

Counting the eigenvalues surrounded by a closed curve

Olivier Bertrand * Bernard Philippe *

Presented at the International Conference honouring academician
Sergei K. Godunov

(Novosibirsk, 25-27 September 1999)

1 Introduction

Many scientific applications require localization of eigenvalues of matrices in some predefined part of the complex plane. The most common approach to the problem consists of computing the eigenvalues but the answer may be misleading for two reasons.

First, the iterative procedure which is used for the calculation of the eigenvalues may lose some eigenvalues especially for large matrices since only a part of the spectrum is computed and one can not be sure that all the relevant eigenvalues are computed. This bad behaviour often occurs from a shift-and-invert transformation of a matrix that normally provides the eigenvalues in the order of their increasing distance from the shift. But sometimes, especially when the matrix is non normal, the order of the appearances of the eigenvalues is not exact and wrong conclusions may be drawn.

The second objection to the direct evaluation of the eigenvalues is motivated by the possible lack of precision on the entries of the matrix. In that case, the matrix must be considered as an element of a neighborhood for which diameter is given by the level of precision. Then, the question is that whether is to localize the eigenvalues of all the matrices of the neighborhood. For that purpose, Godunov and Trefethen simultaneously but independently defined the notion of ϵ -spectrum or pseudospectrum (see section 2.1).

In this paper, we propose some procedures for counting all the eigenvalues that belong to a bounded domain of the complex plane ; the boundary of the domain may be defined either by an a priori curve or by the boundary of an ϵ -spectrum which is built step by step by a path following procedure. Other authors proposed different methods for localizing roots of analytic functions

*INRIA/IRISA, Campus de Beaulieu, 35042 RENNES Cedex FRANCE

[8, 10, 11, 15, 16]. This paper is devoted to the special case of the characteristic polynomial of a matrix and is specially concerned with the reliability of the numerical computation.

In Section 2, we introduce the mathematical tools that are necessary to solve our problem; they are the ϵ -spectrum and the Cauchy formula applied to the characteristic polynomial. In Section 3, we describe a method to count the number of eigenvalues in a ϵ -spectrum. The procedure, which is based on a series expansion of the logarithm, is combined with an existing path following method. In Section 4, we present some other procedures based on quadratures for the case of parametrized curves. Roundoff errors occurring in the evaluation of the characteristic polynomial and of its derivatives are examined in Section 5. Finally, in Section 6, we present some numerical results for a common test matrix, and for a matrix obtained from the discretization of a Navier-Stokes equation.

2 Mathematical tools

2.1 The ϵ -pseudospectrum and its determination

The concept of pseudospectrum, or ϵ -pseudospectrum was introduced by Trefethen [14] and Godunov [4].

Let $A \in \mathbb{C}^{n \times n}$, and $\epsilon \geq 0$. The ϵ -pseudospectrum of the matrix A is defined by :

$$\Lambda_\epsilon = \{z \in \mathbb{C} : z \text{ is an eigenvalue of } A + \Delta \text{ where } \|\Delta\|_2 \leq \epsilon\} \quad (1)$$

$$= \{z \in \mathbb{C} : \|(A - zI)^{-1}\|_2 \geq \epsilon^{-1}\} \quad (2)$$

when z is an eigenvalue of A , $\|(A - zI)^{-1}\|_2 = \infty$.

When A is a normal matrix, the set, Λ_ϵ is the union of the discs of radius ϵ centered at every eigenvalue. When A is non-normal, the pseudospectrum may be much bigger.

From (2), we obtain an equivalent definition of Λ_ϵ :

$$\Lambda_\epsilon = \{z \in \mathbb{C} : \sigma_{min}(zI - A) \leq \epsilon\}$$

where $\sigma_{min}(zI - A)$ is the smallest singular value of the matrix $(zI - A)$.

Therefore, the construction of the pseudospectrum of the matrix A corresponds to the determination of the function

$$s : z \longrightarrow \sigma_{min}(zI - A).$$

A common way to compute the pseudospectrum of a matrix A in a given region of the complex plane, is to compute the function s on a grid discretizing the domain. Other approaches described by Brühl [2] and Bekas and Gallopoulos [1] consist of following the level curve of the function s corresponding to a given ϵ .

2.2 Cauchy formula and complex logarithm

To compute the number of eigenvalues surrounded by a given closed curve, we use the following result from the complex analysis [5, 13]:

Theorem 2.1 *Let Γ be a closed piecewise regular Jordan curve (piecewise C^1 and of winding number 1) in the complex plane which does not include eigenvalues of A . (In this paper, Γ is even assumed to be a piecewise- C^∞ curve.)*

The number N_Γ of eigenvalues surrounded by Γ can be expressed by :

$$N_\Gamma = \frac{1}{2i\pi} \int_\Gamma \frac{\frac{d}{dz}f(z)}{f(z)} dz \quad (3)$$

where $f(z) = \det(zI - A)$ is the characteristic polynomial of A .

Let $\gamma(t)_{0 \leq t \leq 1}$ be a parametrization of Γ . Then equation (3) can be transformed into

$$N_\Gamma = \frac{1}{2i\pi} \int_0^1 \frac{\frac{d}{dt}(f \circ \gamma(t))}{f \circ \gamma(t)} dt \quad (4)$$

The primitive φ which is defined by

$$\varphi(u) = \int_0^u \frac{\frac{d}{dt}(f \circ \gamma(t))}{f \circ \gamma(t)} dt \quad u \in [0, 1]$$

is a continuous function which is piecewisely equal to determinations of $\ln(f \circ \gamma)$:

$$\begin{aligned} \forall t \in \mathbb{R} \quad \Re(\varphi(t)) &= \ln |f(\gamma(t))| \\ \Im(\varphi(t)) &\equiv \arg(f(\gamma(t))) \quad (2\pi) \end{aligned}$$

where \Re and \Im denote, respectively, the real and the imaginary parts. The principal argument of z ($0 \leq \arg(z) < 2\pi$) is denoted by $\arg(z)$.

