

About the asymptotic numerical method

Ph. CAUSSIGNAC, IMB/FSB-EPFL

October 2003

1 Introduction

In this note, we want to recall some features of the Asymptotic Numerical Method (ANM), report some numerical experiments we made and present some points which, in our opinion, need to be investigated.

The ANM method was designed for solving PDE problems with one parameter in structure or in fluid mechanics [8], [7]. After discretization of these problems, usually with finite elements, one has to deal with a nonlinear algebraic system

$$\mathbf{F}(\mathbf{x}) = \mathbf{0}, \quad (1)$$

where $\mathbf{F} : \mathbf{x} = (\mathbf{u}, \lambda) \in \mathbb{R}^{n+1} \mapsto \mathbf{F}(\mathbf{x}) \in \mathbb{R}^n$ is defined by

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}_0 + \mathbf{L}\mathbf{x} + \mathbf{Q}(\mathbf{x}, \mathbf{x}). \quad (2)$$

The operator \mathbf{L} is linear, \mathbf{Q} is quadratic and λ is a scalar parameter. In fact, any problem which can be brought in this form can be solved with the ANM. A solution of such a system is a curve or path or branch $\mathbf{u}(\lambda)$ in \mathbb{R}^n .

The principle of the method is based on analytic perturbation. Starting from a known point $\mathbf{u}_0(\lambda_0)$, one looks for a solution (still denoted $\mathbf{u}(\lambda)$ for simplicity) of the form

$$\mathbf{u}(\lambda) = \sum_{i=0}^N \mathbf{u}_i \lambda^i \quad (3)$$

and, by comparison of the powers of λ in (1), we obtain a sequence of N linear systems for the coefficients $\mathbf{u}_i \in \mathbb{R}^n$.

However, if the path $\mathbf{u}(\lambda)$ has singularities (folds or bifurcations), one has to introduce another parametrization:

$$\mathbf{x}(s) = \sum_{i=0}^N \mathbf{x}_i s^i, \quad (4)$$

where s is some sort of arclength-like parameter; for example, for the well-known pseudo-arclength [3], we have:

$$s = \dot{\mathbf{x}}_0^t (\mathbf{x} - \mathbf{x}_0), \quad (5)$$

where $(\dot{\mathbf{x}}_0, \dot{\lambda}_0)$ is a unit tangent vector to the curve $(\mathbf{x}(s), \lambda(s))$ at $(\mathbf{x}_0, \lambda_0)$. In place of (1), we will have to solve the augmented system

$$\mathbf{G}(\mathbf{x}, s) = \begin{pmatrix} \mathbf{F}(\mathbf{x}) \\ \dot{\mathbf{x}}_0^t (\mathbf{x} - \mathbf{x}_0) - s \end{pmatrix} = \mathbf{0}, \quad (6)$$

and now, the linear systems to solve read

$$J(\mathbf{x}_0)\mathbf{x}_i = \begin{pmatrix} (\delta_{1i} - 1) \sum_{k=1}^{i-1} \mathbf{Q}(\mathbf{x}_k, \mathbf{x}_{i-k}) \\ \delta_{1i} \end{pmatrix}, \quad (7)$$

with

$$J(\mathbf{x}_0) = \begin{pmatrix} \mathbf{F}_{\mathbf{x}}(\mathbf{x}_0) \\ \dot{\mathbf{x}}_0^t \end{pmatrix}. \quad (8)$$

Of course the Jacobian of \mathbf{F} is given by $\mathbf{F}_{\mathbf{x}}(\mathbf{x}_0) = \mathbf{L} + \mathbf{Q}(\mathbf{x}_0, \cdot) + \mathbf{Q}(\cdot, \mathbf{x}_0)$.

The interesting feature of the method is that the expansion of \mathbf{x} is valid in a whole interval which does not need to be small [5], contrarily to the standard pseudo-arclength method [3]. Thus, proceeding

by successive intervals, we get a (continuous) piecewise polynomial approximation of a solution of the problem. The usual criterion for determining the interval length requires that the first nonzero term of the residual is smaller than a given (small) number; one observes that, while using this criterion, very small steps could be needed, especially close to a bifurcation. But, this does not allow to locate precisely the bifurcation point.

