the position of the turning point of the nose response and its convergence w.r.t. the ansatz order n of the associated HBM solution x, are investigated. As will be seen, not all of the examined measures are useful with respect to the task of determining artifacts. As the first attempt, we consider in Fig. 2a and 2b for our test case the convergence of the nose solutions turning point denoted as $(\Omega_{tp}, \|c_{tp}\|_2)$. It can be observed that both the solution amplitude $\|c_{tp,n}\|_2$ and the excitation frequency $\Omega_{tp,n}$ appear to converge for increasing ansatz orders n. Nevertheless, as can be observed in Fig. 1, convergence of the turning point does not necessarily imply convergence of all other solution points of the nose branch, where in general larger ansatz orders are necessary. The iteration error of the ansatz orders $n = 75, 91$ yields $|\|c_{tp,91}\|_2 - \|c_{tp,75}\|_2| \approx 6.67 \cdot 10^{-7}$ and $|\Omega_{tp,91} - \Omega_{tp,75}| \approx 2.48 \cdot 10^{-4}$ which we deem to be small. Hence we assume that the HBM solution x_{91} converged sufficiently close to the exact solution x of (1.1) at least at the turning point. Although for $n = 19$ the iteration error at the turning point is similiar to the one for $n = 91$, sufficient convergence is not yet achieved in a large part of the frequency range of the nose branch. This is why we consider x_{91} over a potential lower order solution as the *reference solution* with the associated ansatz order $n = 91$ of the test case (2.9).

Next, in Fig. 3, the frequency response of the ansatz order $n = 1$ is compared to the reference response of the ansatz order $n = 91$ in more detail. The frequency values of the turning points of the nose branches for $n = 91$ and $n = 1$ are found to be approximately $\Omega_{tp,91} = 0.21$ and $\Omega_{tp,1} = 0.47$, respectively. The two turning points at $\Omega_{tp,91}$ and $\Omega_{tp,1}$ can be considered to divide the entire frequency range into the sub-intervals $\mathbb{F}_A := [0, \Omega_{tp,91}]$, $\mathbb{F}_B := [\Omega_{tp,91}, \Omega_{tp,1}]$ and $\mathbb{F}_C := [\Omega_{tp,1}, 1.6]$ s.t. $\mathbb{F} = \mathbb{F}_A \cup \mathbb{F}_B \cup \mathbb{F}_C$. Now note that the reference nose response only covers $\mathbb{F}_A$ but the lower order nose response of $n = 1$ covers $\mathbb{F}_A$ and also $\mathbb{F}_B$. Apparently, the solutions of the response of $n = 1$ for all frequencies in $\mathbb{F}_B$ "vanish" upon an increase of the ansatz order to $n = 91$. From this, we conclude that the solution of the frequency response of $n = 1$ in the frequency range $\mathbb{F}_B$ are *artifact solutions* as defined in more detail in Section 3.

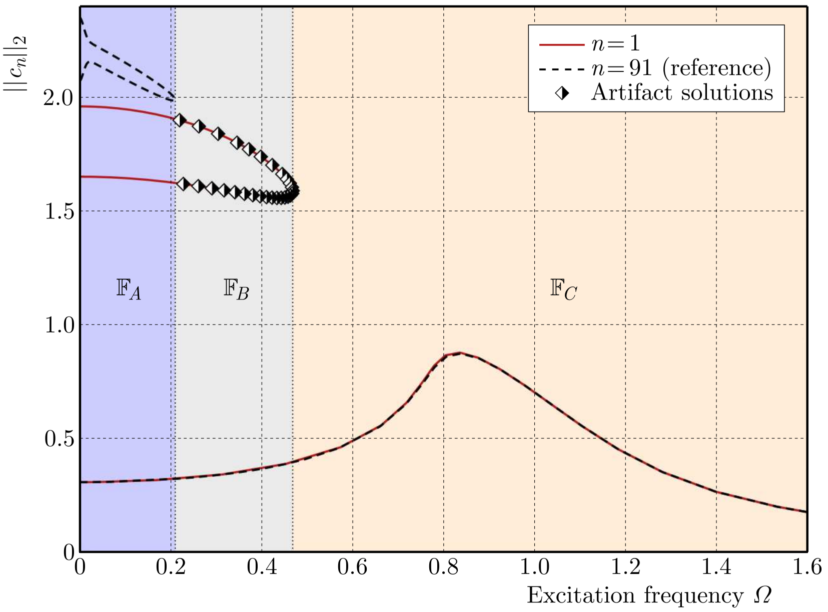

Fig. 3. Approximated frequency response $\Gamma_n(\Omega)$ for ansatz orders $n = 1, 91$ of the case (2.9). Solutions in the range $\mathbb{F}_B$ are denoted as artifact solutions, cf. Definition 1 in Section 3

3. Error measures

As can be seen from the results in Fig. 1 and 3, HBM solutions of low ansatz orders exhibit, especially for the nose solutions, a deviation from the reference solution ($n = 91$). This deviation is considered to be an error due to HBM and is divided into two types of errors. The first refers to the amplitude error which we refer to as *quantitative error* which can be measured, e.g., by the convergence error $\|\mathbf{c}_n - \mathbf{c}_{n_{ref}}\|$. The second error type is the artifact behavior which we refer to as *qualitative error*. In order to be able to measure the qualitative error, a mathematical definition of the artifact behavior is required. The definition is presented and discussed in the following. After that we discuss potential qualitative error measures based on the residual as well as algebraic, geometric and solver-related properties.

3.1. Artifact definition

In the following, we provide and discuss a mathematical definition of the artifact behavior which we base on the turning points of the solution branches. We start by providing the required mathematical lingua. Let $n, n_{ref} \in \mathbb{N}$ be two HBM ansatz orders with $n < n_{ref}$ as well as $B_n \subset \Gamma_n(\mathbb{F})$ and $B_{n_{ref}} \subset \Gamma_n(\mathbb{F})$ two computed solution branches, respectively. Again, we refer to the solutions of the ansatz order n_{ref} as reference (solutions). Recall that $B_n = \{P_1, \dots, P_N\}$ and $B_{n_{ref}} = \{Q_1, \dots, P_M\}$ are *ordered* point sets with points $P_i = (\Omega_n^i, (\|\mathbf{c}_n\|_2)^i) \in \mathbb{R}^2$, $i = 1, \dots, N$, and $Q_j = (\Omega_{n_{ref}}^j, (\|\mathbf{c}_{n_{ref}}\|_2)^j) \in \mathbb{R}^2$, $j = 1, \dots, M$, where $N, M \in \mathbb{N}$ are the number of points of the solution branches B_n and $B_{n_{ref}}$, respectively. We assume that B_n and $B_{n_{ref}}$ each exhibit a turning point denoted as $P_{tp} \in B_n$ and $Q_{tp} \in B_{n_{ref}}$ and we denote the associated excitation frequency as $\Omega_{n,tp}$ and $\Omega_{n_{ref},tp}$, respectively. In order to compare the similarity of the turning points P_{tp} and Q_{tp}, we consider their local curvature w.r.t. the frequency component. For this, let X be a turning point on a curve $B \in \mathbb{R}^2$ and let $\mathcal{B}_s(X) \subset B$ denote an arclength-parameterized neighborhood[1] of B at X with arclength $s > 0$. With this, we define the *signed normalized curvature w.r.t.* Ω *at* X as

[1] That is, all points Y that lie on B and that are closer to X than the arclength $s > 0$.

$$\kappa_\Omega(X) := \begin{cases} +1 & \forall Y \in \mathcal{B}_s(X): \quad (Y)_\Omega > (X)_\Omega \\ -1 & \forall Y \in \mathcal{B}_s(X): \quad (Y)_\Omega < (X)_\Omega \end{cases} \tag{3.1}$$

where $(X)_\Omega, (Y)_\Omega$ denote the Ω-component of $X, Y \in B$, respectively. In (3.1), $\kappa_\Omega(X) = \pm 1$ basically means that at the turning point X the curve B "opens" to the right (resp. left). With this, we can present the following

Definition 1 (Artifact solution). *Let two solution branches B_n and $B_{n_{ref}}$ be given. Assume they each exhibit a single turning point $P_{tp} \in B_n$ and $Q_{tp} \in B_{n_{ref}}$. If $\kappa(P_{tp}) = \kappa(Q_{tp}) = \pm 1$ and $\Omega_{n_{ref},tp} \gtrless \Omega_{n,tp}$ then all points $P \in B_n$ with frequency components Ω in the frequency range $[\Omega_{n,tp}, \Omega_{n_{ref},tp}]$ (respectively $[\Omega_{n_{ref},tp}, \Omega_{n,tp}]$) are called* artifact solutions.

