Global smooth solution curves using rigorous branch following

Jan Bouwe van den Berg∗ Jean-Philippe Lessard†
Konstantin Mischaikow‡

Abstract
In this paper, we present a new method to rigorously compute smooth branches of zeros of nonlinear operators $f : \mathbb{R}^{l_1} \times B_1 \to \mathbb{R}^{l_2} \times B_2$, where $B_1$ and $B_2$ are Banach spaces. The method is first introduced for parameter continuation and then generalized to pseudo-arclength continuation. Examples in the context of ordinary, partial and delay differential equations are given.

1 Introduction
Finding solutions of a nonlinear functional differential equation

$$G(p, u) = 0,$$

where $p$ is a set of parameters, is central in mathematics. In particular, when (1) takes the form of a partial differential equation or a delay equation, finding explicit solutions becomes a real challenge, due to the nonlinearity of $G$ and to the fact that the state space is infinite dimensional.

Several computer-assisted approaches to solve rigorously systems of nonlinear equations have been proposed since the early 1990’s [2, 4, 5, 7, 10, 11, 12, 13, 14, 15, 17, 18]. A combination of topological methods (Banach fixed point theorem, Conley index theory), a priori analytic estimates and use of interval arithmetic have led to new theorems about the existence of solutions. In early works like in [17, 18], the proofs of existence were done for fixed parameters. In [5, 7] these arguments were put in a context of continuation, where a premium was placed on minimizing computational cost, the focus remained on discrete parameter values only. This method was referred to as validated continuation. In [4, 10], continuous branches of solution curves were obtained, in the context of a predictor-corrector algorithm. The idea was to directly work with small intervals of parameters (using interval arithmetic) and then draw conclusions about solution branches for these intervals of parameters. However, the computational cost of such methods is high, since trivial predictors were used, leading to very small step sizes in the parameter.

∗VU University Amsterdam, Department of Mathematics, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands.
†Rutgers University, Department of Mathematics, Hill Center-Busch Campus, 110 Frelinghuysen Rd, Piscataway, NJ 08854-8019, USA. and VU University Amsterdam, Department of Mathematics, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands.
‡Rutgers University, Department of Mathematics, Hill Center-Busch Campus, 110 Frelinghuysen Rd, Piscataway, NJ 08854-8019, USA.
In [1], validated continuation was adapted to prove the existence of piecewise continuous solution curves of (1). At each step of the algorithm, first order predictors were used to prove the existence of small continuous solutions curves, allowing significantly larger step sizes. With this in mind, we now aim to develop a method that allow us to rigorously obtain the existence of *global* smooth solution curves, in the context of both parameter and pseudo-arclength continuation.

Before proceeding, it is worth mentioning that this method might as well be applied to finite dimensional systems. However, the motivation for applying rigorous numerical techniques to such problem is less appealing, as the confidence in getting reliable outputs from classical numerical methods is high, since the main source of error is often due to round-off. In the context of infinite dimensional problems, the numerical methods must be applied to some finite dimensional approximation, which raises questions concerning the validity of the output. With this in mind, we develop a method that provides an internal check of consistency on the dimension of truncation from the infinite to finite dimensional problem, hence delivering rigorous mathematical proofs.

When looking for solutions of (1) with a periodic profile, one may apply a Fourier transformation to the a priori unknown solution \( u \) and then solve for the Fourier coefficients. This transforms (1) into an equivalent problem in Fourier space. We will turn to concrete examples quickly, where we also specify the parameters and spaces involved, but we first introduce the general setting and notation. Denote by \( g : \mathbb{R}_{l1} \times B_{l1} \rightarrow B_{l2} \) the Fourier transformation of \( G \), where \( \mathbb{R}_{l1} \) is the parameter space and \( B_{l1}, B_{l2} \) are Banach spaces. Sometimes, we will be interested in finding solutions of \( g = 0 \) satisfying additional conditions (see Examples 1 and 2 below). An extra set of \( l_2 \) equations will then ensure that the additional conditions are satisfied, i.e., \( h = 0 \) with \( h : \mathbb{R}_{l1} \times B_{l1} \rightarrow \mathbb{R}_{l2} \). Hence, consider the infinite dimensional system of equations

\[
F : \mathbb{R}_{l1} \times B_{l1} \rightarrow \mathbb{R}_{l2} \times B_{l2} : (p, \xi) \mapsto F(p, \xi) = (h(p, \xi), g(p, \xi)) = 0,
\]

(2)

where \( p = (p_1, \ldots, p_{l1}) \in \mathbb{R}_{l1} \) are the original parameters of (1) and \( \xi \) consists of the Fourier coefficients of \( u \). To be more specific, we denote these by \( \xi = (\xi_k)_{k \geq 0} \) with, in general, \( \xi_k \in \mathbb{R}^n, n \geq 1 \) (see below for examples where \( n = 1, 2 \); one may also think of systems of equations and higher dimensional spatial settings, leading to larger \( n \)). In this paper, we do not deal with the details of the equivalence (for periodic solutions) of (1) and (2), which will be context dependent. Let us remark that although in the present paper we restrict attention to periodic solutions, extensions to non-periodic (boundary value) problems are possible within this setting.

Since the periodic solutions of (1) we are looking for, are reasonably smooth, we choose our Banach spaces such that the Fourier coefficients \( \xi = (\xi_k)_{k}, \xi_k \in \mathbb{R}^n \) decay quickly. There are of course many possibilities. We only deal with one popular choice, used in [1, 4, 5, 6, 7], mostly in the context of validated continuation. We choose weight functions \( (q > 0) \)

\[
\omega_k^q = \begin{cases} 
1, & k = 0; \\
k^q, & k \geq 1,
\end{cases}
\]

(3)

which are used to define the norm

\[
\|\xi\|_q = \sup_{k \geq 0} \omega_k^q|\xi_k|_{\infty},
\]

(4)

and the Banach space

\[
\Omega^q = \{\xi, \|\xi\|_q < \infty\},
\]

(5)
consisting of sequences with algebraically decaying tails. We finally let \( B_1 = \Omega^{q_1} \) and \( B_2 = \Omega^{q_2} \). Throughout we assume that \( \mathcal{F} \) is a \( \mathcal{C}^1 \) function.

**Example 1.** Consider the problem of computing periodic solutions (with a special symmetry) of the fourth order Swift-Hohenberg ordinary differential equation

\[
-u^{iv} - \nu u'' + u - u^3 = 0, \quad \nu \in \mathbb{R}.
\tag{6}
\]

This ODE has a conserved quantity (first integral), called the energy, which is given by

\[
E = u'''u' - \frac{1}{2} u''^2 + \frac{\nu}{2} u'^2 + \frac{1}{4} (u^2 - 1)^2.
\]

We restrict our attention to finding periodic solutions at the zero energy level \( E = 0 \). Plugging the cosine Fourier expansion

\[
u(y) = \xi_0 + 2 \sum_{k \geq 1} \xi_k \cos kLy
\]

into (6), the problem \( g = (g_k)_{k \geq 0} = 0 \), where

\[
g_k \overset{\text{def}}{=} \left[ 1 + \nu L^2 k^2 - L^4 k^4 \right] \xi_k - \sum_{k_1 + k_2 + k_3 = k} \xi_{k_1} \xi_{k_2} \xi_{k_3}, \quad k \geq 0,
\tag{7}
\]

corresponds to finding periodic solutions \( u \) of (6), see [1]. Here, \( p = (p_1, p_2) \in \mathbb{R}^2 \), where \( p_1 = \nu \) and \( p_2 = L \) is the frequency of \( u \) (\( \frac{2\pi}{L} \) is its period). The extra equation

\[
h \overset{\text{def}}{=} -2L^2 \sum_{l=1}^{\infty} l^2 \xi_l - \frac{1}{\sqrt{2}} \left[ \xi_0 + 2 \sum_{l=1}^{\infty} \xi_l \right]^2 + \frac{1}{\sqrt{2}} = 0
\]

is added in order to ensure that \( E = 0 \) (one evaluates the energy at \( y = 0 \), where \( u' = 0 \)). Letting \( \mathcal{F} = (h, g) \), the problem \( \mathcal{F}(p, \xi) = 0 \) is considered, with \( \mathcal{F} : \mathbb{R}^2 \times \Omega^q \to \mathbb{R} \times \Omega^{q-4} \), \( q > 3 \), see also Section 4.2.