In order to determine N_Γ , only the imaginary part φ_I of φ is needed since $\varphi(1) = 2i\pi N_\Gamma$:

$$N_\Gamma = \frac{1}{2\pi} \varphi_I(1).$$

Lemma 2.1 *Let $0 \leq t_1 < t_2 \leq 1$. Under the condition*

$$|\varphi_I(t_1) - \varphi_I(t_2)| < \pi \quad (5)$$

the knowledge of $\varphi_I(t_1)$ is sufficient to know $\varphi_I(t_2)$.

Proof. There exists $k_1 \in \mathbb{N}$ such that :

$$\varphi_I(t_1) = \arg(f(\gamma(t_1))) + 2k_1\pi$$

Since

$$\varphi_I(t_2) \equiv \arg(f(\gamma(t_2))) \quad (2\pi)$$

and

$$\varphi_I(t_2) \in]\varphi_I(t_1) - \pi, \varphi_I(t_1) + \pi[$$

the real $\varphi_I(t_2)$ is uniquely determined. \diamond

3 Counting the eigenvalues by an expansion of the logarithm

In this section, the objective is to count the number of eigenvalues surrounded by a level curve of the function s when the curve is constructed step by step with a path following method. This method is based on a predictor-corrector scheme. It uses the left and right singular vectors corresponding to the smallest singular value of the matrix $(zI - A)$. The existing implementations of this approach involve a priori given constant stepsize. We look now for a stepsize control which implies

$$|\varphi_{\mathcal{I}}(t + \Delta t) - \varphi_{\mathcal{I}}(t)| < \pi. \quad (6)$$

Let $z = \gamma(t)$ and $\Delta z = \gamma(t + \Delta t) - z$. We assume that neither z nor $z + \Delta z$ are eigenvalues of A . Let $R(z) = (zI - A)^{-1}$. From the following decomposition of the characteristic polynomial :

$$\begin{aligned} f(z + \Delta z) &= \det((z + \Delta z)I - A) \\ &= f(z) \det(I + \Delta z(zI - A)^{-1}) \end{aligned}$$

and therefore we may write

$$\begin{aligned} \varphi(t + \Delta t) - \varphi(t) &\equiv \ln \left(\frac{f(z + \Delta z)}{f(z)} \right) && (2\pi i) \\ &\equiv \ln [\det(I + \Delta z R(z))] && (2\pi i) \\ &\equiv \text{trace} [\ln(I + \Delta z R(z))] && (2\pi i) \end{aligned} \quad (7)$$

Under the condition $|\Delta z| \cdot \|R(z)\|_2 < 1$ which is equivalent to :

$$|\Delta z| < \sigma_{\min}(zI - A) \quad (8)$$

we may consider the series expansion of (7) :

$$\varphi(t + \Delta t) - \varphi(t) \equiv \sum_{j \geq 1} \frac{(-1)^{j+1}}{j} \text{trace}(\Delta z^j R(z)^j) \quad (2\pi i) \quad (9)$$

Actually, we may replace the modulo sign by an equality sign since both sides of the equation are continuous with respect to Δt provided (8) holds. In conclusion, we can claim that

Proposition 3.1 *Assume that the curve which is parametrized by γ does not include any eigenvalue.*

Let $t \in [0, 1]$ and Δt such that

$$|\Delta z| < \sigma_{\min}(zI - A) \quad (10)$$

where $z = \gamma(t)$ and $\Delta z = \gamma(t + \Delta t) - z$, then :

$$\varphi_{\mathcal{I}}(t + \Delta t) - \varphi_{\mathcal{I}}(t) = \sum_{j \geq 1} \frac{(-1)^{j+1}}{j} \Im [\text{trace}(\Delta z^j R(z)^j)]. \quad (11)$$

Since we look for constraints on Δz such that (6) is satisfied, we shall consider the stronger condition :

$$\sum_{j \geq 1} \frac{|\Delta z|^j}{j} |\text{trace}(R(z)^j)| < \pi \quad (12)$$

3.1 Case I: all the singular values are available

When a full SVD decomposition is performed during the continuation method, a bound can easily be obtained. Let $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n = \sigma_{min} > 0$ be the singular values of $(zI - A)$.

A bound is obtained from the following inequality (see [9], p176):

$$\begin{aligned} \forall j \in \mathbb{N} \quad , \quad |\text{trace}(R(z)^j)| &\leq \sum_{k=1}^n |\lambda_k(R(z))|^j \\ &\leq \sum_{k=1}^n (\sigma_k(zI - A))^{-j} \end{aligned} \quad (13)$$

Therefore:

$$\begin{aligned} |\varphi_{\mathcal{I}}(t + \Delta t) - \varphi_{\mathcal{I}}(t)| &\leq \sum_{j \geq 1} \frac{|\Delta z|^j}{j} \sum_{k=1}^n \frac{1}{\sigma_k^j} \\ |\varphi_{\mathcal{I}}(t + \Delta t) - \varphi_{\mathcal{I}}(t)| &\leq \sum_{k=1}^n \sum_{j \geq 1} \frac{1}{j} \frac{|\Delta z|^j}{\sigma_k^j} \\ |\varphi_{\mathcal{I}}(t + \Delta t) - \varphi_{\mathcal{I}}(t)| &\leq - \sum_{k=1}^n \ln \left(1 - \frac{|\Delta z|}{\sigma_k} \right) \end{aligned} \quad (14)$$

which is well defined under condition (8). Let $u(x) = \sum_{k=1}^n \ln \left(1 - \frac{x}{\sigma_k} \right)$ for $x \in]0, \sigma_n[$. Since this is a sum of increasing functions, we can easily compute the zero of $u(x) - \pi$ and thus obtain an upper-bound of $|\Delta z|$ which satisfies the condition of Lemma 2.1 for $\gamma(t_1) = z$ and $\gamma(t_2) = z + \Delta z$.

3.2 Case II: only a few of the smallest singular values are known

Because only the smallest singular value is needed for the continuation method, we can use a method (e.g. Block-Davidson [6]) for computing only the m ($1 \leq m \leq n$) smallest singular values and their associated singular vectors. Therefore, from (13), the following relation holds :

$$\forall j \in \mathbb{N} \quad , \quad |\text{trace}(R(z)^j)| \leq \sum_{k=n-m+1}^n |\sigma_k(zI - A)|^{-j} + \frac{n - m}{\sigma_{n-m+1}(zI - A)^j} \quad (15)$$

It implies :

$$|\varphi_{\mathcal{I}}(t + \Delta t) - \varphi_{\mathcal{I}}(t)| \leq - \sum_{k=n-m+1}^n \ln \left(1 - \frac{|\Delta z|}{\sigma_k} \right) - (n - m) \ln \left(1 - \frac{|\Delta z|}{\sigma_{n-m+1}} \right) \quad (16)$$

From this last relation, an upper-bound of Δz can be obtained as in case I.