Two possible situations for artifact solutions to occur as given in Definition 1 are depicted in Fig. 4a and 4b.

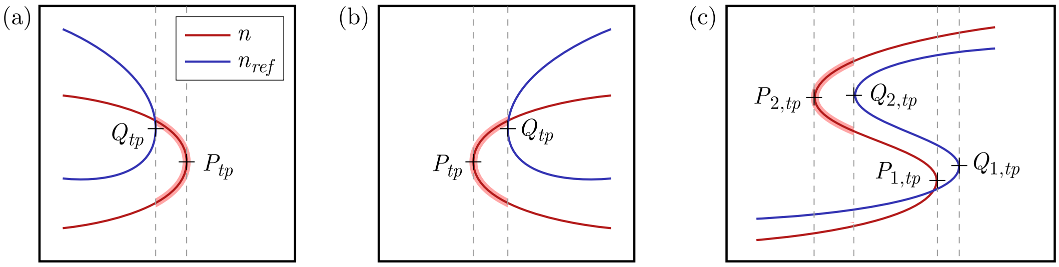

Fig. 4. Three representative cases of artifact solutions (▬) on solution branches as identified by applying Definition 1: (a) and (b) single turning point per solution branch B_n and $B_{n_{ref}}$; (c) multiple turning points per solution branch B_n and $B_{n_{ref}}$, however artifact solutions are only existent between the turning points $P_{2,tp}$ and $Q_{2,tp}$

In case both solution branches B_n and $B_{n_{ref}}$ exhibit multiple turning points then the above definition may be applied in succession according to the following scheme:

1. Assume that B_n and $B_{n_{ref}}$ exhibit the same number $H \in \mathbb{N}$ of turning points denoted as $P_{tp,1}, \ldots, P_{tp,H} \in B_n$ and $Q_{tp,1}, \ldots, Q_{tp,H} \in B_{n_{ref}}$, respectively. Further assume that the aforementioned turning points ordering coincides with the ordering of the points of B_n and $B_{n_{ref}}$, i.e. $B_n = \{\ldots, P_{tp,1}, \ldots, P_{tp,H}, \ldots\}$ and $B_{n_{ref}} = \{\ldots, Q_{tp,1}, \ldots, Q_{tp,H}, \ldots\}$, respectively.
2. If now $\kappa(P_{tp,i}) = \kappa(Q_{tp,i})$ for all $i = 1, \ldots, H$, then Definition 1 can be applied to each pair of the turning point $(P_{tp,i}, Q_{tp,i})$ successively in order to identify artifact solutions.

Figure 4c shows an exemplary case of two turning points per solution branch where the above scheme can be applied. Although there are two turning points per branch, there is only a single frequency region in which solution artifacts occur. For an algorithmic detection of artifact solutions based on Definition 1, a robust detection of turning points is critical. This can be done within the employed GQGNM by means of a readily available cubic Hermite interpolation (Deuflhard *et al.*, 1987). An alternative approach of robust computation of the turning points would be available upon implementation of the ANM (Woiwode *et al.*, 2020).

The objective in the following is to consider a number of error measures in general and at its best to find a way to distinguish between artifacts and other types of errors, and to avoid both of them. To this end, various methods of measuring the error of an HBM solution will be introduced and applied to the test case described in Subsection 2.3. Of course, other classifications of errors are possible, e.g. errors due to the HBM itself, comparing the HBM solution with the exact solution and numerical errors while applying the HBM. These errors have been studied in detail in e.g. Urabe (1965), Kogelbauer and Breunung (2021), García--Saldaña and Gasull (2013), Woiwode and Krack (2023). These studies do not investigate the qualitative error type, i.e. the artifact behavior described above is not considered. However we want to point out that in the work of Woiwode and Krack (2023), the suggested n-adaptive error measure appears to us to be a potential tool of detecting artifacts. Within their approach of a numerical continuation method, the HBM ansatz order is adaptively refined or coarsened based on an error measure. The adaptive switching of the ansatz order could possibly speed up the detection of turning points which is a mandatory step to successfully apply the solution artifact Definition 1. To our understanding, this approach could, in principle, avoid the computation of possible artifact solutions, although a confirmation of this hypothesis would require further research.

3.2. Residual

An obvious way to measure the error of a HBM solution is to consider the terms neglected in equation (2.6), which are evaluated in the following. As can be inferred from the definition of this expression, the value of this residual must be zero if the HBM solution exactly satisfies the underlying differential equation. The residual thus represents a measure of the non-fulfillment of the differential equation, but does not provide direct information about the extent to which the HBM solution deviates from the exact solution. Nevertheless, it has been shown in previous works, e.g. Ferri and Leamy (2009), von Wagner and Lentz (2018, 2019), Lentz and von Wagner (2020), that the residual can be used to determine whether the approximation order of an HBM solution needs to be further increased to achieve the HBM solution that accurately represents the exact solution. A drawback of this procedure, as shown e.g. in von Wagner and Lentz (2018, 2019), Lentz and von Wagner (2020) is that the residual is not suitable for distinguishing between the quantitative and qualitative error type. A high value merely indicates that the examined HBM solution has a high error. Therefore, the solution might be an artifact, or it could be a solution that exists but provides a poor approximation of the exact amplitude due to an insufficient order of approximation. Therefore, as the residual was considered in several earlier publications of the authors, it is not further considered in the subsequent analysis in the present paper.

3.3. Algebraic measures

Another approach to investigate the error associated with a solution involves a direct examination of the underlying algebraic system of equations. The approach consists of searching within the algebraic system equations F_n for indications that finding a solution will be problematic. If such indications are present, it is reasonable to assume that the solutions found are flawed. Therefore, various methods will be enumerated in the following, which can be used to estimate the quality of a solution based on the solved system of equations. Since the properties of the algebraic equation system are largely determined by the linear component, the Jacobian matrix — in short *Jacobian* — at the solution point is used for this purpose. Since the equation system can be considered as a function of the coefficients or the excitation frequency, the following three Jacobian can be defined

$$J_{\mathbf{c},n}(\mathbf{c},\Omega) := D_{\mathbf{c}}F_n(\mathbf{c},\Omega) \qquad J_{\Omega,n}(\mathbf{c},\Omega) := D_{\Omega}F_n(\mathbf{c},\Omega)$$
$$J_n(\mathbf{c},\Omega) := \big[J_{\mathbf{c},n}(\mathbf{c},\Omega), J_{\Omega,n}(\mathbf{c},\Omega)\big]$$

Here, $J_{\mathbf{c},n}(\mathbf{c},\Omega)$ denotes the Jacobian of F_n w.r.t. $\mathbf{c}$, $J_{\Omega,n}(\mathbf{c},\Omega)$ denotes the Jacobian of F_n w.r.t. Ω and $J_n(\mathbf{c},\Omega)$ denotes the full Jacobian of F_n.