**Example 2.** Consider the problem of finding periodic solutions of the so-called Wright’s delay equation

\[
y'(t) = -\alpha y(t-1)[1 + y(t)], \quad \alpha > \frac{\pi}{2}
\tag{8}
\]

considered in [16]. Plugging the Fourier expansion

\[
y(t) = \xi_{0,1} + 2 \sum_{k=1}^{\infty} \left[ \xi_{k,1} \cos kLt - \xi_{k,2} \sin kLt \right]
\]

into (8) and letting \( \xi_k = (\xi_{k,1}, \xi_{k,2}) \in \mathbb{R}^2 \) (with \( \xi_{0,2} = 0 \)), consider

\[
g_k \overset{\text{def}}{=} R_k \left( \begin{array}{l}
\xi_{k,1} \\
\xi_{k,2}
\end{array} \right) + \alpha \sum_{k_1 + k_2 + k_3 = k} \Theta_{k_1} \left( \begin{array}{l}
\xi_{k_1,1} \xi_{k_2,1} - \xi_{k_1,2} \xi_{k_2,2} \\
\xi_{k_1,1} \xi_{k_2,2} + \xi_{k_1,2} \xi_{k_2,1}
\end{array} \right), \quad k \geq 0,
\]

where

\[
R_k = \left( \begin{array}{cc}
\alpha \cos kL & -kL + \alpha \sin kL \\
kL - \alpha \sin kL & \alpha \cos kL
\end{array} \right) \quad \text{and} \quad \Theta_{k_1} = \left( \begin{array}{cc}
\cos k_1 L & \sin k_1 L \\
-\sin k_1 L & \cos k_1 L
\end{array} \right).
\]
Solving \( g = (g_h)_{h \geq 0} = 0 \) corresponds to finding periodic solutions of (8), see [9]. In order to eliminate arbitrary shifts of the periodic solution \( y \), the normalizing condition \( y(0) = 0 \) is imposed. Hence,

\[
h \underset{\text{def}}{=} y(0) = \xi_{0,1} + 2 \sum_{k=1}^{\infty} \xi_{k,1} = 0
\]

is appended to \( g = 0 \). Letting \( F = (h, g) \) and \( p = (\alpha, L) \in \mathbb{R}^2 \), the problem \( F(p, \xi) = 0 \) is considered, with \( F: \mathbb{R}^2 \times \Omega^1 \to \mathbb{R} \times \Omega^{n-1}, q \geq 2 \).

**Example 3.** Consider the problem of looking for stationary solutions of nonlinear partial differential equations of the form

\[
u u = \nu u - (1 + \Delta)^2 u - u^3 = 0, \quad \text{in } D = \prod_{i=1}^{N} \left[ 0, \frac{2 \pi}{L_i} \right]
\]

defined on \( N \)-dimensional rectangular spatial domains, where \( \mathcal{L} \) is a linear differential operator in \( u \). In particular, consider the two-dimensional problem \( N = 2 \), \( \mathcal{L}(u) = (\nu - (1 + \Delta)^2)u, \quad P = 3, \quad c_2 = 0, \quad c_3 = -1 \) with periodic boundary conditions, see [6]. More precisely, consider

\[
u u = \nu u - (1 + \Delta)^2 u - u^3 = 0, \quad \text{in } D = \left[ 0, \frac{2 \pi}{L_1} \right] \times \left[ 0, \frac{2 \pi}{L_2} \right]
\]

\[
u u = \nu u - (1 + \Delta)^2 u - u^3 = 0, \quad \text{in } D = \left[ 0, \frac{2 \pi}{L_1} \right] \times \left[ 0, \frac{2 \pi}{L_2} \right]
\]

\[
u u = \nu u - (1 + \Delta)^2 u - u^3 = 0, \quad \text{in } D = \left[ 0, \frac{2 \pi}{L_1} \right] \times \left[ 0, \frac{2 \pi}{L_2} \right]
\]

Plugging the expansion of the time independent a priori unknown solution

\[u(x, y) = \sum_{k_1, k_2 \in \mathbb{Z}} c_{k_1, k_2} e^{ik_1 L_1 y} e^{ik_2 L_2 y}\]

into (10), we need to solve

\[g_{i,j}(\nu, \xi) \underset{\text{def}}{=} \mu_{i,j}(\nu) \xi_{i,j} - \sum_{i_1 + i_2 + i_3 = i, j_1 + j_2 + j_3 = j} \xi_{i_1,j_1} \xi_{i_2,j_2} \xi_{i_3,j_3} = 0, \quad i, j \geq 0 \tag{11}\]

where \( \xi_{i,j} \) is the real part of \( c_{i,j}, \xi = (\xi_{i,j})_{i,j \geq 0}, \xi_{i,j} = (1 - j^2 L_1^2 + j^2 L_2^2)^{\frac{1}{2}}. \) Letting \( F = (g_{i,j})_{i,j \geq 0} \), solving \( F(\nu, \xi) = 0 \) corresponds to finding solutions of (10).

1.1 Parameter continuation

We want to develop a computational method to rigorously continue the zeros of \( F: \mathbb{R}^{1} \times \Omega^1 \to \mathbb{R}^{2} \times \Omega^2 \), as we move one of the parameters of \( p \), say \( p_1 \). We introduce only the main ideas here, and discuss the method in detail in Section 2. Fixing the parameters \( p_2, \ldots, p_t \) and considering \( \nu \underset{\text{def}}{=} p_1 \) as the continuation parameter, we define the infinite dimensional vector of variables \( x = (p_2, \ldots, p_t, \xi) \) and the new map

\[f: \mathbb{R} \times \left[ \mathbb{R}^{1} \times \Omega^1 \right] \to \mathbb{R}^{2} \times \Omega^2 : (\nu, x) \mapsto f(\nu, x). \tag{12}\]
Under the assumption that $D_x f(\nu, x)$ is nonsingular along the branch of zeros that we are computing, we vary the parameter $\nu$. In this case, the implicit function theorem implies that the branch of zeros can be viewed globally as the graph of a function of the parameter $\nu$. The idea is to transform the problem $f(\nu, x) = 0$ into a fixed point equation and to apply the Banach fixed point theorem. Since we want to develop this idea in a computational setting, consider a finite dimensional projection $f^{(m)}$ of (12). First, using a Newton-like iterative scheme on $f^{(m)}$, we compute an approximate zero $\bar{x}$ of (12) at the parameter value $\nu = \nu_0$. Next, we compute a tangent vector $\dot{x}$ such that $D_x f(\nu_0, \bar{x}) \dot{x} + D_x f(\nu_0, \bar{x}) \approx 0$. Using the vectors $\bar{x}$ and $\dot{x}$, we define the set of predictors by

$$ x_\nu = \bar{x} + \Delta_\nu \dot{x}, \quad (13) $$

where $\Delta_\nu$ is small. Consider the Banach space $\Phi = \mathbb{R}^{l-1} \times \Omega^n$ (with the induced product norm). We compute an approximate inverse $A$ of the linear operator $D_x f(\nu_0, \bar{x})$. For $\nu = \nu_0 + \Delta_\nu$ close to $\nu_0$, we define $T_\nu : \Phi \to \Phi$ by

$$ T_\nu(x) = x - A f(\nu, x), \quad (14) $$

and look for a fixed point of $T_\nu$ using the Banach fixed point theorem. Note that it is sufficient that $A$ is injective to ensure that fixed points of $T$ are in bijection with zeros of $f$.

**Example 4.** For the problem introduced in Example 1, the approximate inverse $A$ may be constructed as follows [1]. Denote by $D_x f^{(m)}(\nu_0, \bar{x})$ the Jacobian matrix of the projection $f^{(m)}$ at the approximate solution $(\nu_0, \bar{x})$, and let $J_m$ be an approximate inverse of $D_x f^{(m)}(\nu_0, \bar{x})$, computed using an LU decomposition. Recalling (7), denote the linear part of $g_k$ by $\mu_k(L, \nu) = 1 + \nu L^2 k^2 - L^4 k^4$. Now we define

$$ A \stackrel{\text{def}}{=} \begin{bmatrix} J_m & 0^T & 0^T & 0^T & \cdots \\ 0_F & \mu_m(L, \nu_0)^{-1} & 0 & 0 & \cdots \\ 0_F & 0 & \mu_{m+1}(L, \nu_0)^{-1} & 0 & \cdots \\ 0_F & 0 & 0 & \mu_{m+2}(L, \nu_0)^{-1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, $$

which acts as an approximate inverse of the linear operator $D_x f(\nu_0, \bar{x})$, provided of course that the projection dimension $m$ is large enough. Note that $A : \mathbb{R} \times \Omega^{q-4} \to \mathbb{R}^2 \times \Omega^q$.