2 Steplength and convergence radius

The expansion (4) is a truncation of the full Taylor serie

$$\mathbf{y}(s) = \sum_{i=0}^{\infty} \mathbf{y}_i s^i, \quad (9)$$

which, in general has a finite radius of convergence, say s_1 . Hence, the interval of validity of this expansion, i.e. the steplength $s_{\max} > 0$, should be determined by:

1. The truncation error, which requires an estimate of the quantity $\sum_{i=N+1}^{\infty} \mathbf{y}_i s^i$,
2. The domain of validity of (9), that is the value of s_1 .

In the standard ANM method, one uses basically two criteria. The first and mostly used one is based on the residual $\mathbf{F}(\mathbf{x}(s))$. More precisely, one assumes that the term of order $N + 1$ is dominant and requiring its absolute value to be smaller than a positive number ϵ :

$$s^{N+1} \left| \sum_{i=1}^N \mathbf{Q}(\mathbf{x}_i, \mathbf{x}_{N+1-i}) \right| \leq \epsilon,$$

we get

$$s_{\max} = \sqrt[N+1]{\frac{\epsilon}{\sum_{i=1}^N \mathbf{Q}(\mathbf{x}_i, \mathbf{x}_{N+1-i})}}. \quad (10)$$

Intuitively, we believe that this criterion has more to do with truncature than with the convergence radius; however, in several cases, using it allows to locate a bifurcation by observing an accumulation of steps of short length [1].

The other criterion simply says that the ratio of the last to the first term of the sum (4) should not be greater than ϵ and yields

$$s_{\max} = \sqrt[N-1]{\epsilon \frac{|\mathbf{x}_1|}{|\mathbf{x}_N}}. \quad (11)$$

Recall that the convergence radius ρ of a power serie

$$\sum_{i=0}^{\infty} a_i s^i$$

is either infinite or can be determined by the square root criterion

$$\rho = \frac{1}{\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}};$$

one can then intuitively understand why the roots in both criteria above are important. In fact, very often the coefficient themselves grow to very large values when N increases.

Of course, the serie (9) is not a standard one; but one can expect the same kind of behavior: either convergence for any s or existence of a finite convergence radius. In the latter case, divergence is due to a singularity on the curve $\mathbf{y}(s)$ which can also be present on $\mathbf{x}(s)$. Here, we will restrict ourselves to one kind of singularity, *the simple bifurcation point* and neglect folds or other bifurcations.

Definition 1 [3] *A point \mathbf{x}_1 on a branch $\mathbf{x}(s)$ is called simple bifurcation point of (1) if $\text{rank}(\mathbf{F}_{\mathbf{x}}(\mathbf{x}_1)) = n - 1$.*

Assuming a technical assumption which can be found in [3], the following theorem holds.

Theorem 1 *If \mathbf{x}_1 is a simple bifurcation point and if 0 is an algebraically simple eigenvalue of $J(\mathbf{x}_1)$ (8), then the determinant of $J(\mathbf{x}_1)$ changes its sign at \mathbf{x}_1 .*

This theorem can be used to detect a simple bifurcation point. In each ANM step, we look to the sign of the determinant in the specific subinterval; then using for example the bisection method or another more efficient method, we can locate more precisely the bifurcation point.

Another way to find bifurcation points is based on the asymptotic behaviour of the serie (9), thanks to the next theorem, quoted in Ref. [5].

Theorem 2 *Assume that \mathbf{x}_1 is the simple bifurcation point closest to $s = 0$. Then, we have, in a neighborhood of s_1 :*

$$\mathbf{y}(s) \approx \alpha(s)(s - s_1)^\nu \quad (12)$$

where $0 < \nu < 1$ is a rational number and α is analytic at 0 and s_1 .

Remark 1 *We have often $\nu = 1/2$ and in general in most cases $\nu = 1/k$, k a positive integer ≥ 2 .*

In fact, we do not compute $\mathbf{y}(s)$, but an approximation of it, namely $\mathbf{x}(s)$ given by (4). The problem is to compute the asymptotic behavior of a function, knowing only a finite expansion approximation of it; this topic has been addressed in Ref.[6] and results in the following theorem, adapted to our particular case.