3.3.1. Condition number

Before starting to assess the qualitative error of solutions, we investigate the ill- or well--posedness of the problem of solving the system $F_n(\mathbf{c}_n, \Omega) = 0$ in the scope of Newton's method as required in the solvers described in Section 2.2. For $A = J_{\mathbf{c},n}(\mathbf{c}, \Omega)$ or $A = J_n(\mathbf{c}, \Omega)$, this is measured by the *condition number of A*

$$\mathrm{cond}_2(A) := \|A\|_2 \|A^{-1}\|_2 \tag{3.2}$$

where $\|\cdot\|_2$ is the matrix norm induced by the Euclidean vector norm. An upper bound for the relative error amplification made during solving of the linear equation system within Newton's method is given by the factor $\mathrm{cond}_2(A) \cdot \delta$, where δ is the relative error in $\mathbf{c}_n$ (Deuflhard and Hohmann, 2019). Since in Newton's method linear systems are solved iteratively, the cummulative relative error is proportional to $\mathrm{cond}_2(A) \cdot \delta \cdot M$, where M is the number of iterations. In our case, $\delta = \varepsilon = 10^{-14}$, and typically $M = 12, \dots, 20$. Consequently, this problem is said to be *well-* or *ill-conditioned*, if $\mathrm{cond}_2(A)$ is small or large, respectively. Furthermore, A being singular is equivalent to $\mathrm{cond}_2(A) = \infty$. This is the case for the turning point of the nose solution branch at which $J_{\mathbf{c},n}(\mathbf{c}, \Omega_{tp})$ is singular. Hence, in numerical practice, large condition numbers can be used as an indicator for singularities of the associated matrix. To illustrate this concept, Fig. 5 depicts values of the condition number for the range $(c_1, c_2) \in [-2.4, 2.4]^2$ and for excitation frequencies $\Omega = 0.0,\ 0.2,\ 0.4,\ 0.6$. Additionally, the standard branch solution (◓) as well as the two nose branch solutions (larger amplitude: ⬗, smaller amplitude: ⬖) existing at each frequency are marked. Based on these graphs, it is possible to assess the values that the condition number of the Jacobian matrix takes in the vicinity of the solutions. As can be seen, solutions with a large amplitude (i.e. ⬗, ⬖) are located in a region with high values, while the solutions with a low amplitude (i.e. ⬗, ⬖) is in a region with low values.

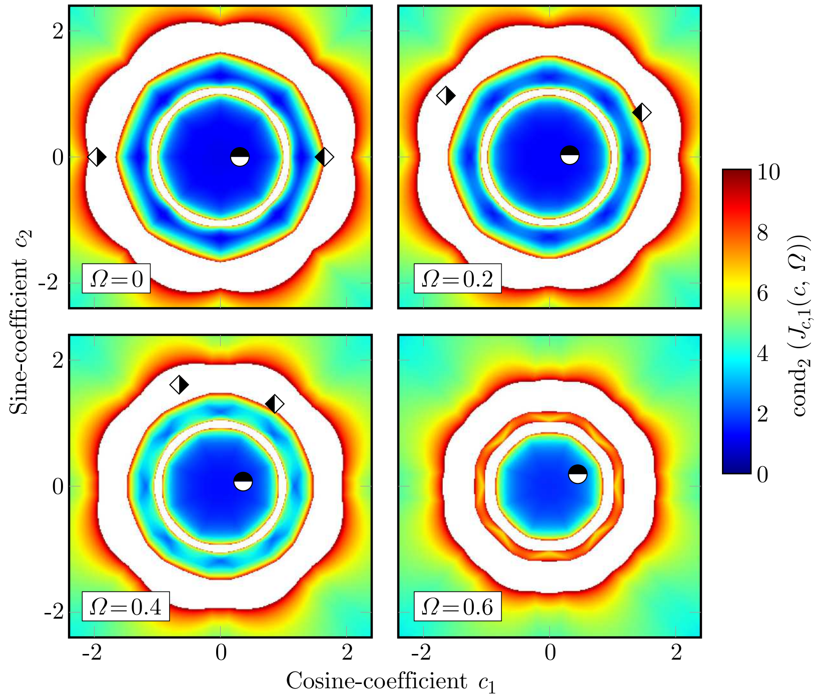

Fig. 5. Condition number $\mathrm{cond}_2\big(J_{\mathbf{c},1}(\mathbf{c}, \Omega)\big)$ for $n = 1$ for four different excitation frequencies of test case (2.9). Values greater than ten are color-coded to white. The symbols ⬗, ⬖ and ◓ denote the two nose solutions and the standard solution, respectively

3.3.2. Jacobian angle

Another possibility is to consider the angles between the columns of the Jacobian. The motivation for this is that these angles can be used as a measure of the linear independence of the linearized equations. For example, an angle of 90° means that the equations are linearly independent, while an angle of 0° means that the equations are linearly dependent. With regard to the solvability of the equation system, it is therefore assumed that small angles may indicate difficulties in computing the solution. For a more precise definition, let $(\mathbf{J}_{\mathbf{c},n}(\mathbf{c},\Omega))_i \in \mathbb{R}^{1,2n+1}$ denote the i-th row vector of $J_{\mathbf{c},n}(\mathbf{c},\Omega)$ for all $i = 0, 1, \dots, 2n$. In particular, for $n = 1$ and $c_0 := 0$ let $\mathbf{J}_1, \mathbf{J}_2 \in \mathbb{R}^{1,2}$ denote the first and second row vector of the Jacobian matrix $J_{\mathbf{c},1}(\mathbf{c},\Omega)$. Then

$$\theta := \arccos \frac{\mathbf{J}_1 \mathbf{J}_2^{\mathrm{T}}}{\|\mathbf{J}_1\| \|\mathbf{J}_2\|} \in \left[0, \frac{\pi}{2}\right] \tag{3.3}$$

is referred to as the Jacobian angle. To illustrate this concept as well, the same method as described above is employed. The corresponding graphs can be found in Fig. 6. These graphs also demonstrate that nose solutions with large amplitude (⬖, ⬗) are located in regions with poor solution properties, characterized by small Jacobian angles, while the standard solution with a small amplitude (◓) is situated in a region with good solution properties, characterized by a large Jacobian angle. In order to extend the concept of the Jacobian angle to ansatz orders $n > 1$, we consider to find the minimal Jacobian angle over all pair-wise distinct row vectors $\mathbf{J}_i \in \mathbb{R}^{1,2n+1}$ which we define as

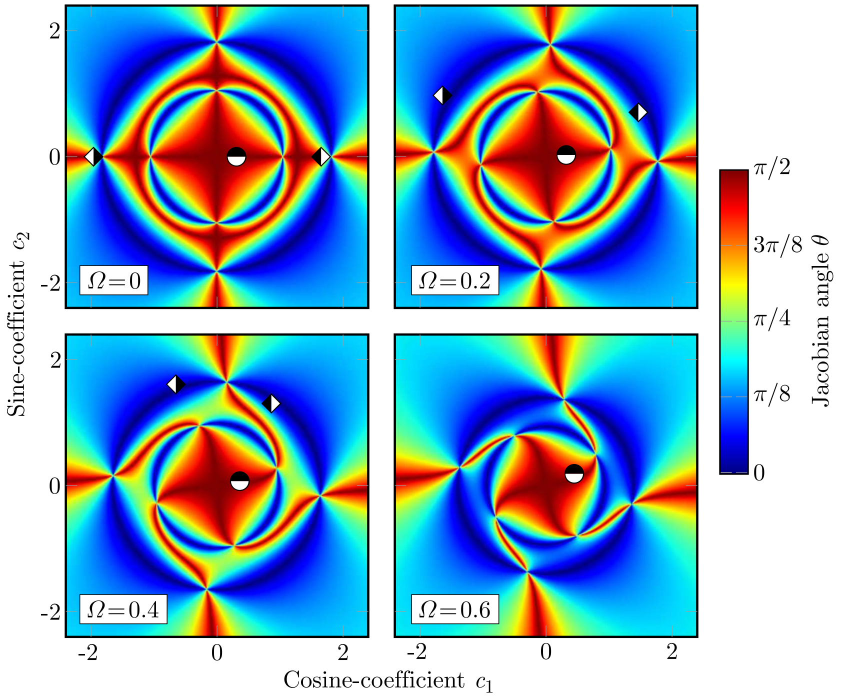

Fig. 6. Jacobian angle θ for $n = 1$ for four different excitation frequencies of the test case (2.9), symbols as in Fig. 5