3.3 The algorithm

For the path following method, it is necessary to compute the smallest singular value and its associated singular vectors in order to follow the tangent to the level curve of the function $s : z \rightarrow \sigma_{\min}(zI - A)$. Iterative methods that return only a few singular values and vectors are powerful methods to compute those quantities. Then, from the inequality (16) we can estimate a step size control in the following way.

We consider the step k , with $z_k = \gamma(t_k)$. Let $(\sigma_i)_{i=n-m+1, n}$ the m smallest singular values of $(z_k I - A)$ sorted in decreasing order. Let

$$h_{\beta}(\Delta t) = - \sum_{k=n-m+1}^n \ln \left(1 - \frac{|\Delta t|}{\sigma_k} \right) - (n - m) \ln \left(1 - \frac{|\Delta t|}{\sigma_{n-m+1}} \right) - \beta \pi$$

where $\Delta t \in [0, \sigma_n]$, and $\beta \in [0, 1]$.

This is an increasing function of Δt , with $h_{\beta}(0) = -\beta\pi$ and $\lim_{\Delta t \rightarrow \sigma_n} h_{\beta}(\Delta t) = +\infty$. By a dichotomy method, we can estimate the largest $\Delta t_{opt} \in [0, \sigma_n]$ such that $h_1(\Delta t_{opt}) \leq 0$ ($\beta = 1$ represents the condition of Lemma 2.1). Let $t_{k+1} = t_k + \Delta t_{opt}$, then t_k and t_{k+1} satisfy the condition of Lemma 2.1 and we can determine $\varphi_{\mathcal{I}}(t_{k+1})$ in an unique way, as describe in algorithm 1 :

ALGORITHM 1: Algorithm to follow a determination of $\arg(\det(\gamma(t_{k+1})I - A))$

```

teta(k + 1) := arg(det(γ(tk+1)I - A)) ;
if teta(k) < π and (teta(k + 1) > teta(k) + π or teta(k + 1) = 0)
    n(k + 1) := n(k) - 1 ;
else if teta(k) > π and (teta(k + 1) ≤ teta(k) + π or
    teta(k + 1) = 0)
    n(k + 1) := n(k) + 1 ;
else
    n(k + 1) := n(k) ;
endif

```

At the end of the pathfollowing method, $n(k_{end})$ is the number of eigenvalues surrounded by the level curve.

Practically, we do not use $\beta = 1$ because of the roundoff error during the computation of $\arg(\det(\gamma(t_{k+1})I - A))$. A bound of this error is given by $\varepsilon \frac{\|A\|}{\sigma_n}$ which is constant along the level curve. Therefore, we have $\beta = 1 - \varepsilon \frac{\|A\|}{\pi \sigma_n}$. In the case of a well-conditioned matrix, β will be close to 1.

4 Quadratures in the case of parameterized curves

In this section, we consider general integrators to compute (3) along a parametrized user-defined curve. Actually, except for the trapezoidal rule with constant step-size, the methods can be adapted to path following techniques as well.

4.1 Trapezoidal quadrature with constant stepsize

Let $\Gamma \subset \mathbb{C}$ a Jordan curve, and $\gamma(t)_{0 \leq t \leq 1}$ a parametrization of the contour Γ . We have to evaluate the integral :

$$N_\Gamma = \frac{1}{2i\pi} \int_{\gamma(0)}^{\gamma(1)} \frac{\frac{d}{dt} \det(\gamma(t)I - A)}{\det(\gamma(t)I - A)} dt \quad (17)$$

Because Γ is closed, we have $\gamma(0) = \gamma(1)$. So, we can extend the parametrization $\gamma(t)_{0 \leq t \leq 1}$ to $\gamma_{ext}(t)$ for $t \in \mathbb{R}$ with $\gamma_{ext}(t) = \gamma(t \bmod 1)$. We assume that $\gamma_{ext}(t) \in C^{m+1}(\mathbb{R})$. We have the following result about quadrature for periodic functions [3] which is a consequence of the Euler-Mac Laurin formula :

Theorem 4.1 *Let f a $(\beta - \alpha)$ -periodic function, and $f \in C^{m+1}(\mathbb{R})$. Let a subdivision of interval $[\alpha, \beta]$ in k intervals with equidistant points $x_i = \alpha + ih$ and $h = \frac{\beta - \alpha}{k}$. We consider the integration of $\int_\alpha^\beta f(x)dx$ by the trapezoidal rule :*

$$T_k(f) = h \left[\frac{1}{2} f(x_0) + f(x_1) + \dots + f(x_{k-1}) + \frac{1}{2} f(x_k) \right]$$

Then, we have

$$\left| \int_\alpha^\beta f(x)dx - T_k(f) \right| \leq C_m h^{m+1}.$$

This integration method ensures a convergence to the exact integral. The convergence rate depends on the regularity of the curve Γ . It is thus very interesting in the case of a circle or an ellipse, which implies that the convergence is faster than any power.

The drawback of the method is the constant step size along Γ that may imply too many evaluations of determinants. Because of the computational cost, this is not a desirable property. The points computed at the previous steps must be reused at the current step. A method which satisfies the constraint, consists of considering a stepsize twice smaller than the previous one for every iteration.

We call it Strategy 1 method ; it leads to the following algorithm :

ALGORITHM 2: Trapezoidal method with constant step size

```

k := 1 ;
h := 0.5 ;
f(1) := u(γ(0)) ; f(2) := u(γ(0.5)) ; f(3) := f(1) ;
repeat
  k := k + 1 ;
  h := h/2 ;
  f[1 : 2 : 2k + 1] := f[1 : 2k-1 + 1] ;
  f[2 : 2 : 2k] := u(γ([1 : 2 : 2k - 1]/h)) ;
  N(k) := (∑j=12k+1 f(j)) / (2πh) ;
until |N(k) - round(N(k))| < ε

```

To evaluate the function $u(t) = \frac{\frac{d}{dt}f(\gamma(t))}{f(\gamma(t))}$ which must be integrated, we consider a first order approximation of the derivative in the direction of contour line (see §5.3).