Theorem 3 *Assume that the function p is analytic at $s = 0$ and has the asymptotic behaviour*

$$p(s) \approx q(s)(s - s_1)^\nu$$

near $s = s_1$, with ν a noninteger real number. Then, if p_k are the coefficients of the Taylor serie of p at $s = 0$, one has:

$$s_1 p_k - \frac{k - \nu - 1}{k} p_{k-1} = O(k^{-\nu-3}). \quad (13)$$

The following trick[6] allows to get an approximation of s_1 in a closed form. We write (13) twice with $k = N, N - 1$, neglect the right-hand-sides for obtaining a linear system with unknowns s_1 and ν ; solving this system yield the approximate value:

$$\bar{s}_1 = \left(N \frac{p_N}{p_{N-1}} - (N - 1) \frac{p_{N-1}}{p_{N-2}} \right)^{-1}. \quad (14)$$

Now, thanks to Teorem 2, each component of \mathbf{x} given by (4) satisfies the hypotheses of Theorem 3. One can choose for example the last one (λ) for applying formula (14). In an ANM computation, we can obtain in this way an approximation of the closest bifurcation in each subinterval; of course, we will have problems when two bifurcation are quite close to the beginning of the subinterval. In this case, the steplength has to be adapted to the location of the bifurcations; this is a topic that we have not yet studied.

2.1 Trivial branches

Let us define a trivial branch of (1) as a path for which

$$\mathbf{u}(\lambda) = \mathbf{u}_0, \quad (15)$$

that is the sum (4) contains only the first term. Of course, in this particular case, both criteria (10) and (11) will give the correct answer $s_{\max} = \infty$; but, this is numerically untractable and furthermore the standard ANM does not allow the detection of bifurcation on this trivial branch.

Due to the definition (2) of \mathbf{F} , the Jacobian of \mathbf{G} on a trivial branch takes the special form:

$$J(\lambda) = \begin{pmatrix} A(\mathbf{u}_0) - \lambda B(\mathbf{u}_0) \\ \mathbf{0} & 1 \end{pmatrix} \quad (16)$$

where the $n \times n + 1$ matrices A and B are defined by

$$A(\mathbf{u}_0) = \mathbf{F}_{\mathbf{x}}(\mathbf{u}_0, 0), \quad B(\mathbf{u}_0) = A(\mathbf{u}_0) - \mathbf{F}_{\mathbf{x}}(\mathbf{u}_0, 1). \quad (17)$$

Thus, a simple bifurcation point corresponds, thanks to Theorem 1, to an eigenvalue of the generalized problem in \mathbb{R}^{n+1} :

$$\mathcal{A}\mathbf{v} = \lambda\mathcal{B}\mathbf{v}, \quad (18)$$

with

$$\mathcal{A} = \begin{pmatrix} A(\mathbf{u}_0) & \\ \mathbf{0} & 1 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} B(\mathbf{u}_0) & \\ \mathbf{0} & 0 \end{pmatrix}.$$

We assume that the eigenvalues are ordered in the following way:

$$0 < |\lambda_0| < |\lambda_1| < \dots < |\lambda_{k-1}| < |\lambda_k| < \dots. \quad (19)$$

If we are looking for the smallest eigenvalue (in modulus) λ_0 , it is given by the largest eigenvalue $\mu_0 = 1/\lambda_0$ of the problem

$$\mathcal{A}^{-1}\mathcal{B}\mathbf{v} = \mu\mathbf{v}. \quad (20)$$

Therefore, we can use the power iteration method to compute μ_0 , which consists, given $\widehat{\mathbf{v}}_0 \in \mathbb{R}^{n+1}$ (for example randomly chosen) such that $\|\widehat{\mathbf{v}}_0\| = 1$, in solving the sequence of linear systems

$$\mathcal{A}\mathbf{v}_k = \mathcal{B}\widehat{\mathbf{v}}_{k-1}, \quad \widehat{\mathbf{v}}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}, \quad k = 1, 2, \dots;$$

then $\pm\|\mathbf{v}_k\| \rightarrow \mu_0$ when $k \rightarrow \infty$ and $\widehat{\mathbf{v}}_k$ is converging to a corresponding eigenvector. In our practical computations, we used a number of iterations equal to the order N of the approximation (3), although we always noticed a good convergence already for 10 iterations.

This method allows us to compute only the first bifurcation; for λ_k we need a shift $\bar{\lambda}_k$ such that

$$|\lambda_k - \bar{\lambda}_k| \ll |\lambda_{k-1} - \bar{\lambda}_k|$$

and then solve the problem

$$(\mathcal{A} - \bar{\lambda}_k\mathcal{B})\mathbf{v} = \lambda\mathcal{B}\mathbf{v}$$

with the previous power method to get the shifted eigenvalue $\lambda_k - \bar{\lambda}_k$. Of course, we need some preliminary knowledge about the spectrum (19).