$$\theta_{min} := \min_{i \neq j = 0,1,\dots,2n} \arccos \frac{\mathbf{J}_i \mathbf{J}_j^{\mathrm{T}}}{\|\mathbf{J}_i\| \|\mathbf{J}_j\|} \in \left[0, \frac{\pi}{2}\right] \tag{3.4}$$

and refer to it as the *minimal Jacobian angle*. Here, we consider the minimum as the measure of choice since of all row vectors of the two associated to the minimal Jacobian angle are the pair closest to be linearly dependent. And, of course, for $n = 1$, the two definitions (3.3) and (3.4) are equivalent.

3.3.3. Number of solutions

As the final algebraic measure, we consider the number of real solutions of x of (1.1) for a given excitation frequency Ω. In the context of the HBM, this requires to investigate *the number of real solutions of* F_n for a given ansatz order n and excitation frequency Ω, which we denote as $\#F_n$. In general, there exists no *a priori* way of determining $\#F_n$ expect for computing *all* real solutions $\mathbf{c}_{n,i}$, $i = 1, \dots, \#F_n$. However, this is impractical since Bézout's theorem provides $\#F_n \leqslant \prod_{i=0}^{2n} \deg(R_i)$ as the upper bound for the number of solutions of F_n where $\deg(R_i)$ is the degree of the multivariate polynomial R_i which is the i-th equation of F_n (Basu *et al.*, 2006). Fortunately, in this work we only consider the standard and nose solution branch as a subset of the entire frequency response of the Duffing system. This reduces the complexity of the problem of determining $\#F_n$ drastically to simply counting the number of solutions of $\Gamma_n(\{\Omega\})$ for each $\Omega \in \mathbb{F}$, i.e. $\#F_n(\Omega) = |\Gamma_n(\{\Omega\})|$, where $|\cdot|$ denotes the cardinality of $\Gamma_n(\{\Omega\})$. Considering only the standard and nose branch, this yields $\max_{\Omega\in\mathbb{F}} |\Gamma_n(\{\Omega\})| = 3$. With this, the number of solutions can be compared for different ansatz orders and different frequencies. In fact, this procedure can be applied to any combination of two ansatz orders $n_1 < n_2$ for which the difference $|\Gamma_{n_2}(\{\Omega\})| - |\Gamma_{n_1}(\{\Omega\})|$ needs to be determined for every $\Omega \in \mathbb{F}$. Intuitively, the main disadvantage of this approach is that the two frequency responses Γ_{n_1} and Γ_{n_2} need to be computed beforehand, i.e. it is not an *a priori* measure that identifies artifacts for a requested ansatz order — it needs the frequency response of the second, higher ansatz order as a reference.

3.4. Geometric measures

Next, we want to investigate two geometric measures. In order to be able to compare the "resemblance" of two solution branches of the frequency response, an adequate measure is required. We already discussed the convergence of the nose branch turning point in Section 2.3 and its disadvantage of not being able to capture the entirety of the solution branches. Instead, we consider two normalized distance measures that measure the distance between the solution branch of the ansatz order n and the corresponding reference solution branch of the ansatz order n_{ref}.

3.4.1. Arclength distance

First, straightforwardly, consider the *arclength* or *length* $L(B) > 0$ of the solution branch $B \in \Gamma(\mathbb{F})$. The approximated solution branch B_n can be interpreted as a polygonal curve — or polyline — represented by N points of the ordered set $\{(\Omega, \|\mathbf{c}_n\|_2)_i\}_{i=1}^{N} \subset \mathbb{R}^2$. A simple approximation $L(B) \approx L(B_n)$ for the approximated solution branch $B_n \subset \Gamma_n(\mathbb{F})$ is obtained by

$$L(B_n) := \sum_{i=0}^{N-1} \|d_{i+1} - d_i\| \qquad d_i := \begin{bmatrix} \Omega \\ \|\mathbf{c}_n\|_2 \end{bmatrix}_i \in \mathbb{R}^2 \tag{3.5}$$

i.e. the line segments of the polygonal curve B_n. With this, we introduce the *arclength distance* between the solution branch of the ansatz order n and n_{ref} as

$$d_L(B_n, B_{n_{ref}}) := |L(B_n) - L(B_{n_{ref}})| \tag{3.6}$$

Furthermore, in order to be able to compare multiple arclength distances, we introduce the *normalized arclength distance*

$$\overline{d}_L(B_n, B_{n_{ref}}) := \frac{d_L(B_n, B_{n_{ref}})}{L(B_{n_{ref}})} \tag{3.7}$$

where $L(B_{n_{ref}})$ is the arclength of the reference solution branch. The arclength is computationally inexpensive and, therefore, $d_L, \overline{d}_L$ readily available. However, both variants are not invariant under translation and rotation of the polygonal curves $B_n, B_{n_{ref}}$.

3.4.2. *Fréchet distance*

An improved but computationally more expensive distance measure is the *Hausdorff distance*. However, it does not consider the course of two compared curves. Fortunately, the so-called Fréchet distance circumvents the disadvantages of both distance measures at the expense of higher computational costs. Let a, b be parametrizations of two polylines A, B, respectively. Then the *Fréchet distance* between A and B is defined as

$$d_F(A, B) := \inf_{a,b} \max_{t\in[0,1]} \{\|A(a(t)) - B(b(t))\|_2\} \tag{3.8}$$

This distance measure captures the similarity between A and B while it takes into account the ordering and position of the curves points. An intuitive understanding of (3.8) might be obtained by the following analogy (Alt and Godau, 1995): "A person is walking a dog on a leash: the person can move on one curve, the dog on the other; both may vary their speed, but backtracking is not allowed. Then the Fréchet distance of the two curves is the minimal required length of the leash". In practice, we are actually interested in computing an approximation of the Fréchet distance for two approximated solution branches B_{n_1}, B_{n_2}. A "good" approximation of $d_F(B_{n_1}, B_{n_2})$ is given by the so-called *discrete Fréchet distance* (DFD) (Alt and Godau, 1995) which we denote by $d_{DF}(B_{n_1}, B_{n_2})$. In order to compute the DFD, we utilize the Python code `discrete-frechet` by Figueira (2023). Finally, we introduce the *normalized DFD*

$$\overline{d}_{DF}(B_n, B_{n_{ref}}) := \frac{d_{DF}(B_n, B_{n_{ref}})}{L(B_{n_{ref}})} \tag{3.9}$$

in order to be able to compare it to the normalized arclength distance.