Now, we must describe the stopping criterion of the algorithm. The study of (3) ensures that the real part of $\varphi(1)$ must be equal to zero, and the imaginary part of $\varphi(1)$ is a multiple of 2π . Those two properties will be used to stop the integration. The method will be said to have converged when

- the difference of quadrature between two steps of the algorithm will be below a given threshold,
- the imaginary and real part will be near a value in accordance with the properties mentioned above.

A drawback of the previous strategy defining the sequence of steps is its lack of progressivity : every step doubles the amount of work performed in the total of the previous steps. A second strategy, called Strategy 2, is a combination of partitioning based on radix two and three such that in five steps, an interval is divided into 18 subintervals whereas Strategy 1 splits it into 32 parts. Strategies 1 and 2 are illustrated in Figure 1. It is expected to see that Strategy 2 is more efficient than Strategy 1, especially, when the number of steps necessary to reach convergence is high.

A weakness of the constant stepsize method is that it does not take into account rapid variations of the function. In order to integrate a local stiffness of the function, the stepsize will be reduced along the whole domain of integration, whereas a local reduction of the step would have been more adequate.

Such local stiffness of the function occur in neighbourhoods of eigenvalues. This method of integration will only be considered for curves far from eigenvalues. This notion of distance between the curve and eigenvalues must be related to the size of the domain. For instance, if the curve turns around an eigenvalue

We get the following two approximations of $I_{k-1,k+1}$:

$$T_{k-1,k+1}^{(1)} = \frac{h_{k-1} + h_k}{2} (u(t_{k-1}) + u(t_{k+1})) \quad (18)$$

$$T_{k-1,k+1}^{(2)} = \frac{h_{k-1}}{2} (u(t_{k-1}) + u(t_k)) + \frac{h_k}{2} (u(t_k) + u(t_{k+1})). \quad (19)$$

Supposing $h_k = O(h)$, we also define the two quadrature errors :

$$\begin{aligned} E_{k-1,k+1}^{T,1} &= I_{k-1,k+1} - T_{k-1,k+1}^{(1)} \\ &= -\frac{1}{12} (h_{k-1} + h_k)^3 u''(t_k) + O(h^4) \\ E_{k-1,k+1}^{T,2} &= I_{k-1,k+1} - T_{k-1,k+1}^{(2)} \\ &= -\frac{1}{12} (h_{k-1}^3 + h_k^3) u''(t_k) + O(h^4). \end{aligned}$$

The quadrature error at step k can be expressed by :

$$E_{k,k+1} = I_{k,k+1} - T_{k,k+1}^{(1)} = -\frac{1}{12} h_k^3 u''(t_k) + O(h^4) \quad (20)$$

where $T_{k,k+1}^{(1)} = \frac{h_k}{2} (u(t_k) + u(t_{k+1}))$. Since

$$\begin{aligned} E_{k-1,k+1}^{T,1} - E_{k-1,k+1}^{T,2} &= T_{k-1,k+1}^{(1)} - T_{k-1,k+1}^{(2)} \\ &= \frac{1}{4} h_{k-1} h_k (h_{k-1} + h_k) u''(t_k) + O(h^4), \end{aligned}$$

the following estimation

$$u''(t_k) = \frac{4(T_{k-1,k+1}^{(1)} - T_{k-1,k+1}^{(2)})}{h_{k-1} h_k (h_{k-1} + h_k)} + O(h),$$

can be plugged into (20)

$$|E_{k,k+1}| \leq \frac{1}{3} \frac{h_k^2}{h_{k-1}^2} |T_{k-1,k+1}^{(1)} - T_{k-1,k+1}^{(2)}| + O(h^4) \quad (21)$$

and then the condition

$$h_k < h_{k-1} \sqrt{\frac{3\epsilon}{|T_{k-1,k+1}^{(1)} - T_{k-1,k+1}^{(2)}|}} \quad (22)$$

maintains the error below the threshold ϵ , as long as the error expansion is valid.

We apply this result to $u(t) = \frac{d}{dt} \frac{f(\gamma(t))}{f(\gamma(t))}$, and we assume that we know exactly the integral on $[0, t_k]$. A stepsize h_k is therefore suggested by (22). We compute $\int_{t_{k-1}}^{t_{k+1}} u(t) dt$ in two different ways by (18) and (19) and obtain by (21) an evaluation of the error $|E_{k,k+1}|$. This error can be used to accept the proposed step h_k since $|E_{k,k+1}| < \pi$ in the quadrature from t_k to t_{k+1} implies the exact knowledge of the determination of the logarithm in t_{k+1} and therefore of the integral on $[0, t_{k+1}]$.

4.3 Integration by interpolating polynomials

We now compute the integral by interpolation polynomials as it is done in the Adams schemes that are ODE integrators.

Let $\mathcal{P}_{k-s,k}(t)$ be the Lagrange polynomial of degree s interpolating $u(t) = \frac{d}{dt} \frac{det(z(t)I-A)}{det(z(t)I-A)}$ in $(t_j)_{j=k-s,\dots,k}$. Therefore :

$$\mathcal{P}_{k-s,k}(t) = \sum_{i=0}^s u(t_{k-i}) \frac{\prod_{j=0, j \neq i}^s (t - t_{k-j})}{\prod_{j=0, j \neq i}^s (t_{k-i} - t_{k-j})},$$

and $I_{k,k+1} = \int_{t_k}^{t_{k+1}} u(t)dt$ can be approximated by the Adams-Bashforth approximation

$$A_{k,k+1}^{(1)} = \int_{t_k}^{t_{k+1}} \mathcal{P}_{k-s,k}(t)dt. \quad (23)$$

By interpolating u with $\mathcal{P}_{k+1-s,k+1}(t)$, a second approximation is obtained (Adams-Moulton approximation) :

$$A_{k,k+1}^{(2)} = \int_{t_k}^{t_{k+1}} \mathcal{P}_{k+1-s,k+1}(t)dt. \quad (24)$$