The goal is to prove that there exists a ball $B(r, \Delta_\nu) = x_\nu + B(r) \subset \Phi$ of radius $r$ using norm (4), centered at $x_\nu$, such that $T_\nu$ maps the ball $B(r, \Delta_\nu)$ into itself and acts as a contraction on $B(r, \Delta_\nu)$, for small $\Delta_\nu = \nu - \nu_0$. To verify these conditions, we need to compute two bounds $Y = Y(\Delta_\nu)$ and $Z = Z(r, \Delta_\nu)$. In essence, $Y$ measures how far the center $x_\nu$ of $B(r, \Delta_\nu)$ is mapped from itself (under $T_\nu$), whereas $Z$ measures the contraction rate of (all components of) $T_\nu$ on $B(r, \Delta_\nu)$. The most computationally demanding part of the method is the construction of the bounds $Y$ and $Z$ (see for instance Sections 3.2 and 3.3 in [1] or Section 6 in [5]). Their construction requires a combination of a priori analytic estimates (bounds on the truncation error terms) and rigorous computations involving interval arithmetic. Once the bounds $Y$ and $Z$ are computed, verifying that

$$ \|Y(\Delta_\nu) + Z(r, \Delta_\nu)\|_\Phi < r \quad (15) $$

is sufficient to conclude that $T_\nu : B(r, \Delta_\nu) \to B(r, \Delta_\nu)$ is a contraction (see Lemma 5 and [17]), yielding a unique zero of $f(\nu, x)$ at $\nu = \nu_0 + \Delta_\nu$. In practice, we use an iterative
procedure (with $\Delta_\nu$ varying) to find the approximate maximal $\Delta_0^\nu$ for which there exists an $r > 0$ such that (15) is satisfied (see Section 2 and [1]). If this step is successful, let $\nu_1 = \nu_0 + \Delta_0^\nu$. We then have a continuum of zeros $C_0 = \{ (\nu, x^0(\nu)) \mid f(\nu, x^0(\nu)) = 0, \quad \nu \in [\nu_0, \nu_1] \}$, see Lemma 7. Since we want to repeat the argument with initial parameter value $\nu_1$, we then put ourself in the context of a continuation method. This involves a predictor and corrector step. Recalling (13), the predictor at the parameter value $\nu_1 = \nu_0 + \Delta_0^\nu$ is given by $\hat{x}_1 \equiv \hat{x} + \Delta_0^\nu \hat{x}$. The corrector step, based on a Newton-like iterative scheme for the projection $f^{(m)}$, takes $\hat{x}_1$ as its input and produces, within a prescribed tolerance, a zero $\bar{x}_1$ at $\nu_1$. We can then compute a new tangent vector $\bar{x}_1$, build the new set of predictors $\bar{x}_1 + \Delta_\nu \bar{x}_1$, compute the bounds $Y, Z$ at the parameter value $\nu_1$ and try to verify (15) again. If we are successful in finding a new $\Delta_0^\nu$, we let $\nu_2 = \nu_1 + \Delta_0^\nu$ and we get the existence of a continua of zeros $C_1 = \{ (\nu, x^1(\nu)) \mid f(\nu, x^1(\nu)) = 0, \quad \nu \in [\nu_1, \nu_2] \}$. Once we have the two continuum of zeros $C_0$ and $C_1$, we ask the natural question: can we prove that $C_0$ and $C_1$ connect at $(\nu_1, x^0(\nu_1)) = (\nu_1, x^1(\nu_1))$ such that $C_0 \cup C_1$ is a smooth one dimensional branch of solutions of $f = 0$? In turns out that there is a simple check that can be added to the continuation step in order to give an answer to this question, see Proposition 8.

### 1.2 Pseudo-arclength continuation

The rigorous continuation introduced in the previous section requires $D_x f(\nu, x)$ to be nonsingular along the branch of zeros we are following. This implies that the continuation method will necessarily fail when trying to continue past a fold. One way to overcome this difficulty is to consider the continuation parameter $\nu$ as a variable and the arclength of the curve as a new parameter [8]. Consider the vector of variables $X = (p, \xi)$ and recall (2). To solve $F(X) = 0$ past folds, we append one equation to the system, namely the equation $E = 0$ of a plane almost perpendicular to the curve we are following. In practice, we do not know exactly the arclength of the curve. The new continuation parameter, denoted by $s$, will then be the pseudo-arclength of the curve. Note that $E$ depends on $s$.

The details of the construction of $E$ are rather technical and are presented in Section 3. In essence, we apply the rigorous continuation method on $F(s, X) = 0$, where

$$
F(s, X) = \begin{pmatrix} E(s, X) \\ F(X) \end{pmatrix}.
$$

With this construction, note that $D_X F(s, X)$ will be nonsingular at a fold point. Hence, we can expect a Newton-like map to contract neighborhoods of the fold point. In Lemma 10 and Proposition 11 we formulate the algorithms to establish the existence of a smooth solution curve.

The paper is organized as follow. In Section 2, we introduce the parameter continuation method to obtain smooth branches of zeros. In Section 3, we show how to modify the continuation method in order to continue past folds: pseudo-arclength continuation. In Section 4, we first present an example of the parameter continuation in the context of periodic solutions of delay-differential equations. We also discuss an application of the pseudo-arclength method to periodic solutions of ordinary differential equations. This example provides an improvement of a result presented in [1].


2 Parameter continuation

In this section, we develop a method to compute smooth solution curves of

\[ F : \mathbb{R}^l \times \Omega^p \rightarrow \mathbb{R}^l \times \Omega^q, \]

as we move one of the parameters of \( p \). Without loss of generality, we consider \( \nu \overset{\text{def}}{=} p_1 \) as the continuation parameter. Hence, we fix all parameters \( p_2, \ldots, p_l \). Defining the infinite dimensional vector of variables \( x = (p_2, \ldots, p_l, \xi) \), we want to do rigorous branch following for the problem \( f(\nu, x) = 0 \). As mentioned before, we transform this problem into a fixed point problem \( T_\nu(x) = x \). With \( x \) as given above, define the norm

\[ \|x\|_{\Phi} = \max \{|p_2|, \ldots, |p_l|, |\xi|\}, \]

and the corresponding Banach space

\[ \Phi = \{x = (p_2, \ldots, p_l, \xi), \|x\|_{\Phi} < \infty\}. \]

Consider \( \nu_0 \) fixed and suppose the existence of \( \bar{x} \in \Phi \) such that \( f(\nu_0, \bar{x}) = 0 \). Assume we have an bijective linear operator \( A : \mathbb{R}^l \times \Omega^q \rightarrow \mathbb{R}^{l-1} \times \Omega^{q_1} \) which acts as an approximation for the inverse of \( D_2 f(\nu_0, \bar{x}) \). Recalling (14), consider the Newton-like operator \( T_\nu(x) = x - A f(\nu, x) \), with \( \nu \) close to \( \nu_0 \). Suppose also that we have computed a tangent vector \( \hat{x} \in \Phi \) such that \( D_x f(\nu_0, \bar{x}) \hat{x} + D_\nu f(\nu_0, \bar{x}) \approx 0 \). The idea is to find balls in \( \Phi \) on which \( T_\nu \) is a contraction mapping, thus leading to solutions of \( f(\nu, x) = 0 \). Recalling that \( \xi_k \in \mathbb{R}^n \), let us define the ball of radius \( r \) in \( \Phi \), centered at the origin,

\[ B^q(r) = [-r, r]^{l-1} \times \prod_{k=0}^{\infty} \left[ \frac{-r}{\omega_k^l}, \frac{r}{\omega_k^l} \right]^n. \]

We will drop \( q_1 \) from the notation whenever this does not compromise clarity. Recalling (13), consider the predictors based at \( \nu_0 \): \( x_\nu = \bar{x} + \Delta_\nu \hat{x} \), with \( \Delta_\nu = \nu - \nu_0 \). For \( \nu \) close to \( \nu_0 \) we define the ball centered at \( x_\nu \) by \( B_{x_\nu}(r) = x_\nu + B(r) \). To simplify the presentation, define \( k_0 = -l_1 + 1 \), so that the indexing of the sets begins at \( k = k_0 \). To show that \( T_\nu \) is a contraction mapping, we need bounds \( Y_k \) and \( Z_k \) for all \( k \geq k_0 \), such that, with \( \Delta_\nu = \nu - \nu_0 \),

\[ \left| T_\nu(x_\nu) - x_\nu \right|_k \leq Y_k(\Delta_\nu), \]

and

\[ \sup_{b, c \in B(r)} \left| DT_\nu(x_\nu + b) c \right|_k \leq Z_k(r, \Delta_\nu). \]

Note that \( Y_k, Z_k \in \mathbb{R} \) for \( k_0 \leq k < 0 \), and \( Y_k, Z_k \in \mathbb{R}^n \) for \( k \geq 0 \). As mentioned earlier, we refer to Sections 3.2 and 3.3 in [1] or in Section 6 in [5] for explicit computations of the bounds (19) and (20). The following lemma was proved in [1].

**Lemma 5.** Consider \( \nu = \nu_0 + \Delta_\nu \). If there exists an \( r > 0 \) such that \( \|Y + Z\|_\Phi < r \), with \( Y = (Y_k)_{k \geq k_0} \) and \( Z = (Z_k)_{k \geq k_0} \), satisfying (19) and (20), respectively, then \( T_\nu \) is a contraction mapping on \( B_{x_\nu}(r) \) with contraction constant at most \( \|Y + Z\|_\Phi / r < 1 \). Furthermore, there is a unique \( \bar{x}_\nu \in B_{x_\nu}(r) \) such that \( f(\nu, \bar{x}_\nu) = 0 \), and \( \bar{x}_\nu \) lies in the interior of \( B_{x_\nu}(r) \).
For sake of simplicity of the presentation, we assume \( n = 1 \). The generalization of the discussion below for the case \( n \geq 1 \) is straightforward, using component-wise comparison for all vector inequalities concerned. An example with \( n = 2 \) can be found in [9].