2.2 Numerical experiments

2.2.1 An enzyme model

This example describes the stationary behavior of a chemical reaction with a two-compartment apparatus [3]. The equations for the concentrations u_i in the compartments read

$$\begin{aligned} 2u_1 - u_2 + 100 \frac{u_1}{1 + u_1 + u_1^2} - \lambda &= 0, \\ 2u_2 - u_1 + 100 \frac{u_2}{1 + u_2 + u_2^2} - \lambda &= 0; \end{aligned}$$

the parameter λ represents the concentration in the reservoir external to the compartments.

First of all, the problem has to be brought in the form (2), which of course is not a unique procedure; although all formulation will be equivalent for the "continuous" problem, they can affect the numerics. Following [2], we introduce the new variables

$$\begin{aligned} u_3 &= u_1 + u_1^2, \\ u_4 &= u_2 + u_2^2, \\ u_5 &= \frac{1}{1 + u_3}, \\ u_6 &= \frac{1}{1 + u_4}; \end{aligned}$$

we then get the algebraic system defined by setting

$$\mathbf{F}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ -1 \end{pmatrix}, \mathbf{L} \begin{pmatrix} \mathbf{u} \\ \lambda \end{pmatrix} = \begin{pmatrix} 2u_1 - u_2 - \lambda \\ 2u_2 - u_1 - \lambda \\ u_3 - u_1 \\ u_4 - u_2 \\ u_5 \\ u_6 \end{pmatrix}, \mathbf{Q}(\mathbf{u}, \mathbf{u}) = \begin{pmatrix} 100u_1u_5 \\ 100u_2u_6 \\ -u_1^2 \\ -u_2^2 \\ u_3u_5 \\ u_4u_6 \end{pmatrix}.$$

Starting from the point $(\mathbf{u}, \lambda) = (0, 0, 0, 0, 1, 1, 0)$ we have computed, using the matlab code of [2] the first branch of this problem. The steplength criterion was given by (10) and we made 22 steps. The result (i.e. $u_2(\lambda)$) is presented in Figure 1, where it is compared to a computation with the code AUTO [3].

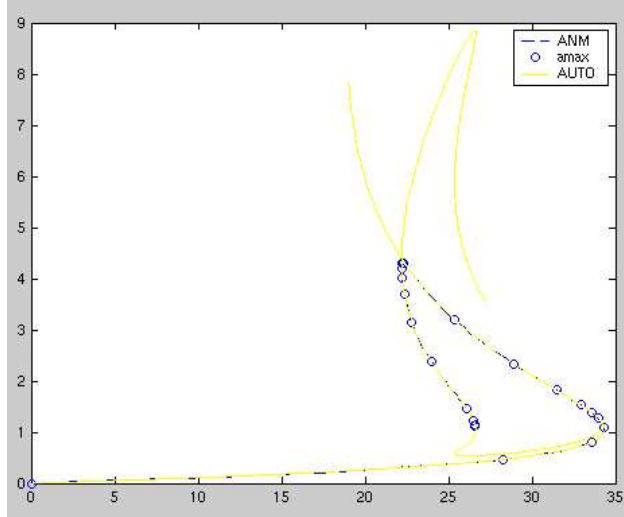


Figure 1: Branch switching

The blue circles indicate the end of the subintervals; notice that the ANM switches to the second branch at the second bifurcation point. For staying onto the first branch, we made a correction after the first bifurcation (after 10 steps) and obtain the right curve, as shown on Figure 2.

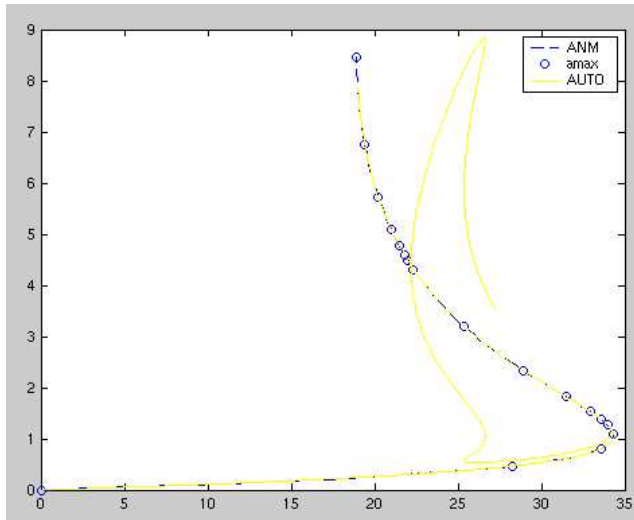


Figure 2: Continuation with correction

In order to compute both branches, we use the strategy of detecting bifurcation with the sign of the determinant of the augmented Jacobian on the first branch. Then, we restart from the first bifurcation point with the tangent orthogonal to the one of the first branch; as explained in [3] the vector tangent