We obtain the quadrature errors (by denoting $h_j = t_{j+1} - t_j = O(h)$):

$$\begin{aligned} E_{k,k+1}^{A,1} &= I_{k,k+1} - A_{k,k+1}^{(1)} \\ &= \frac{u^{(s+1)}(t_k)}{(s+1)!} \int_{t_k}^{t_{k+1}} \prod_{i=0}^s (t - t_{k-i})dt + O(h^{s+2}), \end{aligned} \quad (25)$$

and

$$\begin{aligned} E_{k,k+1}^{A,2} &= I_{k,k+1} - A_{k,k+1}^{(2)} \\ &= \frac{u^{(s+1)}(t_k)}{(s+1)!} \int_{t_k}^{t_{k+1}} \prod_{i=0}^s (t - t_{k+1-i})dt + O(h^{s+2}). \end{aligned} \quad (26)$$

Therefore, we obtain the following approximation of $u^{(s+1)}(t_k)$:

$$u^{(s+1)}(t_k) = \frac{(A_{k,k+1}^{(1)} - A_{k,k+1}^{(2)})(s+1)!}{(t_{k+1} - t_{k-s}) \int_{t_k}^{t_{k+1}} \prod_{i=1}^s (t - t_{k+1-i})dt} + O(h). \quad (27)$$

By plugging (27) into (26), we get an evaluation of the quadrature error $E_{k,k+1}^{A,2}$. Then, similarly to the procedure used with the Trapezoidal rule, when this error is lower than a given threshold the predicted stepsize $h_k = t_{k+1} - t_k$ is accepted. Otherwise, the procedure is rerun with a smaller stepsize.

Now, we discuss the way to compute (23) and (24). We consider the polynomials expressed in their divided difference form as in [12].

$$\begin{aligned} \mathcal{P}_{k-s,k}(t) = & u[t_k] + (t - t_k)u[t_k, t_{k-1}] + (t - t_k)(t - t_{k-1})u[t_k, t_{k-1}, t_{k-2}] + \dots + \\ & (t - t_k)(t - t_{k-1})\dots(t - t_{k-s+2})u[t_k, t_{k-1}, \dots, t_{k-s+1}] \end{aligned}$$

with

$$u[t_k, t_{k-1}, \dots, t_{k-s+1}] = \sum_{i=k-s+1}^k u(t_i) \prod_{j=k-s, j \neq i}^{k-1} \frac{1}{t_i - t_j}.$$

We now consider the following formulae :

$$\begin{aligned} h_i &= t_i - t_{i-1}, \\ \psi_i(k+1) &= h_{k+1} + h_k + \dots + h_{k+2-i} & i \geq 1, \\ \alpha_i(k+1) &= \frac{h_{k+1}}{\psi_i(k+1)} & i \geq 1, \\ \phi_1(k) &= u[t_k] = u(t_k), \\ \phi_i(k) &= \psi_1(k)\psi_2(k)\dots\psi_{i-1}(k)u[t_k, t_{k-1}, \dots, t_{k+1-i}] & i > 1, \\ \phi'_i(k) &= \psi_1(k+1)\psi_2(k+1)\dots\psi_{i-1}(k+1)u[t_k, t_{k-1}, \dots, t_{k+1-i}] & i > 1, \\ \phi_i^e(k) &= \sum_{j=i}^s \phi'_j(k-1) & i > 1, \end{aligned} \tag{28}$$

and

$$g_{i,q} = \begin{cases} \frac{1}{q} & i = 1, \\ \frac{1}{q(q+1)} & i = 2, \\ g_{i-1,q} - \alpha_{i-1}(k+1)g_{i-1,q+1} & i \geq 3, \end{cases} \tag{29}$$

from which

$$\begin{aligned} A_{k,k+1}^{(1)} &= \int_{t_k}^{t_{k+1}} \mathcal{P}_{k-s,k}(t) dt \\ &= h_{k+1} \sum_{i=1}^s g_{i,k} \phi_i(k+1) \\ A_{k,k+1}^{(2)} &= \int_{t_k}^{t_{k+1}} \mathcal{P}_{k+1-s,k+1}(t) dt \\ &= A_{k,k+1}^{(1)} + h_{k+1} g_{s,k} (u(t_{k+1}) - \phi_1^e(k+1)). \end{aligned}$$

The use of the polynomials through the divided difference form is very attractive because the quantities defined in (28) can be updated at a low computational cost for the next step. The computation of $E_{k,k+1}^{A,2}$ relies on the estimation of $\int_{t_k}^{t_{k+1}} \prod_{i=1}^s (t - t_{k+1-i}) dt$. The sequence $c_i(j)$ for $0 \leq i \leq s$ and $0 \leq j \leq s+1$ which is defined by :

$$c_i(j) = \begin{cases} c_i(0) &= 0 \\ c_i(i+1) &= 1 \\ c_i(j) &= c_{i-1} - c_{i-1}(j) \times t_{k-s+i}, \end{cases} \tag{30}$$

provides the constants $c_s(j)$ ($j = 1, \dots, s + 1$) which are used in :

$$\int_{t_k}^{t_{k+1}} \prod_{i=1}^s (t - t_{k+1-i}) dt = \sum_{j=1}^{s+1} c_s(j) \frac{t_{k+1}^j - t_k^j}{j}. \quad (31)$$

5 Computation of $u(t) = \frac{\frac{d}{dt}f(\gamma(t))}{f(\gamma(t))}$

The derivative will be evaluated by its first order approximation in the neighbourhood of t . Several facts will impact on the accuracy of the estimation:

- the roundoff error in the determinant evaluation;
- the roundoff error in the first order approximation of the derivative;
- the approximation error due to the first order approximation.

In this section, the three cases above are investigated in order to lower the final error.

5.1 Roundoff error in the determinant evaluation

The quantity $f(\gamma(t)) = \det(\gamma(t)I - A)$ can be computed through a LU factorization of the matrix $(\gamma(t)I - A)$ with partial pivoting. In order to get rid of underflow or overflow, the quantity is computed by

$$\frac{f(\gamma(t))}{|f(\gamma(t))|} = \prod_{k=1}^n \left(\frac{u_{k,k}}{|u_{k,k}|} \right), \quad (32)$$

where $u_{k,k}$ is the k -th diagonal entry of U in the LU factorization. The quantity $|f(\gamma(t))|$ is implicitly recorded by the modulus of the main diagonal of U .