The bounds/functions \( Y_k(\Delta_\nu) \) and \( Z_k(r, \Delta_\nu) \) can be constructed so that they are polynomials in \( r \) and \( |\Delta_\nu| \) (note the absolute value) with non-negative coefficients. Of course, in parameter continuation, at each step one is interested in either \( \Delta_\nu > 0 \) or \( \Delta_\nu < 0 \), but we stick with the general setting since using the sign of \( \Delta_\nu \) will only marginally improve the bounds and step size. Also, for sufficiently large \( k \), say \( k \geq M \), one may choose

\[
Y_k = 0, \quad \text{and} \quad Z_k = \tilde{Z}_M \left( \frac{M}{k} \right)^{\gamma_1},
\]

for some \( \tilde{Z}_M = \tilde{Z}_M(r, \Delta_\nu) > 0 \), where \( M \) is a computational parameter (to be discussed in the example presented in Section 4.2). The reason why one can choose \( Y_k = 0 \) for \( k \) large enough is because the quantity \( [T_\nu(x_\nu) - \tilde{x}_\nu]_k \) eventually vanishes. This is due to the fact that \( x_\nu \) has only finitely many non-zero entries (e.g. see Section 3.2 in [1]). In order to verify the hypotheses of Lemma 5 in a computationally efficient way, we introduce the following notion of radii polynomials.

**Definition 6.** Let \( Y_k(\Delta_\nu) = 0 \) and \( Z_k(r, \Delta_\nu) = \tilde{Z}_M(r, \Delta_\nu)(\frac{M}{k})^{\gamma_1} \) for all \( k \geq M \). We define the radii polynomials \( \{p_{k_0}, \ldots, p_{M-1}, p_M\} \) by

\[
p_k(r, |\Delta_\nu|) \overset{\text{def}}{=} \begin{cases} Y_k(\Delta_\nu) + Z_k(r, \Delta_\nu) - \frac{r}{\omega_k}, & k = k_0, \ldots, M - 1; \\ \tilde{Z}_M(r, \Delta_\nu) - \frac{r}{\omega_M}, & k = M, \end{cases}
\]

where we recall that \( Y_k(\Delta_\nu) = Y_k(\Delta_\nu) \) and \( Z_k(r, \Delta_\nu) = Z_k(r, \Delta_\nu) \) are polynomials with non-negative coefficients. In particular, \( p_k \) is increasing in \( |\Delta_\nu| \geq 0 \) and convex in \( r \geq 0 \).

Here, we repeat the discussion presented in [1], as it sheds light on the reason why the radii polynomials \( p_k \) are useful. Some terms of the polynomials \( Y_k \) and \( Z_k \) are close to zero. More precisely,

\[
Y_k \sim \delta_1 + \delta_2 |\Delta_\nu| + O(\Delta_\nu^2), \\
Z_k \sim \delta_3 r + O(\Delta_\nu r, r^2),
\]

where \( \delta_1, \delta_2 \) and \( \delta_3 \) are very small: \( \delta_1 \approx 0 \) because of the choice of \( \bar{x} \), \( \delta_2 \approx 0 \) because the choice of \( \hat{x} \), and \( \delta_3 \approx 0 \) because of the choice of the linear operator \( A \) and the Newton-like map \( T_\nu \). Therefore, the radii polynomials are roughly of the form

\[
p_k(r, \Delta_\nu) \sim (\delta_1 + |\Delta_\nu| \delta_2) - \left( \frac{1}{\omega_k} - \delta_3 \right) r + O(\Delta_\nu^2, r^2).
\]

Hence, for a reasonably large range of \( \Delta_\nu \), one may anticipate to find a small \( r > 0 \) (but not too small) at which all radii polynomials are negative. The following is a slight modification of a result presented in [1].

**Lemma 7.** Recall (2) and suppose that \( F \in C^\ell \left( \mathbb{R}^{1+} \times B_1, \mathbb{R}^{1+} \times B_2 \right), \ell \geq 1 \). If there exists an \( r > 0 \) and a small \( \Delta_\nu \) such that \( p_k(r, |\Delta_\nu|) < 0 \) for all \( k = k_0, \ldots, M \), then there exists a \( C^\ell \) function \( \bar{x} : [v_0 - \Delta_\nu, v_0 + \Delta_\nu] \rightarrow \Phi : \nu \mapsto \bar{x}(\nu) \) such that \( f(\nu, \bar{x}(\nu)) = 0 \) for all \( \nu \in [v_0 - \Delta_\nu, v_0 + \Delta_\nu] \). Furthermore, these are the only solutions of \( f(\nu, x) = 0 \) in the tube \( \{ |\nu - v_0| \leq \Delta_\nu, x - x_\nu \in B(r) \} \).
Proof. Since \( p_k \) is increasing in \(|\Delta_\nu| \geq 0\), existence and uniqueness of a solution \( \tilde{x}(\nu) \) for \( \nu \in [\nu_0 - \Delta_\nu, \nu_0 + \Delta_\nu] \) follows from the definition of the radii polynomials and Lemma 5. In particular, for every fixed \( \nu \in [\nu_0 - \Delta_\nu, \nu_0 + \Delta_\nu] \), \( T_\nu : B_{\nu_0} \to B_{\nu_0} \) is a contraction. Consider the change of variable \( y = x - \nu_0 \). Then, the operator

\[
\tilde{T} : [\nu_0 - \Delta_\nu, \nu_0 + \Delta_\nu] \times B(r) \to B(r) : (\nu, y) \mapsto \tilde{T}(\nu, y) \overset{\text{def}}{=} T_\nu(y + \nu_0)
\]

is a uniform contraction on \( B(r) \). Since \( F \in C^l \left( \mathbb{R}^l \times B_1, \mathbb{R} \times B_2 \right) \), we have that \( \tilde{T} \in C^l \left( [\nu_0 - \Delta_\nu, \nu_0 + \Delta_\nu] \times B(r), B(r) \right) \). By the uniform contraction principle, see e.g. [3], we conclude that \( \tilde{x}(\nu) \) is a \( C^l \) function of \( \nu \).

After one successful step, based at \((\nu, x) = (\nu_0, \bar{x}_0)\) with predictor \( \bar{x}_0 \) and step size \( \Delta_\nu \), we find the corrector \( \tilde{x}_1 \) at \( \nu = \nu_1 = \nu_0 + \Delta_\nu \) using a Newton iteration, and we rebuild the radii polynomials, now based at \((\nu, x) = (\nu_1, \tilde{x}_1)\). Suppose now that we have performed two successful continuation steps, i.e., in both steps we have found radii \( r_0 \) and \( r_1 \), respectively, where the radii polynomials are negative. We thus have two continuous solution graphs over intervals \([\nu_0, \nu_1]\) and \([\nu_1, \nu_2]\). Lemma 7 implies the existence of two functions \( x^0(\nu) \) and \( x^1(\nu) \) of class \( C^l \) such that \( C_0 \overset{\text{def}}{=} \{ (\nu, x^0(\nu)) \mid \nu \in [\nu_0, \nu_1] \} \) and \( C_1 \overset{\text{def}}{=} \{ (\nu, x^1(\nu)) \mid \nu \in [\nu_1, \nu_2] \} \) are smooth branches of solutions of \( f(\nu, x) = 0 \). The question is to determine whether or not \( C_0 \) and \( C_1 \) connect at the parameter value \( \nu_1 \) to form a smooth continuum of zeros \( C_0 \cup C_1 \). In other words, can we prove that \( x^0(\nu_1) = x^1(\nu_1) \) and that the connection is smooth? It turns out that validated continuation is well suited to answer this question in a nice fashion. At the parameter value \( \nu_1 \), we have two sets enclosing a unique zero namely

\[
B_0 \overset{\text{def}}{=} \bar{x}_0 + (\nu_1 - \nu_0)\bar{x}_0 + B(r_0),
\]

and

\[
B_1 \overset{\text{def}}{=} \tilde{x}_1 + B(r_1).
\]

We want to prove that the solutions in \( B_0 \) and \( B_1 \) are the same. We return now to the radii polynomials \( p_k(r, |\Delta_\nu|), k = k_0, \ldots, M \) constructed at basepoint \((\nu, x) = (\nu_1, \tilde{x}_1)\), and evaluate them at \( \Delta_\nu = 0 \):

\[
\tilde{p}_k(r) = p_k(r, 0).
\]

Since \( \tilde{p}_k(r_1) < 0 \), we find a non empty interval \( I \overset{\text{def}}{=} [r_1^-, r_1^+] \) containing \( r_1 \) such that \( \tilde{p}_k(r) \) are all strictly negative on \( I \). We now have two additional sets enclosing a unique zero at parameter value \( \nu_1 \), namely

\[
B_1^\pm \overset{\text{def}}{=} \tilde{x}_1 + B(r_1^\pm).
\]

Proposition 8. If \( B_0 \subset B_1^+ \) or \( B_1^- \subset B_0 \), then \( C_0 \cup C_1 \) consists of a continuous branch of solutions of \( f(\nu, x) = 0 \), and \( C_0 \cap C_1 = \{ (\nu_1, x^0(\nu_1)) \} = \{ (\nu_1, x^1(\nu_1)) \} \in B_0 \cap B_1 \). Moreover, if \( T(\nu, x) = T_\nu(x) \) is of class \( C^l \), then \( C_0 \cup C_1 \) is a \( C^l \) smooth curve.