to the orthogonal branch is in the kernel of the Jacobian. The first bifurcation is detected at $\lambda = 34.29$ (34.22 with AUTO), the second one at $\lambda = 22.49$ (22.18 in AUTO). The resulting diagram is shown on Figure 3.

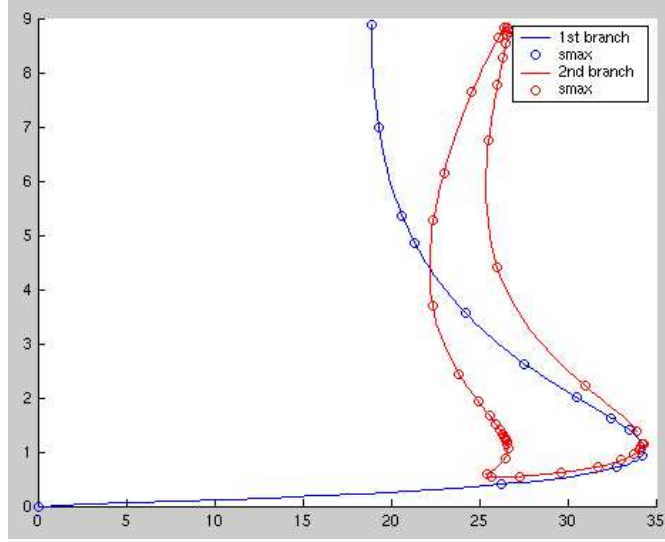


Figure 3: Complete diagram

2.2.2 A planar clamped elastic rod

This problem and its solution are presented in [4]. An elastic rod of length 1 lying on the z -axis, is clamped at the origin; one applies a horizontal force λ and a vertical force $-\mu$ at the end $z = 1$ and also a moment in the y -direction equal to ν . This problem has a trivial branch and has the interesting feature that there is a path connecting the first bifurcation on the first nontrivial branch to the second nontrivial branch.

Using the centerline vector \mathbf{r} and the director \mathbf{d}_3 :

$$\mathbf{r} = \begin{pmatrix} x \\ z \end{pmatrix} \quad \mathbf{d}_3 = \begin{pmatrix} d_{3x} \\ d_{3z} \end{pmatrix},$$

one can write the problem as an ODE system for the unknowns $(\mathbf{r}, \mathbf{d}_3) \in C^1([0, 1])^2 \times C^1([0, 1])^2$:

$$\begin{aligned} \mathbf{r}' &= \mathbf{d}_3, \\ \mathbf{d}_3' &= \begin{pmatrix} (\nu - \lambda x - \mu z)d_{3z} \\ -(\nu - \lambda x - \mu z)d_{3x} \end{pmatrix}, \end{aligned}$$

with the boundary conditions

$$\mathbf{r}(0) = \mathbf{0}, \quad \mathbf{d}_3(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad x(1) = 0, \quad d_{3x}(1) = 0. \quad (21)$$

The first step consists in introducing the unknown functions

$$p = d_{3x}, \quad q = d_{3z}, \quad w = \nu - \lambda x - \mu z,$$

and writing the equations as

$$\begin{aligned} x' &= p, \\ z' &= q, \\ p' &= qw, \\ q' &= -pw, \\ w &= \nu - \lambda x - \mu z. \end{aligned} \quad (22)$$

Next, we have to discretize this system; we are using the Keller-box finite difference scheme. Given the regular partition

$$h = 1/(m - 2), \quad t_i = (i - 3/2)h, \quad i=1,2,\dots,m,$$

we denote by q_i the approximation of a function $q \in C^1((-\frac{h}{2}, 1 + \frac{h}{2}))$ at t_i ; its derivative will be approximated by

$$q'(t_i + h/2) \simeq \frac{q_{i+1} - q_i}{h}, \quad i=1,2,\dots,m-1;$$

if needed, like for example for a boundary condition, we use

$$q(t_i + h/2) \simeq \frac{q_i + q_{i+1}}{2}, \quad i = 1, m - 1.$$

Using this scheme and the vector of unknowns

$$(\mathbf{u}, \lambda)^t = (x_1, x_2, \dots, x_m, z_1, z_2, \dots, z_m, p_1, p_2, \dots, p_m, q_1, q_2, \dots, q_m, w_1, w_2, \dots, w_m, \nu, \mu, \lambda),$$

the system (22) with boundary conditions (21) can be written in the form (1)-(2) with $n = 5m + 2$.