For sake of simplicity in the notation of this subsection, we assume the following notations :

- The matrix $P(\gamma(t)I - A)$, where P is the permutation matrix obtained in the LU factorization, is simply denoted A . Therefore, in this subsection : $A = LU$.
- A_k, L_k, U_k are the principal submatrices of order k ($k = 1, \dots, n$) of A, L, U .
- $\chi(M)$ denotes the condition number of the matrix M for the 2-norm and for any matrix M .
- ε is the machine precision parameter.
- \tilde{L}, \tilde{U} are the obtained factors in finite arithmetic :

$$A + \Delta A = \tilde{L}\tilde{U} \quad (33)$$

From the backward error analysis of the LU factorization [7], we obtain that:

$$|\Delta A| \leq \varepsilon \gamma_n |\tilde{L}| |\tilde{U}| \quad (34)$$

with $\gamma_n = \frac{n}{1-n\varepsilon}$. We assume that $n\varepsilon \ll 1$ and therefore $\gamma_n \simeq n$. The matrices $\Delta L = \tilde{L} - L$ and $\Delta U = \tilde{U} - U$ are, respectively, a strictly lower triangular and an upper triangular matrices.

We assume that the pivots $(\tilde{u}_{k,k})$ are not too small with respect to the arithmetic precision in such a way that we may assume that :

$$\|\Delta L\| = O(\varepsilon \|L\|) \quad \text{and} \quad \|\Delta U\| = O(\varepsilon \|U\|).$$

From (33) we derive :

$$\begin{aligned} I + L^{-1} \Delta A U^{-1} &= (I + L^{-1} \Delta L)(I + \Delta U U^{-1}) \\ L^{-1} \Delta A U^{-1} &= L^{-1} \Delta L + \Delta U U^{-1} + L^{-1} (\Delta L) (\Delta U) U^{-1} \end{aligned}$$

We consider the k -th diagonal entry; since $\text{diag}(L^{-1} \Delta L) = 0$, we get :

$$\begin{aligned} e_k^t L^{-1} \Delta A U^{-1} e_k &= e_k^t \Delta U U^{-1} e_k + e_k^t L^{-1} (\Delta L) (\Delta U) U^{-1} e_k \\ &= \frac{\Delta u_{k,k}}{u_{k,k}} + O(\varepsilon^2 \chi(L_k) \chi(U_k)) \end{aligned}$$

which implies

$$\frac{|\Delta u_{k,k}|}{|u_{k,k}|} \leq e_k^t |L_k^{-1}| |\Delta A_k| |U_k^{-1}| e_k + O(\varepsilon^2 \chi(L_k) \chi(U_k)) \quad (35)$$

where ΔA_k is the principal submatrix of order k of ΔA . From (34), we derive the following bound :

$$\begin{aligned} \frac{|\Delta u_{k,k}|}{|u_{k,k}|} &\leq \gamma_k \varepsilon |L_k^{-1}| |\tilde{L}_k| |\tilde{U}_k| |U_k^{-1}| + O(\varepsilon^2 \chi(L_k) \chi(U_k)) \\ \frac{|\Delta u_{k,k}|}{|u_{k,k}|} &\leq \gamma_k \varepsilon |L_k^{-1}| |L_k| |U_k| |U_k^{-1}| + O(\varepsilon^2 \chi(L_k) \chi(U_k)) \\ \left| \frac{\Delta u_{k,k}}{u_{k,k}} \right| &\leq \gamma_k \varepsilon \chi(L_k) \chi(U_k) + O(\varepsilon^2 \chi(L_k) \chi(U_k)) \end{aligned}$$

Since the function $k \rightarrow \chi(L_k) \chi(U_k)$ is an increasing function, we obtain the following result.

Lemma 5.1 *The roundoff error which occurs at the k -th step in the factorization is bounded by :*

$$\left| \frac{\Delta u_{k,k}}{u_{k,k}} \right| \leq \gamma_k \varepsilon \chi(L) \chi(U) + O(\varepsilon^2 \chi(L) \chi(U)) \quad (36)$$

which provides the following bound for the error on the argument :

$$\sin(|\Delta \theta_k|) \leq \gamma_k \varepsilon \chi(L) \chi(U) + O(\varepsilon^2 \chi(L) \chi(U)) \quad (37)$$

where $\Delta \theta_k = \arg\left(\frac{\tilde{u}_{k,k}}{|\tilde{u}_{k,k}|}\right) - \arg\left(\frac{u_{k,k}}{|u_{k,k}|}\right)$.

Proof. It remains to prove (37). It directly comes from the simple geometric illustration 3. \diamond

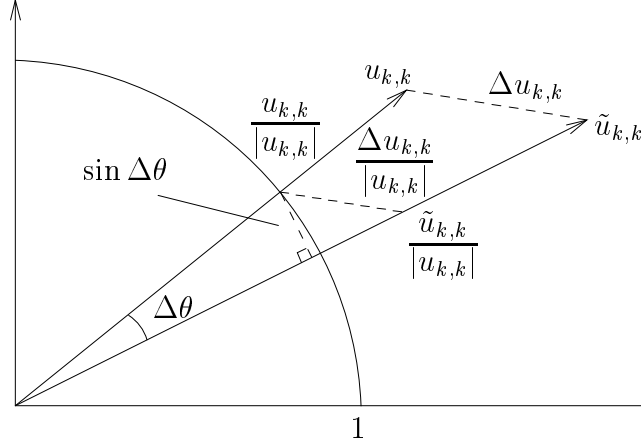


Figure 3: Errors at step k of the LU factorization

We deduce from that the roundoff error on the approximation of $\arg(f(\gamma(t)))$.

Proposition 5.1 *Let e_a the roundoff error*

$$e_a = |\arg(\tilde{f}(\gamma(t))) - \arg(f(\gamma(t)))|.$$

Then

$$e_a \leq \varepsilon \chi(L) \chi(U) \sum_{k=1}^n \gamma_k + O(\varepsilon^2 \chi(L) \chi(U)) \quad (38)$$

where the parameters γ_k are defined in (34).