Proof. For a geometric representation of the proof, we refer to Figure 1. The sets \( B_1^- \), \( B_1^+ \) and \( B_1 \) all contain a unique zero of \( f(\nu_1, \cdot) \). Since the balls are nested, these zeros are one and the same, namely \( x^1(\nu_1) \). Furthermore, \( B_0 \) also contains exactly one zero of \( f(\nu_1, \cdot) \), namely \( x^0(\nu_1) \). The assertion implies that either \( B_0 \) and \( B_1^+ \), or \( B_0 \) and \( B_1^- \) are nested, hence \( x^0(\nu_1) = x^1(\nu_1) \). This means that \( C_0 \cup C_1 \) consists of a one dimensional continuous branch of zeros of \( f \). It remains to prove smoothness at \( \nu = \nu_1 \). By Lemma 7, \( x^1(\nu) \) is a smooth \( C^l \) function on the interval \([\nu_1 - \Delta_\nu, \nu_1 + \Delta_\nu] \). Moreover, we assert
Figure 1: $B_0 \cap B_1$ contains a unique zero of (12) and $C_0 \cup C_1$ consists of a continuum of zeros. This picture illustrates the proof of Proposition 8 that $x^0(\nu)$ and $x^1(\nu)$ coincide on $[\nu_1 - \epsilon, \nu_1]$ for $\epsilon > 0$ sufficiently small. Namely, $x^1(\nu)$ lies in the interior of the tube $\{(\nu, x), |\nu - \nu_1| \leq \Delta_x, x - (\bar{x}_1 + (\nu - \nu_1)\dot{x}_1) \in B(r_1)\}$, and $(\nu, x^1(\nu))$ are the only zeros of $f$ inside this tube. On the other hand, the solution curve $x^0(\nu)$ must enter the tube for $\nu$ close to $\nu_1$, since $x^0(\nu_1)$ is in the interior. From uniqueness of solutions inside the tube (Lemma 7) it follows that indeed $x^0(\nu)$ and $x^1(\nu)$ coincide on $[\nu_1 - \epsilon, \nu_1]$ for $\epsilon > 0$ sufficiently small. Hence, we conclude that the union $C_0 \cup C_1$ is $C^1$ smooth.

In practice, the hypotheses of Proposition 8 are verified as follows. The center points $\bar{x}_0 + (\nu_1 - \nu_0)\dot{x}_0$ of $B_0$ and $\bar{x}_1$ of $B_1, B_1^\pm$ are computed using the finite dimensional approximations $f^{(m_0)}$ and $f^{(m_1)}$ of $f$, respectively. This means that $\bar{x}_0 + (\nu_1 - \nu_0)\dot{x}_0 \in \mathbb{R}^{m_0}$ and $\bar{x}_1 \in \mathbb{R}^{m_1}$. Let $\bar{m} = \max\{m_0, m_1\}$. Recalling (18), let $q_0$ and $q_1$ be the decay rates of the tails of $B_0$ and $B_1$, respectively. Note that $B_1$, $B_1^-$ and $B_1^+$ have the same decay rate. If $q_0 < q_1$, the tail of $B_1^+$ decays faster than the tail of $B_0$, which clearly means that $B_0 \not\subset B_1^+$. Hence, we have to check whether or not $B_1^- \subset B_0$ by verifying that the product of the first $\bar{m}$ intervals of $B_1^-$ is a subset of the product of the $\bar{m}$ first intervals of $B_0$ (this consist of checking $2\bar{m}$ inequalities on a computer) and checking that $r^-_1 < r^-_0$. If $q_0 > q_1$, we can only investigate that $B_0 \subset B_1^+$. We proceed as before; that is we verify the inclusion of the $\bar{m}$ dimensional finite part of the sets and then check that $r^-_0 < r^+_1$. If $q_0 = q_1$, we have the choice. For instance, we can start by verifying that $B_0 \subset B_1^+$. If it is true, we stop. If not, we determine whether or not $B_1^- \subset B_0$. If we can show that $B_1^- \subset B_0$, then we have the wanted continuum. If not, we cannot conclude about the continuity of the branch.

3 Pseudo-arclength continuation

In this section, we adapt the continuation method presented in Section 2 to pseudo-arclength continuation. In general, there may be no preferred parameter in which one wants to continue, or if there is, one would like to continue past folds. This is where
pseudo-arclength continuation comes into the picture [8]. The first step is to reformulate the problem so that $D_x f(\nu, x)$ being singular is no longer an obstruction for the method.

### 3.1 Avoiding the singularity of the derivative

Considering $X = (p, \xi)$, where all parameters $p$ are now variables, we want to solve $F(X) = 0$, where $F$ is given by (2), restricted to a plane *almost* perpendicular to the branch of zeros we are following (see [8]). Suppose that we have a predictor $\hat{X}$ and some guess about the direction $\dot{X}$ of the curve, then one can define the plane $(X - \hat{X}) \cdot \dot{X} = 0$. This plane is transverse to the curve and contains the predictor. Appending the equation of the plane to $F$, we consider

\[
F(X) \overset{\text{def}}{=} \left( \frac{(X - \hat{X}) \cdot \dot{X}}{F(X)} \right) = 0.
\]

In this setting, a generic fold point $\bar{X}$ is hyperbolic, that is, $D_\mathbf{X} F(\bar{X})$ is nonsingular. Hence, we can expect a contraction mapping argument to be successful. For a geometric representation, we refer to Figure 2.

### 3.2 Piecewise smooth solution curves

We now incorporate the discussion of Section 3.1 in the context of a predictor-corrector algorithm. From a previous step, we have a direction vector $\dot{X}_0$, and suppose we have computed an approximate solution $\bar{X}_1$ of $F(X) = 0$ in a plane perpendicular to $\dot{X}_0$. We want to construct the radii polynomials based at $\bar{X}_1$. We numerically compute $\dot{X}_1$ such that $D F(\bar{X}_1) \dot{X}_1 \approx 0$. Then, fixing $\Delta_s > 0$ (to be determined later), we define the predictors

\[
\begin{align*}
X_s &= \bar{X}_1 + s \Delta_s \dot{X}_1, \\
X'_s &= \dot{X}_0 + s (\dot{X}_1 - \dot{X}_0),
\end{align*}
\]

where $X \cdot Y$ denotes an inner product (in practice we use the usual dot product in Euclidian space, since $X_s$ and $X'_s$ only have finitely many nonzero components). The

\[
\Pi_s = \{(s, X) \mid E(s, X) \overset{\text{def}}{=} (X - X_s) \cdot X'_s = 0\},
\]

Figure 2: Solving $F(X) = 0$ on the plane $(X - \hat{X}) \cdot \dot{X} = 0$. 

\[
\begin{align*}
&1.9 \quad 1.92 \quad 1.94 \quad 1.96 \quad 1.98 \quad 2 \quad 2.02 \quad 2.04 \\
&0.92 \quad 0.925 \quad 0.93 \quad 0.935 \quad 0.94 \quad 0.945
\end{align*}
\]
family \{\Pi_s \mid s \in [0,1]\} is an interpolation between the plane \Pi_0 from the previous step and the plane \Pi_1 perpendicular to the predictors \mathbf{X}_s, see Figure 3. Note that we can choose \mathbf{X}_1 to be approximately of unit length and such that \mathbf{X}_0 \cdot \mathbf{X}_1 is positive, so that \mathbf{X}_0 and \mathbf{X}_1 point roughly in the same direction (and we do not back trace on the solution curve). When we set \( P = (s,p) \) and \( H = (E,h) \), then we are in the setting of parameter continuation introduced in Section 2, for zeros of \( \mathbf{F}(P,\xi) = (H,g) \), except that the first equation \( E = 0 \) changes at each step in the iterative continuation process (which has some consequences for matching the piecewise continuous solution curves, as discussed in Section 3.3). The set of equations is more conveniently written as

\[
\mathbf{F}(s,X) = \begin{pmatrix}
E(s,X) \\
\mathcal{F}(X)
\end{pmatrix}.
\tag{23}
\]

We point out one difference in notation compared to parameter continuation, namely, a single continuation step is always described by \( s \in [0,1] \), while \( \Delta s \) controls the length of the step (pseudo-arclength). As in parameter continuation, we do not need to fix the step size a priori, allowing us to choose a near optimal \( \Delta s \) at each continuation step.