We have computed parts of the trivial branch, of the first and second bifurcated branches and the path linking the first branch to the second one. For the trivial branch, we applied the power method and for the other branches, we used the sign of the determinant of the Jacobian to detect bifurcations and, at occurrence of a change of sign, we used the value predicted by (14) in the current interval for the bifurcation location. The values of λ at the bifurcation points are given below, the ones obtained with AUTO are in parentheses:

- Trivial branch: 39.6 (39.5), 81.3 (80.8)
- First branch: 84.7 (86.2)
- Second branch: -56.8 (-56.2).

The bifurcation diagram is shown in Figure 4 together with the symmetric part computed with AUTO; the agreement is excellent.

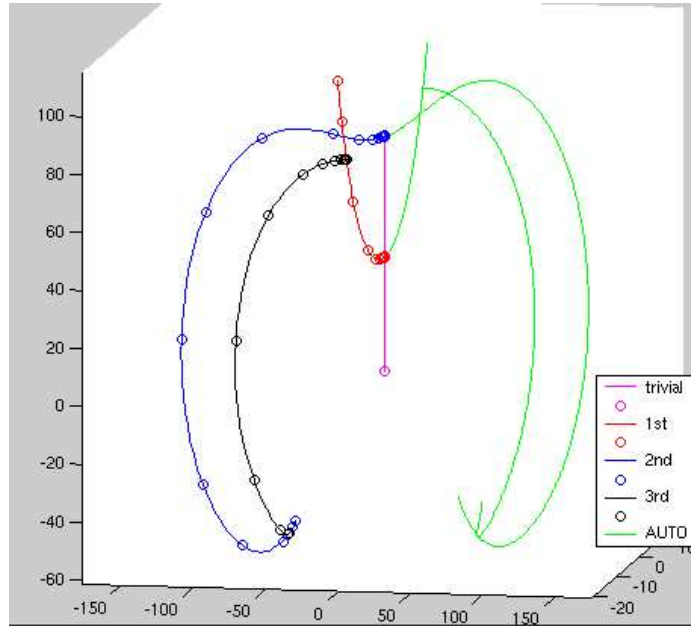


Figure 4: 2D elastic rod diagram

2.3 Conclusion

Our numerical experiments, especially the second one, show that bifurcation detection and switching, in quite complicated diagrams, can be handled with the ANM. The algorithms we used have to be improved; in particular, the step length has to be adapted to the truncation error and to the location of the bifurcation points.

References

- [1] S. Baguet and B. Cochelin, *On the behaviour of the ANM continuation in presence of bifurcations*, Commun. Numer. Meth. Engng **00**, 1-6 (2000).
- [2] B. Cochelin and F. Perignon, *Résolution de petits systèmes algébriques par la MAN sous Matlab*, Report REEF-11/2003 (2003), Université de Marseille.
- [3] E. Doedel, H. B. Keller and J.P. Kernevez, *Numerical analysis and control of bifurcation problems (I) Bifurcation in finite dimensions*, Int. J. Bifurcation and Chaos **1**, 493-520 (1991).
- [4] G. Domokos, *Global description of elastic bars*, ZAMMM **74**, 4, T289-291 (1994).
- [5] J-J. Gervais and H. Sadiky, *A new steplength control for continuation with the asymptotic numerical method*, IMA J. Numer. Analysis **22**, 207-229 (2002).
- [6] C. Hunter and B. Guerrieri, *Deducing the properties of singularities of functions from their Taylor series coefficients*, SIAM J. Appl. Math. **39**, 248-263 (1980).
- [7] H. Lahmam, J.M. Cadou, H. Zahrouni, N.Damil and M. Potier-Ferry, *High-order predictor-corrector algorithms*, J. Numer. Meth. Engng. **55**, 685-704 (2002).
- [8] P. Vanucci, B. Cochelin, N. Damil and M. Potier-Ferry, *An asymptotic-numerical method to compute bifurcating branches*, J. Numer. Meth. Engng. **41**, 1365-1389 (1988).