Proof. In the previous proposition we identify $\sin(\Delta\theta)$ and $\Delta\theta$ since the difference is of the third order, and we cumulate the bounds obtained at all the elementary steps. The roundoff error which occurs when summing all the elementary arguments is omitted since it is of order $O(\varepsilon)$ with a small constant. \diamond

Practically we shall consider that the pivoting strategy insures that $\chi(L) \chi(U)$ is of the same order of magnitude as $\chi(A)$, and that $\gamma_k \simeq k$ for $k = 1, \dots, n$. A practical bound is then :

$$e_a \leq \frac{n^2}{2} \varepsilon \chi(L) \chi(U) \quad (39)$$

5.2 Roundoff error in the derivative estimation

The derivative $\frac{d}{dt}f(\gamma(t))$ is approximated by its first order approximation

$$g_h(t) = \frac{f(\gamma(t+h)) - f(\gamma(t))}{h} \quad (40)$$

where h is some small quantity of which order of magnitude will be discussed in the next subsection.

Actually, $g_h(t)$ is not computed but

$$v_h(t) = \frac{1}{h} \left(\frac{f(\gamma(t+h))}{f(\gamma(t))} - 1 \right)$$

which approximates the logarithmic derivative of $f \circ \gamma$. The practical computation of $v_h(t)$ is obtained by

$$v_h(t) = \frac{1}{h} \left(\frac{f(\gamma(t+h))}{|f(\gamma(t+h))|} \frac{|f(\gamma(t))|}{f(\gamma(t))} \frac{|f(\gamma(t+h))|}{|f(\gamma(t))|} - 1 \right) \quad (41)$$

which only involves quantities that can effectively be computed. $\frac{f(\gamma(s))}{|f(\gamma(s))|}$ ($s = t$ or $t+h$) corresponds to the already discussed determinant evaluation, whereas $\frac{|f(\gamma(t+h))|}{|f(\gamma(t))|}$ can be computed explicitly from the records of the modulus of the pivots in the LU factorization.

A bound for the rounding error in the computation of $\frac{|f(\gamma(t+h))|}{|f(\gamma(t))|}$ is given by:

Proposition 5.2

$$\frac{|\tilde{f}(\gamma(t+h))|}{|\tilde{f}(\gamma(t))|} = \frac{|f(\gamma(t+h))|}{|f(\gamma(t))|} (1 + \eta)$$

with

$$|\eta| \leq \sum_{k=1}^n \theta_k e^{\sum_{i=k+1}^n \theta_i} + O(\varepsilon^2) \quad (42)$$

where

$$\theta_k = \frac{|\Delta u_{k,k}(t)|}{|u_{k,k}(t)|} + \frac{|\Delta u_{k,k}(t+h)|}{|u_{k,k}(t+h)|} \quad (43)$$

($\varepsilon = \text{machine precision parameter}$)

Proof. Because

$$\frac{|\tilde{u}_{k,k}(s) - |u_{k,k}(s)||}{|u_{k,k}(s)|} \leq \frac{|\Delta u_{k,k}(s)|}{|u_{k,k}(s)|} \quad \text{for } s = t, t+h \text{ and } k = 1, \dots, n$$

we get $|\tilde{u}_{k,k}(s)| = |u_{k,k}(s)|(1 + \alpha_k(s))$ with $|\alpha_k(s)| \leq \frac{|\Delta u_{k,k}(s)|}{|u_{k,k}(s)|}$ and therefore

$$\begin{aligned} \frac{|\tilde{u}_{k,k}(t+h)|}{|\tilde{u}_{k,k}(t)|} &= \frac{|u_{k,k}(t+h)|}{|u_{k,k}(t)|} \frac{(1 + \alpha_k(t+h))}{(1 + \alpha_k(t))} \\ &= \frac{|u_{k,k}(t+h)|}{|u_{k,k}(t)|} (1 + \beta_k) \end{aligned}$$

with $|\beta_k| \leq \theta_k + O(\varepsilon^2)$. In the previous expression, we omitted the error on each division of modulus, since their impact is negligible. The final error is evaluated by

$$\prod_{k=1}^n \frac{|\tilde{u}_{kk}(t+h)|}{|\tilde{u}_{kk}(t)|} = \left(\prod_{k=1}^n \frac{|u_{kk}(t+h)|}{|u_{kk}(t)|} \right) \prod_{k=1}^n (1 + \beta_k)$$

It can easily be proved by recursion that :

$$\left| \prod_{k=1}^n (1 + \beta_k) - 1 \right| \leq \sum_{k=1}^n |\beta_k| e^{\sum_{i=k+1}^n |\beta_i|}$$

which ends the proof of the proposition. \diamond

Coming back to the evaluation of v_h , from formula (41), we consider the relative errors which occurred in each factor ¹

$$\tilde{v}_h(t) = \frac{1}{h} \left[\frac{f(\gamma(t+h))}{|f(\gamma(t+h))|} (1 + \delta_1) \cdot \frac{|f(\gamma(t))|}{f(\gamma(t))} (1 + \delta_2) \cdot \frac{|f(\gamma(t+h))|}{|f(\gamma(t))|} (1 + \delta_3) - 1 \right]$$

where

$$\begin{aligned} |\delta_1| &\leq 2\varepsilon \chi(L(\gamma(t+h))) \chi(U(\gamma(t+h))) \left(\sum_{k=1}^n \gamma_k \right), \\ &\quad + O(\varepsilon^2 \chi(L(\gamma(t+h))) \chi(U(\gamma(t+h)))) , \\ |\delta_2| &\leq 2\varepsilon \chi(L(\gamma(t))) \chi(U(\gamma(t))) \left(\sum_{k=1}^n \gamma_k \right) + O(\varepsilon^2 \chi(L(\gamma(t))) \chi(U(\gamma(t)))) , \\ |\delta_3| &\leq \sum_{k=1}^n |\theta_k| e^{\sum_{i=k+1}^n |\theta_i|} + O(\varepsilon^2) \quad \text{where } \theta_k \text{ is defined by (43).} \end{aligned}$$

Therefore

$$|\tilde{v}_h(t) - v_h(t)| \leq \frac{1}{h} (|\delta_1| + |\delta_2| + |\delta_3|) + O(K\varepsilon^2) \quad (44)$$

where $K = 2 \max \chi(L(\gamma(s))) \chi(U(\gamma(s)))$, with $s \in [t, t+h]$.