3.5. Solver measures

In addition to the aforementioned residual, algebraic and geometric measures, we also investigate measures obtainable from the employed solvers in order to test whether information available by the solvers can indicate artifact behavior or not. For this, we consider the *number of correction steps k per prediction step* as well as the *computation time per prediction step.* Since both measures correlate strongly, we only present the number of correction steps k as a representative quantity of the solver behavior. Additionally, in order to benchmark the performance of the employed numerical continuation method we also provide data on the following solver-related quantities:

- The *computation time* in seconds t_{comp} required to obtain each solution branch per given ansatz order n.
- The *number of prediction steps k_{pred}* of each solution branch per given ansatz order n, i.e. the number of points that constitute each solution branch.
- The *total number of correction steps k_{corr}* in order to compute each solution branch per given ansatz order n, i.e. the sum of the number of correction steps over all prediction steps.
- The *average number of required correction steps $\overline{k}_{corr} = k_{corr}/k_{pred}$* per prediction step.

4. Error measures applied to test case

In this Section, the error measures described in Section 3 are applied to the test case given in Section 2.3. This is intended to assess the extent to which these error measures are suitable for identifying artifacts and errors in general. As clarified in Section 2.3, regarding the test case, it is known that all nose solutions, i.e. those with large amplitudes, for an excitation frequency

$\Omega > 0.21$ are artifacts. Hence, an error measure that is suitable for distinguishing artifacts from regular solutions is expected to yield significantly different values for solutions with large amplitudes for excitation frequencies $\Omega > 0.21$ ($\mathbb{F}_B$ in Fig. 3) compared to excitation frequencies $\Omega \leqslant 0.21$ ($\mathbb{F}_A$ Fig. 3). Consequently, we expect a discontinuity at $\Omega = 0.21$ of such an error measure. Recall that in the present work, the focus is on ansatz functions x_n with a vanishing mean value, i.e. $c_0 = 0$, and thus $\mathbf{c}_n \in \mathbb{R}^{2n}$.

4.1. Algebraic measures

4.1.1. Condition number

The first algebraic measure results we present are the condition numbers $\mathrm{cond}_2(J_{\mathbf{c},n}(\mathbf{c}_n, \Omega))$ and $\mathrm{cond}_2(J_n(\mathbf{c}_n, \Omega))$ given in Fig. 7a and 7b, respectively. Both figures plot the respective condition number over the excitation frequency for the standard and nose response for ansatz orders $n = 1, 7, 91$. We first discuss Fig. 7a. First of all, the expected increase of the condition number for an increase of the ansatz order can be observed. For $n = 1$ the nose response exhibits a condition number of one to three orders of magnitude larger than the condition number of the standard response. In particular, towards the turning point of the nose the condition number rapidly increases towards values of $\mathrm{cond}_2(J_{\mathbf{c},1}(\mathbf{c}_1, \Omega_{tp,1})) \approx 10^3$. Similar behavior can be observed for the ansatz order $n = 7, 91$ with a maximum condition number towards the turning point at around $10^5, 10^7$, respectively. For the largest condition number associated with $n = 91$, we have $\mathrm{cond}_2(J_{\mathbf{c},91}(\mathbf{c}_{91}, \Omega_{tp,91})) \cdot \varepsilon \approx 10^7 \cdot 10^{-14} = 10^{-7} \ll 1$ for a single Newton step. Consequently, with a typical value of around $N = 15$ Newton steps until convergence we may extrapolate to $\mathrm{cond}_2(J_{\mathbf{c},91}(\mathbf{c}_{91}, \Omega_{tp,91})) \cdot \varepsilon \cdot N = 5 \cdot 10^{-6} \ll 1$. From this we conclude that, from the numerical practical standpoint, the problem of solving the linear system within Newton's method is still considered to be well-conditioned. However, note that in case the numerical continuation method reaches an excitation frequency that is numerically close to the turning point of the system, the condition number becomes unbounded and the problem ill-conditioned.

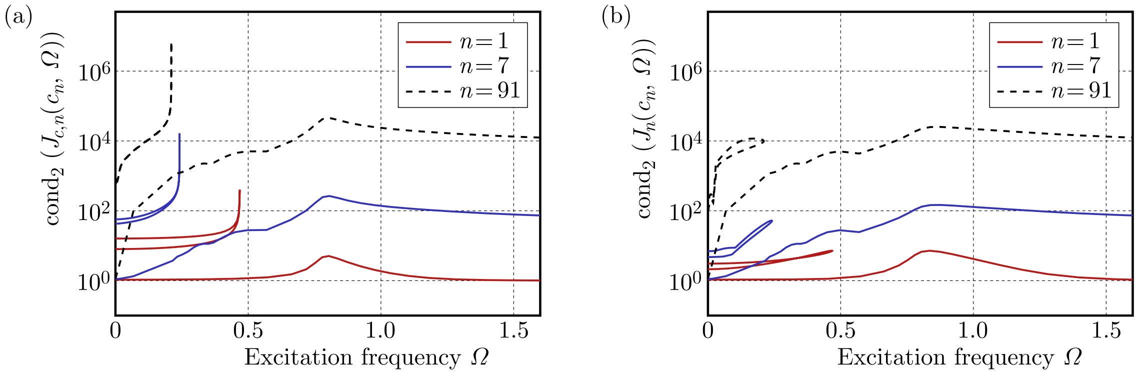

Fig. 7. Condition numbers $\mathrm{cond}_2(J_{\mathbf{c},n}(\mathbf{c}_n, \Omega))$ and $\mathrm{cond}_2(J_n(\mathbf{c}_n, \Omega))$ of the system F_n for approximation orders $n = 1, 3, 91$ for the test case (2.9)

Next, we discuss Fig. 7b. It shows a similar qualitative behavior for the condition number of the extended Jacobian matrix $\mathrm{cond}_2(J_n(\mathbf{c}_n, \Omega))$. However, the largest condition number values at the turning points of the nose branch of ansatz orders $n = 1, 7, 91$ are approximately 7.21, 52.4 and $1.16 \cdot 10^4$, respectively. A comparison of the condition number of $J_{\mathbf{c},n}$ to J_n yields that the values of the condition number of the extended Jacobian are three orders of magnitude smaller. This can be explained by the additional information due to the existence of the additional matrix column $J_{\Omega,n}$, i.e. additionally considering the derivative w.r.t. the excitation frequency improves the conditioning of the original problem. This is in fact used by the pseudo-arclength method or the GQGNM, as presented in Section 2.2. Interestingly, near the standard branch resonance peak at $\Omega = 0.8$, the condition number of the extended Jacobian $\mathrm{cond}_2(J_n(\mathbf{c}_n, \Omega))$ for $n = 7, 91$ is three (respectively two) times larger. However, upon returning to the question of artifact solutions, for $n = 1, 7$ in the frequency range around the turning point $\Omega_{tp,91} = 0.21$ no noticeable change in the condition numbers $\mathrm{cond}_2(J_{\mathbf{c},n}(\mathbf{c}_n, \Omega))$ and $\mathrm{cond}_2(J_n(\mathbf{c}_n, \Omega))$ can be observed. This is why we do not consider either of these two condition numbers to be indicative of artifact behavior.

4.1.2. Jacobian angle

The second algebraic measure result we discuss is the minimal (extended) Jacobian angle $\theta_{min}(J_{\mathbf{c},n})$ (resp. $\theta_{min}(J_n)$) given in Fig. 8a and 8b as an extension of the Jacobian angle for an ansatz order $n \geqslant 1$.