**Remark 9.** Alternative choices of (21) can be made. For example, here we describe how to obtain a \( C^1 \) representation of the curve. One can compute two nearby approximate solutions \( \mathbf{X}_0 \) and \( \mathbf{X}_1 \) on the solution curve (and thereby thus also fixing the step size), as well as corresponding direction vectors \( \mathbf{X}_0' \) and \( \mathbf{X}_1' \). Then, for \( s \in [0,1] \), we set

\[
\mathbf{X}_s = s^3[\mathbf{X}_1 - \mathbf{X}_0 - 2(\mathbf{X}_1 - \mathbf{X}_0)] + s^2[3(\mathbf{X}_1 - \mathbf{X}_0) - (\mathbf{X}_1 - \mathbf{X}_0)] + s\mathbf{X}_0 + \mathbf{X}_0
\]

and \( \mathbf{X}_s' = \frac{d}{ds}\mathbf{X}_s \). Hence, \( \mathbf{X}_0' = \mathbf{X}_0 \) and \( \mathbf{X}_1' = \mathbf{X}_1 \). We can then look for zeros of (23), with \( E(s,X) \overset{\text{def}}{=} (X - \mathbf{X}_s) \cdot \mathbf{X}_s' \). The advantage of such a choice is the global \( C^1 \) representation of the predictors \( \mathbf{X}_s \), whereas the downside are a significantly larger amount of terms in the estimates, as well as the need to fix a priori the distance between successive points.

We look to enclose uniquely zeros of (23) in sets of the form

\[
B_{\mathbf{X}_s}(r) = X_s + B(r),
\]

where

\[
B(r) = [-r,r]^n \times \prod_{k=0}^{\infty} \left[ -\frac{r}{\omega_k}, \frac{r}{\omega_k} \right]^n.
\]
As before, we set up an equivalent fixed point problem. Suppose that numerically, we found an approximation $A$ of the inverse of $D_X F(0, \dot{X}_1)$. We then define the fixed point problem

$$T_s(X) = X - A F(s, X).$$

(24)

Using the same construction as in Section 2, we construct bounds $Y$ and $Z$, as well as the radii polynomials $p_k(r, \Delta_s)$, $k = k_0, \ldots, M$, uniform in $s \in [0, 1]$, where $k_0 = -l_1$, since we now consider $l_1$ parameters as variables. We use the radii polynomials to find the approximate maximum value $\Delta_s \geq 0$ such that there exists an $r_1 > 0$ satisfying $p_k(r_1, \Delta_s) < 0$, for all $k = k_0, \ldots, M$. Hence, for every $s \in [0, 1]$, the set $B_{\chi_s}(r_1)$ encloses a unique zero $\bar{X}(s)$ of (23). Assuming that $F$ defined in (2) is of class $C^l$, we can conclude that the function $\bar{X}(s)$ is of class $C^l$ (see Lemma 7). We now address the question of the smoothness of the curve

$$C \overset{\text{def}}{=} \left\{ \bar{X}(s) \mid F(s, \bar{X}(s)) = 0, \ s \in [0, 1] \right\}.$$  

Lemma 10. Recall (21) and suppose that $\dot{X}_0, \dot{X}_1 \in \mathbb{R}^{m+1}$. Define

$$\kappa_1 \overset{\text{def}}{=} \sum_{k=k_0}^{-1} |(\dot{X}_1 - \dot{X}_0)_k| + \sum_{k=0}^{m-1} \frac{1}{\omega_k} |(\dot{X}_1 - \dot{X}_0)_k|$$

$$\kappa_2 \overset{\text{def}}{=} \min\{\dot{X}_1 \cdot \dot{X}_0, \dot{X}_1 \cdot \dot{X}_1\},$$

where $\omega_k$ is the decay rate of the set $B(r)$. Let $r_1 > 0$ and $\Delta_s > 0$ such that $p_k(r_1, \Delta_s) < 0$, for all $k = k_0, \ldots, M$. If

$$\kappa_1 r_1 < \Delta_s \kappa_2,$$

then $C$ is a smooth curve.

Proof. We will show that the parametrization $\bar{X}(s)$ is such that $\frac{d\bar{X}}{ds}(s)$ never vanishes, implying that $C$ is a smooth curve. Note that $\kappa_1, \kappa_2 \geq 0$, since $\dot{X}_1$ is chosen so that $\dot{X}_1 \cdot \dot{X}_0 \geq 0$. We prove that $\frac{d\bar{X}}{ds}(s) \neq 0$, for all $s \in [0, 1]$. The definition of $C$ implies that $E(s, \bar{X}(s)) = 0$, for all $s \in [0, 1]$. Hence, for all $s \in [0, 1]$, we get that

$$\frac{\partial E}{\partial s}(s, \bar{X}(s)) + \frac{\partial E}{\partial X}(s, \bar{X}(s)) \frac{d\bar{X}}{ds}(s) = 0.$$  

(26)

Recalling (21) and (22), we show that the first term does not vanish

$$\frac{\partial E}{\partial s}(s, \bar{X}(s)) = -\Delta_s X_1 \cdot X'_s + (X - X_s) \cdot (X_1 - X_0) \neq 0.$$  

Let us estimate the two terms separately. Since $s \in [0, 1]$ and $\Delta_s > 0$,

$$\Delta_s X_1 \cdot X'_s = \Delta_s [X_1 \cdot \dot{X}_0 + s(\dot{X}_1 \cdot \dot{X}_1 - \dot{X}_1 \cdot \dot{X}_0)]$$

$$\geq \Delta_s \min\{\dot{X}_1 \cdot \dot{X}_0, \dot{X}_1 \cdot \dot{X}_1\}$$

$$= \Delta_s \kappa_2.$$  

Since $\bar{X}(s) - X_s \in B(r_1)$,

$$\left| (\bar{X}(s) - X_s) \cdot (\dot{X}_1 - \dot{X}_0) \right| \leq \sum_{k=k_0}^{-1} |(\dot{X}_1 - \dot{X}_0)_k| r_1 + \sum_{k=0}^{m-1} \frac{1}{\omega_k} |(\dot{X}_1 - \dot{X}_0)_k| r_1$$

$$= \kappa_1 r_1.$$  

13
Figure 4: Two smooth solution curves $C^0$ and $C^1$ that we want to connect smoothly. Note that the ordering of the planes $\Pi^1_i$ and $\Pi^0_i$ may be different, but this does not influence the arguments.

It follows that $\frac{\partial E}{\partial s}(s, \tilde{X}^1(s)) \leq -\Delta_s \kappa_2 + \kappa_1 r_1 < 0$, for all $s \in [0, 1]$. We conclude from (26) that $\frac{d\tilde{X}^1}{ds}(s) \neq 0$, for all $s \in [0, 1]$. By the implicit function theorem, $C$ is a smooth curve.

In practice, we verify condition (25) at the end of the continuation step, that is when we have found an $r_1 > 0$ and the approximately maximal $\Delta_s$ such that $p_k(r_1, \Delta_s) < 0$ for all $k = k_0, \ldots, M$. We compute $\kappa_1$ and $\kappa_2$ and then check that $\kappa_1 r_1 - \Delta_s \kappa_2 < 0$.