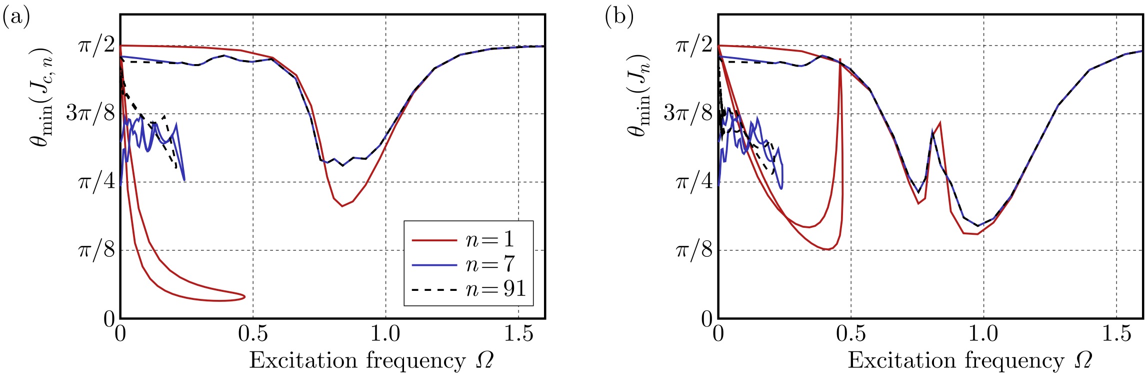

Fig. 8. Minimal (extended) Jacobian angle $\theta_{min}(J_{\mathbf{c},n})$ (resp. $\theta_{min}(J_n)$) per excitation frequency for ansatz orders $n = 1, 7, 91$ for the test case (2.9)

The two figures plot the angles over the excitation frequency for the standard and nose response for ansatz orders $n = 1, 7, 91$. We start by discussing Fig. 8a. The curves are mostly close to $\pi/2$ and have minimal values around $\pi/4$ close to the small amplitudes resonance peak. The minimum angles over all frequencies can be found close to the respective standard branch resonance peak with values of 41%, 56% and 55% of $\pi/2$ for $n = 1, 7, 91$, respectively. As could be expected, the minimal Jacobian angle for standard branches for $n = 7$ appears to be mostly converged against the reference solution branch of $n_{ref} = 91$. On the other hand, the nose branch minimal angles are 6.3%, 50% and 55% for $n = 1, 7, 91$, respectively, and are observed to be close to the respective turning points. Note, that there is a noticeable difference in the angles of roughly 30° when comparing the nose branch of ansatz orders $n = 1$ to $n = 7, 91$. Next, we turn our focus to Fig. 8b. The curves look mostly similar, expect for two noticeable outliers. First, for all three ansatz orders the standard solution branches exhibit a similar but somewhat remarkable increase of the angle close to the resonance peak of the standard frequency response curves. However, in contrast, when comparing the respective nose solution branches only for $n = 1$, a noticeable increase at the turning point can be observed. Similar to the case of the condition number of the extended Jacobian, we attribute these two frequency-wise "local" increases of the Jacobian row vector angles to including the additional column vector of $J_{\Omega,n}$ in the extended Jacobian. However, it would require further investigation in order to answer why this is only observed in such a local manner for the standard solution branch. Upon returning to the original question of artifact detection capabilities, we conclude that neither of the two discussed measures is suitable for detecting artifacts since for $n = 1$ or $n = 7$ there is no observable change of the minimal (extended) Jacobian angle at $\Omega = 0.21$, i.e. the frequency value of the turning point of the reference solution branch.

4.1.3. Number of solutions

Next, we investigate the number of solutions, as defined in Section 3.3, as a potential artifact measure. For this, consider the frequency range partition $\mathbb{F} = \mathbb{F}_A \cup \mathbb{F}_B \cup \mathbb{F}_C$, as introduced in Section 2.3, which we obtained by identifying the turning points of the frequency response curves of the ansatz order $n = 1$ and the reference ansatz order $n_{ref} = 91$. Upon counting the number of solutions over each frequency range, $\mathbb{F}_A, \mathbb{F}_B, \mathbb{F}_C$ yields for $n = 1$

$$|\varGamma_1(\mathbb{F}_A \cup \mathbb{F}_B)| = 3 \qquad \text{and} \qquad |\varGamma_1(\mathbb{F}_C)| = 1$$

and for $n_{ref} = 91$

$$|\varGamma_{91}(\mathbb{F}_A)| = 3 \qquad \text{and} \qquad |\varGamma_{91}(\mathbb{F}_B \cup \mathbb{F}_C)| = 1$$

Comparing the number of solutions of both ansatz orders, yields the differences

$$|\varGamma_{91}(\mathbb{F}_A)| - |\varGamma_1(\mathbb{F}_A)| = 0 \qquad |\varGamma_{91}(\mathbb{F}_C)| - |\varGamma_1(\mathbb{F}_C)| = 0 \qquad |\varGamma_{91}(\mathbb{F}_B)| - |\varGamma_1(\mathbb{F}_B)| = 2$$

That is, on $\mathbb{F}_A$ and $\mathbb{F}_C$ the number of solutions matches, but not on $\mathbb{F}_B$ since there is a difference of two. This is exactly where the above introduced artifact solutions can be observed. However, this approach has two disadvantages. First, it is an *a postiori* measure, i.e. computation of two frequency responses of different ansatz orders is required. Second, the way we presented the counting of solutions requires to count for *all* frequencies $\varOmega \in \mathbb{F}$ which is, of course, not feasible in finite precision. Instead, either a sampling of the frequency range $\mathbb{F}$ or a comparison of the frequency components of all points of the two sets $\varGamma_1(\mathbb{F})$ and $\varGamma_{91}(\mathbb{F})$ subject to a given frequency tolerance $\varepsilon_\varOmega > 0$ would be required. However, this approach is not likely to be numerically robust, which is why we consider it to be of rather academic nature.

4.2. Geometric measures

In this part, we discuss if the two normalized distance measures introduced in Section 3.4 applied to the test case can be used as artifact solution identifiers. Since both the normalized arclength distance and the normalized discrete Fréchet distance require the arclength of the solution branches we start by considering convergence of the arclength, as depicted in Fig. 9a. It shows the arclength of the standard and nose branch over the approximation order n. Apparently, the arclength of the standard branch B_n^s converged quite quickly to a value of $L(B_{91}^s) = 2.16$ with an error $|L(B_{91}^s) - L(B_{75}^s)| < \varepsilon = 10^{-14}$. In contrast, the arclength of the nose branch B_{91}^n converged noticeably slower to a value of $L(B_{91}^n) = 0.77$ with an error $|L(B_{91}^n) - L(B_{75}^n)| \approx 1.03 \cdot 10^{-2}$.

Next, we discuss the results of the normalized arclength distance $\overline{d}_L$ and the normalized discrete Fréchet distance $\overline{d}_{DF}$ plotted over the approximation order n, as depicted in Fig. 9b. For both distance measures, it can be observed that, again, the standard branch converges noticeably faster than the nose branch. For this reason, we choose a linear scale of the diagram for both distance measures in order to be able to better compare the qualitative behavior of each of the curves. Upon comparing the standard branch of ansatz orders $n = 75, 91$, the distance measures yield $\overline{d}_L(B_{75}^s, B_{91}^s) < 10^{-14}$ and $\overline{d}_{DF}(B_{75}^s, B_{91}^s) = 1.14 \cdot 10^{-13}$. However, for the nose branch, the two distance measures yield noticeably larger errors of $\overline{d}_L(B_{75}^n, B_{91}^n) = 1.32 \cdot 10^{-2}$ and $\overline{d}_{DF}(B_{75}^n, B_{91}^n) = 8.08 \cdot 10^{-3}$. Note that for $n = 1$, the value of the two distance measures of the nose branch are noticeably larger compared to the values of the standard branch, i.e. $\overline{d}_L(B_1^n, B_{91}^n) = 0.49$ versus $\overline{d}_L(B_1^s, B_{91}^s) = 0.003$ and $\overline{d}_{DF}(B_1^s, B_{91}^s) = 0.65$ versus $\overline{d}_{DF}(B_1^s, B_{91}^s) = 0.03$. This amounts to a difference of roughly one order of magnitude. Since both distance measures are normalized by the arclength of the respective reference solution branch, this difference is noticeable. However, it is not yet clear if this is characteristic behavior