### 3.3 Matching the piecewise smooth solution curves

In Section 3.2, we introduced the theory to compute smooth pieces of solution curves. In this section, we show how to *glue* these pieces to form a global smooth solution curve. Suppose that we have performed two successful pseudo-arclength continuation steps and obtained two smooth pieces of solution curves $C^0 \triangleq \{\tilde{X}^0(s), s \in [0, 1]\}$ and $C^1 \triangleq \{\tilde{X}^1(s), s \in [0, 1]\}$ of $F(X) = 0$, with $C^i$ originating in $\Pi^0_i$ and ending in $\Pi^1_i$ ($i = 0, 1$), see Figure 4. Consider the sets

$$B_0 \triangleq B(r_0) + \tilde{X}_1 \quad \text{and} \quad B_1 \triangleq B(r_1) + \tilde{X}_1,$$

(27)
each enclosing a unique zero of $F$ on $\Pi^0_1$ and on $\Pi^1_0$ respectively. Note that there might be a small distance between the planes $\Pi^0_1 : (X - \tilde{X}_1) \cdot \dot{X}_0 = 0$ and $\Pi^1_0 : (X - \tilde{X}_1) \cdot \dot{X}_0 = 0$. Indeed, $\tilde{X}_1$ was numerically computed such that

$$\delta \triangleq ||(\tilde{X}_1 - \tilde{X}_2) \cdot \dot{X}_0|| \approx 0,$$

(28)
but exact equality cannot be guaranteed. We remark that if $\dot{X}_0$ is computed so that $||\dot{X}_0|| \approx 1$, then $\delta$ is a very good approximation of the distance between the parallel planes $\Pi^0_1$ and $\Pi^1_0$ (see Figure 4). We need to fill the gap between the planes. Consider
\[ \overline{\Pi}_\tau : \overline{E}(\tau, X) \overset{\text{def}}{=} (X - \bar{X}_1) \cdot \dot{X}_0 + \tau, \quad \tau \in [-\delta, \delta], \]

the interpolation with parallel planes between \( \Pi^0_0 \) for \( \tau = 0 \) and \( \Pi^0_1 \) for \( \tau = \pm \delta \) (depending on the sign of \((\bar{X}_1 - \bar{X}_1) \cdot \dot{X}_0\)). As in Section 3.2, we would like to find uniform \( r^+_1 > r^-_1 > 0 \) such that \( B_{\bar{X}_1}(r^-_1) \) and \( B_{\bar{X}_1}(r^+_1) \) both contain, for all \( \tau \in [-\delta, \delta] \), a unique zero of

\[ \overline{\mathbf{F}}(\tau, X) \overset{\text{def}}{=} \begin{pmatrix} \overline{E}(\tau, X) \\ \mathcal{F}(X) \end{pmatrix}. \tag{29} \]

Let \( A \) be the operator used in the construction of the radii polynomials based at \( \bar{X}_1 \). In other words, \( A \) was used to define the uniform contraction \( T \) that yielded the existence of \( \mathcal{C}^1 \). Define \( \overline{T}_\tau(X) = X - A\overline{\mathbf{F}}(\tau, X) \) and consider the uniform predictor \( \bar{X}_1 \) for all \( \tau \in [-\delta, \delta] \). For every \( \tau \in [-\delta, \delta] \), we want to enclose a unique fixed point of \( \overline{T}_\tau \) in \( B_{\bar{X}_1}(r) \), for some \( r > 0 \). Consider the radii polynomials \( \hat{p}_k(r, |\tau|) \), \( k = k_0, \ldots, M \), associated to this problem. Recalling (19) and (20), we note that the bound \( Z \) does not depend on \( \tau \), while the bound \( Y \) depends on \( |\tau| \) linearly, see equations (31) and (32) below. Note that \( \hat{p}_k(r, |\tau|) \leq \hat{p}_k(r, \delta) \), for all \( k \) and for all \( \tau \in [-\delta, \delta] \). Suppose that there exist \( r^+_1 > r^-_1 > 0 \) such that \( \hat{p}_k(r^+_1, \delta) < 0 \) and \( \hat{p}_k(r^-_1, \delta) < 0 \), for all \( k = k_0, \ldots, M \). Hence, for any given \( \tau \in [-\delta, \delta] \), the sets

\[ B^+_1 \overset{\text{def}}{=} B_{\bar{X}_1}(r^+_1) = \bar{X}_1 + B(r^+_1) \tag{30} \]

contain a unique zero of (29). By Lemma 7, we get the existence of

\[ \mathcal{C}^{0,1} \overset{\text{def}}{=} \left\{ \bar{X}^{0,1}(\tau) \mid \overline{\mathbf{F}}(\tau, \bar{X}^{0,1}(\tau)) = 0, \quad \tau \in [-\delta, \delta] \right\}, \]

where \( \bar{X}^{0,1}(\tau) \) is a smooth function. In the context of pseudo-arclength continuation, the following result is the analogue of Proposition 8.

**Proposition 11.** Suppose that \( \mathcal{C}^0 \) and \( \mathcal{C}^1 \) are smooth curves. If \( B_0 \subset B^+_1 \) or \( B^-_1 \subset B_0 \), then \( \mathcal{C}^{0} \cup \mathcal{C}^{0,1} \cup \mathcal{C}^1 \) consists of a smooth solution curve of \( \mathcal{F}(X) = 0 \).

**Proof.** We show that \( \mathcal{C}^0 \) and \( \mathcal{C}^1 \) connect smoothly via \( \mathcal{C}^{0,1} \). First note that \( B^+_1 \) and \( B_1 \) all uniquely enclose a zero of \( \mathcal{F} \) in the plane \( \Pi^0_0 \). Since these balls are nested, these zeros are the same, namely \( \bar{X}^{0,1}(0) = \bar{X}^1(0) \). Note also that \( B^+_1 \) and \( B_0 \) all uniquely enclose a zero of \( \mathcal{F} \) in the plane \( \Pi^0_1 \). By hypothesis, \( B_0 \) and \( B^+_1 \) are nested or \( B^-_1 \) and \( B_0 \) are nested, implying that \( \bar{X}^{0,1}(\pm \delta) = \bar{X}^0(1) \). This settles continuity of \( \mathcal{C}^0 \cup \mathcal{C}^{0,1} \cup \mathcal{C}^1 \). Smoothness of \( \mathcal{C}^{0,1} \) follows immediately (as in the proof of Lemma 10), since \( \frac{\partial \hat{p}}{\partial \tau} = 1 \). Furthermore, combining the continuity of the polynomials \( \hat{p}_k \) and the fact that \( \hat{p}_k(r^+_1, \delta) < 0 \), we infer the existence of an \( \varepsilon > 0 \) such that \( \hat{p}_k(r^-_1, \delta + \varepsilon) < 0 \). The smooth solution curve \( \{ (\tau, \bar{X}^{0,1}(\tau)), |\tau| \leq \delta + \varepsilon \} \), which slightly elongates \( \mathcal{C}^{0,1} \), overlaps (at the ends) with \( \mathcal{C}^0 \) and \( \mathcal{C}^1 \), by arguments analogous to those used in the proof of Proposition 8. Hence, we conclude that \( \mathcal{C}^0 \) and \( \mathcal{C}^1 \) connect smoothly via \( \mathcal{C}^{0,1} \). \( \square \)

In practice, the construction of the radii polynomials \( \hat{p}_k(r, |\tau|) \) is very little extra work. Indeed, consider the radii polynomials \( p_k(r, \Delta_s) \), \( k = k_0, \ldots, M \), based at \( \bar{X}_1 \), which were used to conclude about the existence of \( \mathcal{C}^1 \). Let \( Y_k(\Delta_s) \) and \( Z_k(r, \Delta_s) \) be the bounds used in the construction of \( p_k(r, \Delta_s) \). Recalling the definition of (29), we first realize that

\[ \sup_{b, c \in B(r)} \left| D\overline{T}_\tau(\bar{X}_1 + b)c \right| \leq Z_k(r, 0), \tag{31} \]
for all $k=k_0,\ldots,M$. This is due to the fact that $D\tilde{F}(\tau,X) = DF(0,X)$. Furthermore, using the triangle inequality, we get that
\[
\left|\left[\tilde{T}_\tau(\bar{X}_1) - \bar{X}_1\right]_k\right| = \left|\left[-A\tilde{F}(\tau,\bar{X}_1)\right]_k\right|
\]
\[
= \left|\left[-A \left(\begin{array}{c} -\tau \\ \mathcal{F}(\bar{X}_1) \end{array}\right) \right]_k\right|
\]
\[
\leq \left|\left[A \left(\begin{array}{c} \tau \\ 0 \end{array}\right) \right]_k\right| + \left|\left[-A \left(\begin{array}{c} 0 \\ \mathcal{F}(\bar{X}_1) \end{array}\right) \right]_k\right|
\]
\[
\leq |\tau||A_{1,k}| + Y_k(0),
\]
by the definition of $Y_k$. Combining (31) and (32), we conclude that
\[
\tilde{p}_k(r,|\tau|) = p_k(r,0) + |\tau||A_{1,k}|.
\]
Thus, the difference between the construction of $\tilde{p}_k$ and $p_k$ is given in (33).

4 Applications

In this section, we introduce two applications of the method, where we compute global smooth solution curves of differential equations. The first application, in the context of delay equations, uses the parameter continuation method of Section 2 and the second one, in the context of ordinary differential equations, uses the pseudo-arclength method of Section 3.

4.1 Periodic solutions of delay equations

In [9], the parameter continuation method introduced in Section 2 is applied to the so-called Wright’s equation
\[
y'(t) = -\alpha y(t-1)[1+y(t)], \quad \alpha > \frac{\pi}{2}.
\]
(34)
The continuation argument is used to compute a continuous branch $\mathcal{F}_0$ of slowly oscillating periodic solutions (SOPS) of (34) and to show (rigorously) that $\mathcal{F}_0$ does not have any fold points on the parameter interval $[\pi/2 + \varepsilon, 2.24]$, where $\varepsilon = 7.3165 \times 10^{-4}$. This result is an attempt to partially answer the conjecture that equation (34) has a unique SOPS for every $\alpha > \pi/2$. A representation of the rigorously computed branch of SOPS is shown Figure 5. The details of the construction of the radii polynomials and the main results of this problem can be found in [9].