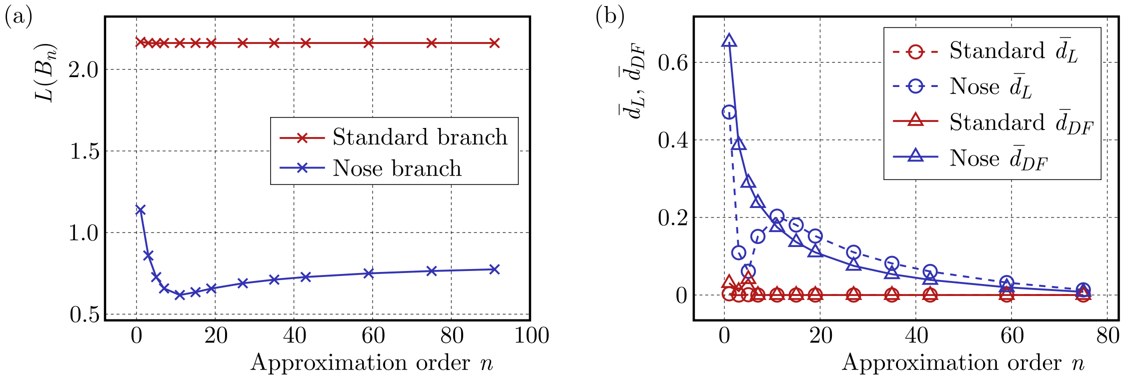

Fig. 9. Geometric measures for the test case (2.9): (a) arclength $L(B_n)$, (b) normalized arclength distance $\bar{d}_L(B_n, B_{n_{ref}})$ and normalized discrete Fréchet distance $\bar{d}_{DF}(B_n, B_{n_{ref}})$

of the artifact solutions. At this point, further studies for different parameter sets for the Duffing system might provide deeper insight. Additionally, consideration of rather large deviations in the amplitudes for frequencies $\Omega < \Omega_{1,tp}$ (cf. Fig. 1) suggests that this is also contributing to somewhat large normalized distance measures of the nose. One would have to filter out the amplitude part of the errors w.r.t. the normalized distances in order to better assess if these distance measures are suitable artifact solution identifiers. A possible way to get the frequency part of the difference of two curve points $A - B$ w.r.t. the Fréchet distance would be to modify the Euclidean norm $\|A - B\|_2$ to a weighted norm, i.e. $\|\mathbf{W}(A - B)\|_2$ with the diagonal matrix $\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix}$ for $0 < \epsilon \leqslant 1$. Computing either of the above distance measures might be more expensive than identifying artifact solutions by the turning points of two compared solution branches within Definition 1. However, a conclusive complexity analysis is yet missing.

4.3. Solver measures

In this last part, we seek to investigate if the solver-related quantities allow for a detection of artifact solutions. For this, we focus on the number of correction steps k per prediction step of the employed numerical continuation method which is depicted in Fig. 10. This figure shows the number of correction (Newton iteration) steps k over the entire frequency range of test case (2.9). Here, we only plotted the curves for the ansatz order $n = 1, 3$ to not clutter the diagram. For the ansatz order $n = 1, 3$, the nose branch exhibits values of k in the range 0 to 20 and 0 to 31, and the standard branch values of 0 to 17 and 0 to 18, respectively. Since the employed numerical continuation method starts at $\Omega = 0$ with pre-computed initial guesses $\mathbf{c}_n^0$, the canonical values

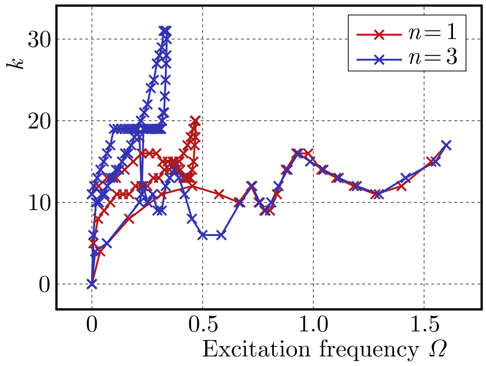

Fig. 10. Number of correction steps k per excitation frequency for the test case (2.9)

of $k = 0$ at this frequency can be observed. Additionally, the smallest, largest and average number of correction steps over all ansatz orders $n = 1, 3, \dots, 91$ and excitation frequencies are 0, 15.2 and 41, respectively. Similar to the diagrams of the condition number in Fig, 7a and 7b, there is a noticeable peak in the number of correction steps at the nose branch turning point. However, in the characteristic frequency range around the turning point of the reference solution at $\Omega = 0.21$, neither for $n = 1$ nor for $n = 3$, there is a noticeable change of the number of correction steps. Hence, this measure is also not indicative for the studied artifact solutions. We end this Section by presenting the solver-related quantities computation time t_{comp} in seconds as well as the number of prediction steps k_{pred}, the total number of correction steps k_{corr} and the average number of correction steps $\overline{k}_{corr}$, all per solution branch and against the approximation order n in Fig. 11. All computations were performed on a 64 bit `Ubuntu 22.04.03 LTS` operating system with an AMD Ryzen 7 Pro 4750U CPU and 32 GB of RAM. For the computation time, an expected exponential increase upon the increase of the ansatz order n is observed. The noticeable outliers of the nose branch at $n = 1, 59$ are attributed to an additional computational load of the operating system. The number of prediction steps k_{pred} and the total number of correction steps k_{corr} of the nose branch exhibit an almost exponential increase as well. However, for the standard branch, the number of prediction steps k_{pred} and the total number of correction steps k_{corr} already converged for $n \geqslant 3$ around values of 31 and 350, respectively. Finally, consider the average number of correction steps $\overline{k}_{corr}$. For the nose and standard branch, this value converges to values around 12.4 and 11.2, respectively. Interestingly, for lower ansatz orders, both solution branches exhibit a larger average number of correction steps compared to the value for the reference ansatz order of $n = 91$. In particular, over the course of the ansatz order increase from $n = 7$ to $n = 91$, the nose branch exhibits a decrease of $\overline{k}_{corr}$ by about a third.