4.2 Forcing theorem and periodic solutions of ordinary differential equations

As was mentioned in Example 1, we are interested in computing periodic solutions of the Swift-Hohenberg equation $-u''' - \nu u'' + u - u^3 = 0$ with a special symmetry at the zero energy level $E = 0$. In [1], a rigorous continuation argument in the parameter $\nu$ is used to prove the following result.

Proposition 12. For every $\nu \in [0,2]$, the dynamics of the Swift-Hohenberg ODE (6) on the energy level $E = 0$ is chaotic in the sense that there exists a two-dimensional Poincaré return map which has a compact invariant set on which the topological entropy is positive.
The reason why the continuation is stopped at \( \nu = 2 \) is the apparent existence of a saddle-node bifurcation (a fold) at \( \nu \approx 2.03165 \).

In what follows, we extend Proposition 12 by using the rigorous pseudo-arclength continuation introduced in Section 3 to continue through the fold. Define \( X = (\nu, L, \xi_0, \xi_1, \xi_2, \ldots) \) and

\[
h(X) \overset{\text{def}}{=} -2L^2 \sum_{l=1}^{\infty} l^2 \xi_l - \frac{1}{\sqrt{2}} \left[ \xi_0 + 2 \sum_{l=1}^{\infty} \xi_l \right]^2 + \frac{1}{2^{\sqrt{2}}},
\]

and for all \( k \geq 0 \),

\[
g_k(X) \overset{\text{def}}{=} [1 + \nu L^2 k^2 - L^4 k^4] \xi_k - \sum_{k_1 + k_2 + k_3 = k} \xi_{k_1} \xi_{k_2} \xi_{k_3},
\]

where \( \xi_{-k} \overset{\text{def}}{=} \xi_k \). Define \( F = (h, g_0, g_1, g_2, \ldots)^T \). Let us describe the algorithm, where we focus on the differences with the parameter continuation in [1] (in particular Procedure 16 and the bounds in Sections 3.2 and 3.3 in [1]).

From a previous step, assume that we computed a smooth solution curve \( C_0 \) and a direction vector \( \dot{X}_0 \). Here are the steps to fulfill in order to prove existence of (and compute) another piece of smooth solution curve \( C_1 \) and to glue it smoothly to \( C_0 \):

1. Using a finite dimensional approximation \( F^{(m)} : \mathbb{R}^{m+2} \rightarrow \mathbb{R}^{m+2} \), we compute an approximate zero \( \bar{X}_1 \) of \( F \) on a plane perpendicular to \( \dot{X}_0 \). We also compute a new direction vector \( \dot{X}_1 \) such that \( DF(\bar{X}_1) \dot{X}_1 \approx 0 \). Knowing \( \bar{X}_0, \bar{X}_1 \) and \( \dot{X}_1 \), we build the predictors defined in (21), the family of planes \( \{ \Pi_s \mid s \in [0,1] \} \) defined in (22) and the augmented map \( F(s, X) \) defined in (23).

2. We compute the derivative \( D_X F^{(M)}(\bar{X}_1) \), where we choose the computational parameter \( M = 3m - 2 \) (see [1]), and a numerical approximation \( J_M \) of its inverse. We define \( \mu_k(L, \nu) = 1 + \nu L^2 k^2 - L^4 k^4 \), the part of \( g_k \) which is linear in the Fourier
modes $\xi_k$. We define the linear operator $A$ on sequence spaces by

$$A \overset{\text{def}}{=} \begin{bmatrix}
J_M & 0 & 0 & 0 & \ldots \\
0 & \mu_M(\tilde{L}, \tilde{\nu})^{-1} & 0 & 0 & \ldots \\
0 & 0 & \mu_{M+1}(\tilde{L}, \tilde{\nu})^{-1} & 0 & \ldots \\
\vdots & \vdots & \vdots & \ddots & \vdots
\end{bmatrix}.$$  \hspace{1cm} (35)

In order to make sure that $A$ is bijective, we verify using interval arithmetic that $\|J_M D_X F^{(M)}(X_1) - I\|_\infty < 1$ (with $I$ the $3m \times 3m$ identity matrix).

3. We set $T_s(X) = X - AF(X, s)$. We construct the bound $Y$ defined component-wise by (19). Let us mention that since $\nu$ is consider a variable (as opposed to a parameter), $\tilde{\nu}$ and $\tilde{\nu}$ (the first components of $X_1$ and $X_1$, respectively) will appear in $Y$. For a complete description of how to compute the bound $Y(\Delta_s)$, we refer to [1]. Notice also that $F_{-2}(X_s) = 0$. Next we construct $Z$ defined by (20), again including $\nu$ as a variable. Otherwise, the only difference with the construction in [1] is the fact that we need to compute an upper bound $Z_{-2}(r, \Delta_s)$. Without repeating the framework of [1] (in particular we refer the reader to [1] for the precise definition of $A^1$, the approximate inverse of $A$), we note that for $k = -2$, 

$$([DF(s, X_s + b) - A^1]c)_{-2} = DE(s, X_s + b)c - \dot{X}_0^T \cdot c$$

$$= (X_s' - \dot{X}_0)T \cdot c$$

$$= s(\dot{X}_1 - \dot{X}_0)T \cdot c$$

$$= \left(\dot{X}_1 - \dot{X}_0\right)^T \cdot v_F \right) rs.$$ 

Defining $C_{-2}^{1,0} = \left|\left(\dot{X}_1 - \dot{X}_0\right)^T \cdot v_F\right|$, we get that for every $b, c \in B(r)$ and $s \in [0, 1]$,

$$\left|([DF(X_s + b, s) - A^1]c)_{-2}\right| \leq C_{-2}^{1,0} r.$$ 

Incorporating $C_{-2}^{1,0}$ in Table 3 in [1], we have all the ingredients to build the $Z(r, \Delta_s)$. Note that Table 3 in [1] contains the coefficients of the polynomials $Z_k(r, \Delta_s)$ defined by (20). We construct the radii polynomials $p_k(r, \Delta_s)$, $k = -2, \ldots, M$ defined in Definition 6. We compute $r_1 > 0$ and an approximately maximal $\Delta_s > 0$ (if they exist and are computable) such that $p_k(r, \Delta_s) < 0$. Recalling Lemma 10, we construct $\kappa_1$ and $\kappa_2$ and verify inequality (25). If the inequality is satisfied, we combine Lemma 7 and Lemma 10 to conclude the existence of the new piece of smooth solution curve $C^1$.

4. We compute $\delta$ defined in (28), recall (33) and construct the radii polynomials $\tilde{p}_k(r_1, |r|)$. If we can show the existence of $r_1^+ > r_1^- > 0$ such that $\tilde{p}(r_1^+, \delta), \tilde{p}(r_1^-, \delta) < 0$, we construct the sets $B_0$, $B_1$ and $B^\pm_1$ defined in (27) and (30). If we can show that the hypothesis of Proposition 11 is satisfied, that is, if we can show that $B_0 \subset B_1^+$ or $B_1^- \subset B_0$, then we conclude that $C^0$ and $C^1$ connect smoothly via $C^{0,1}$.

We have successfully iterated the above steps for the Swift-Hohenberg problem. This proves the existence of a global smooth branch of periodic solutions of (6) at the energy level $E = 0$, see Figure 6 (the additional geometric property needed in [1] is also satisfied). We thus obtain the following Corollary, generalizing Proposition 12.
Corollary 13. Let $\nu_* = 2.0316$. For every parameter value $\nu \in [0, \nu_*]$, the Swift-Hohenberg equation (6) is chaotic at the energy level $\mathcal{E} = 0$.

Using the rigorous pseudo-arclength continuation, we also obtained that the branch of periodic solutions we followed has a fold for a parameter value $\nu \in [2.031647, 2.031657]$.

References


Figure 6: A smooth branch of periodic solutions of (6) at the energy level $\mathcal{E} = 0$. 
Published for the Tata Institute of Fundamental Research, Bombay, 1987. With
notes by A. K. Nandakumaran and Mythily Ramaswamy.

[9] J.-P. Lessard. Recent advances about the uniqueness of the slowly oscillating periodic
solutions of Wright’s equation. Preprint.

Schriften zur Mathematik, Physik und Informatik [Augsburger Publications of Math-
a preface by Stanislaus Maier-Paape.

local uniqueness of a double turning point for a radially symmetric solution of the

equations using residual iterations with higher order finite elements, J. Comput.

311–323.

Special issue on linear algebra in self-validating methods. Linear Algebra Appl., 324

[15] M. Plum. Existence and enclosure results for continua of solutions of parameter-


problems with local uniqueness by Banach’s fixed-point theorem. SIAM J. Numer.

[18] P. Zgliczyński and K. Mischaikow. Rigorous numerics for partial differential equa-