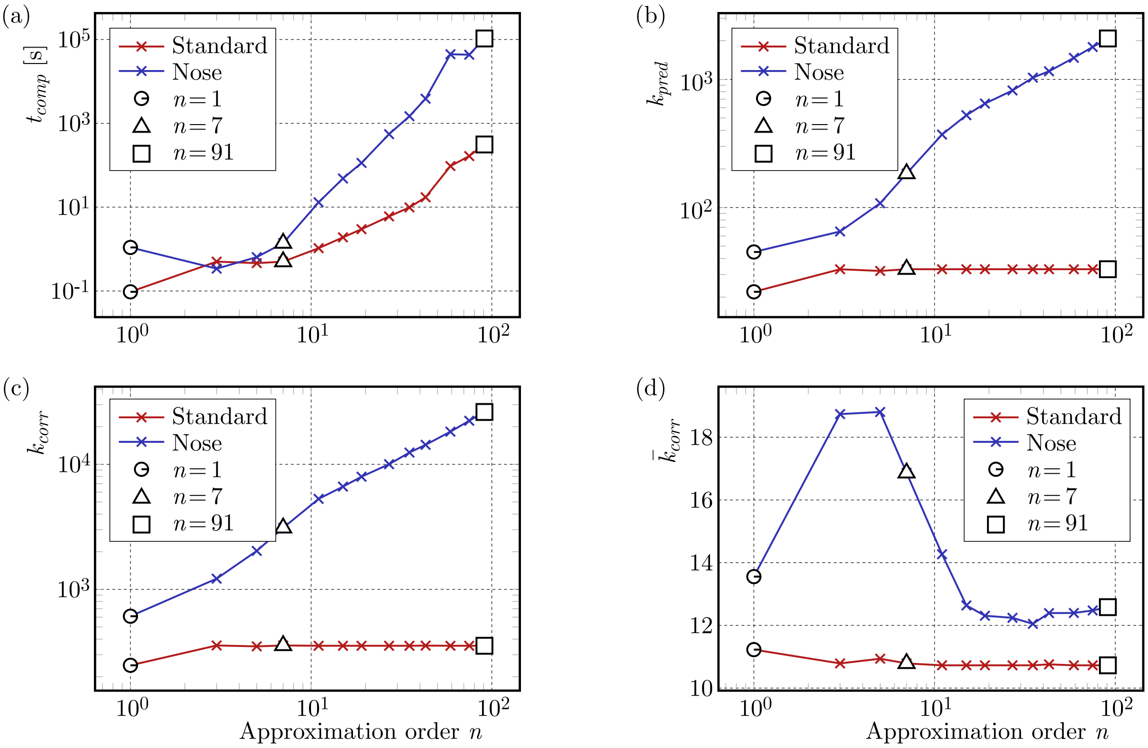

Fig. 11. Computation time t_{comp}, number of prediction steps k_{pred}, total number of correction steps k_{corr} and average number of correction steps $\overline{k}_{corr}$ for the test case (2.9)

5. Conclusions

In this work, we discuss several qualitative error measures in order to characterize the so-called artifact behavior that occurs during computation of HBM solutions for the softening Duffing oscillator. In particular, we provide a mathematical definition of artifact solutions in which the turning points of two solution branches of different ansatz orders are compared. This allows for an *a posteriori* identification of artifact solutions based solely on a robust computation of turning points. Additionally, of the residual, geometric, algebraic and solver-related error measures, investigating only the approach of counting the number of computed solutions, yields the desired discontinuity at the frequency value $\Omega = 0.21$ of the turning point of the reference solution branch. However, this *a posteriori* approach is of rather academic nature and expected to be not robust in the numerical practice. Furthermore, unfortunately, none of the examined error measures potentially showed to be *a priori* indicative of artifact behavior but only *a posteriori*. A possible explanation for the lack of an artifact-related characteristic behavior of the investigated measure lies in the fact that up to now, static error measures have been considered. This means that the value of quantities under examination was always evaluated for a specific order of development only. What remained unconsidered is the dependence of the error measures on the rate of change of the truncation order. Additionally, further studies are required to connect the concept of artifact solutions with the existing error measures such as, e.g., Urabe (1965), Kogelbauer and Breunung (2021), Woiwode and Krack (2023). Consequently, the following question may be raised: Do artifact solutions exist for truncation orders that can be deemed "sufficiently large" as by the measures of the aforementioned authors? Furthermore, the authors plan to publish further studies on their Python codes for the HBM algebraic system generation and the employed numerical path continuation solver. Among others, the presented definition of solution artifacts as a theoretical foundation as well as the Fréchet distance between solution branches of different truncation orders should be further studied as *a posteriori* artifact identifiers. In this context, application to different Duffing parameter sets as well as other nonlinear systems needs investigation to further assess the robustness in numerical practice.

Acknowledgements

This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number LE 4694/1-1. The authors wish to thank Volker Mehrmann and the members of his group for their advice during many helpful discussions.

References

1. Alt H., Godau M., 1995, Computing the Fréchet distance between two polygonal curves, *International Journal of Computational Geometry and Applications*, **5**, 75-91

2. Basu S., Pollack R., Roy M., 2006, *Algorithms in Real Algebraic Geometry*, 2nd ed., Springer Berlin, Heidelberg

3. De Terán F., Dopico F.M., Pérez J., 2013, Condition numbers for inversion of Fiedler companion matrices, *Linear Algebra and its Applications*, **439**, 4 944-981

4. Deuflhard P., 2011, *Newton Methods for Nonlinear problems: Affine Invariance and Adaptive Algorithms*, Springer Series in Computational Mathematics, Springer Berlin, Heidelberg

5. Deuflhard P., Fiedler B., Kunkel P., 1987, Efficient numerical pathfollowing beyond critical points, *SIAM Journal on Numerical Analysis*, **24**, 4, 912-927

6. Deuflhard P., Hohmann A., 2019, *Eine algorithmisch orientierte Einführung*, De Gruyter, Berlin, Boston

7. Duffing G., 1918, *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung*, Samlung Vieweg

8. Edelman A., Murakami H., 1995, Polynomial roots from companion matrix eigenvalues, *Mathematics of Computation*, **64**, 763-776

9. Ferri A.A., Leamy M.J., 2009, Error estimates for harmonic-balance solutions of nonlinear dynamical systems, *Collection of Technical Papers – AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, May, 1-15

10. Figueira J.P., 2023, *Discret-frechet*, https://github.com/joaofig/discrete-frechet

11. García-Saldaña J.D., Gasull A., 2013, A theoretical basis for the Harmonic Balance Method, *Journal of Differential Equations*, **254**, 1, 67-80

12. Hagedorn P., 1981, *Non-linear Oscillations*, Clarendon Press, Oxford and New York

13. Herman R.L., 2016, *An Introduction to Fourier Analysis*, CRC Press

14. Holmes P.J., Rand D.A., 1976, The bifurcations of Duffing's equation: An application of catastrophe theory, *Journal of Sound and Vibration*, **44**, 2, 237-253

15. Kogelbauer F., Breunung T., 2021, When does the method of harmonic balance give a correct prediction for mechanical systems, *Applicable Analysis*, **102**, 2, 425-443

16. Kovacic I., Brennan M.J., 2011, *The Duffing Equation: Nonlinear Oscillators and their Phenomena*, Wiley

17. Krack M., Gross J., 2019, *Harmonic Balance for Nonlinear Vibration Problems*, Springer

18. Lentz L., von Wagner U., 2020, Avoidance of artifacts in harmonic balance solutions for nonlinear dynamical systems, *Journal of Theoretical and Applied Mechanics*

19. Nayfeh A.H., Mook D.T., 1979, *Nonlinear Oscillations*, Wiley

20. Novak S., Frehlich R.G., 1982, Transition to chaos in the Duffing oscillator, *Physical Review A*, **26**, 6, 3660-3663

21. Strogatz S.H., 1994, *Nonlinear Dynamics and Chaos*, Westview Press

22. Ueda Y., 1991, Survey of regular and chaotic phenomena in the forced Duffing oscillator, *Chaos, Solitons and Fractals*, **1**, 3, 199-231

23. Urabe M., 1965, Galerkin's procedure for nonlinear periodic systems, *Archive for Rational Mechanics and Analysis*, **20**, 120-152

24. Von Wagner U., Lentz L., 2016, On some aspects of the dynamic behavior of the softening Duffing oscillator under harmonic excitation, *Archive of Applied Mechanics*, **86**, 1383-1390

25. Von Wagner U., Lentz L., 2018, On artifact solutions of semi-analytic methods in nonlinear dynamics, *Archive of Applied Mechanics*, **88**, 1713-1724

26. Von Wagner U., Lentz L., 2019, On the detection of artifacts in Harmonic Balance solutions of nonlinear oscillators, *Applied Mathematical Modelling*, **65**, 408-414

27. Woiwode L., Balaji N.N., Kappauf J., Tubita F., Guillot L., *et al.*, 2020, Comparison of ANM and predictor-corrector method to continue solutions of harmonic balance equations, [In:] *Conference Proceedings of the Society for Experimental Mechanics Series*

28. Woiwode L., Krack M., 2023, Are Chebyshev-based stability analysis and Urabe's error bound useful features for Harmonic Balance?, *Mechanical Systems and Signal Processing*, **194**, 110265

Manuscript received February 14, 2024; accepted for print February 29, 2024