In practice, the bound will be explicitly computed from the assumptions $2 \sum_{k=1}^n \gamma_k \simeq n^2$ and $\chi(L(\gamma(s))) \chi(U(\gamma(s))) \simeq \chi(A(\gamma(s)))$ when $\sigma_{\min}(A(\gamma(s)))$ is computed. If this quantity is not available, we estimate it roughly by the quantity $\frac{|u_{kmax}| |l_{max}|}{|u_{kmin}|}$ where u_{kmax} and u_{kmin} are the diagonal entries of $U(\gamma(s))$ with maximum and minimum modulus, and l_{max} is the entry of $L(\gamma(s))$ with maximum modulus.

¹For sake of simplicity, we neglect the roundoff error which occurs in the subtraction by 1 and the division by h .

5.3 Search for an acceptable h

Because there are two origins of errors in the evaluation of the derivative, we would like to select a stepsize h for which roundoff error and approximation error are of the same order of magnitude. Figure (4) displays the computed value of the derivative as a function of h for the matrix 3906×3906 introduced in section 6. We may split the evolution into three phases :

- I : for h greater than 10^{-8} , the numerical value \tilde{u}_h is a coarse approximation of u .
- II : $h \in [10^{-11}, 10^{-8}]$, the approximation is acceptable.
- III: for h lower than 10^{-11} , rounding errors become dominant and lead to unexploitable results.

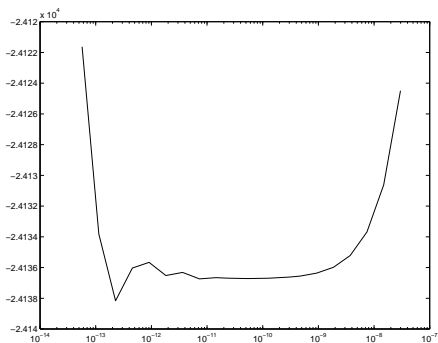


Figure 4: $h \rightarrow \tilde{u}_h(0.532)$

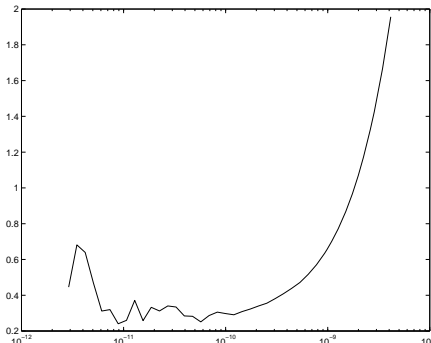


Figure 5: $h \rightarrow \tilde{u}_h(0.532) + 24137$

By zooming phase II (figure (5)), we notice that the range of acceptable results is quite narrow and the goal is to determine an approximation of the lower bound of this area. For that purpose it remains to evaluate the approximation error. By its first order expansion in t , we can write (in exact arithmetic)

$$v_h(t) = \frac{\frac{d}{dt}(f \circ \gamma)(t)}{(f \circ \gamma)(t)} + \frac{h}{2} \left[\frac{\frac{d^2}{dt^2}(f \circ \gamma)(t)}{(f \circ \gamma)(t)} \right] + O(h^2).$$

Therefore the error is expressed by :

$$E_h(t) = \frac{h}{2} \left[\frac{\frac{d^2}{dt^2}(f \circ \gamma)(t)}{(f \circ \gamma)(t)} \right] = \frac{h}{2} \left[\frac{d}{dt} \left(\frac{\frac{d}{dt}(f \circ \gamma)(t)}{(f \circ \gamma)(t)} \right) + \left(\frac{\frac{d}{dt}(f \circ \gamma)(t)}{(f \circ \gamma)(t)} \right)^2 \right]$$

This quantity is estimated from $v_h(t)$ and by a first order approximation of its derivative obtained from the last computed points. Let $K(t) = \frac{d}{dt} \left(\frac{\frac{d}{dt}(f \circ \gamma)(t)}{(f \circ \gamma)(t)} \right) +$

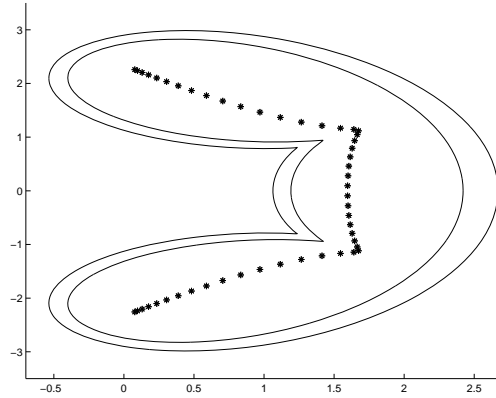


Figure 6: Pseudospectra of G_{50} for $\epsilon = 1.10^{-2}$ and 5.10^{-2}

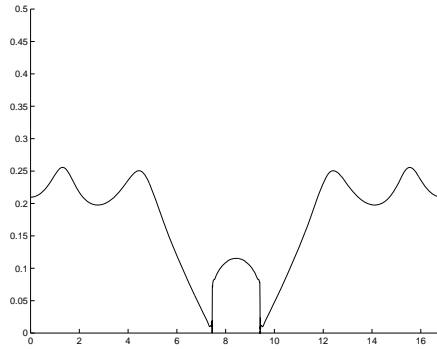


Figure 7: Argument increase between two steps for $\epsilon = 1.10^{-2}$

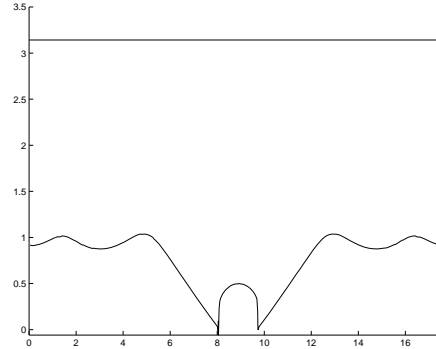


Figure 8: Argument increase between two steps for $\epsilon = 5.10^{-2}$

for $\epsilon = 1.10^{-2}$ and $\epsilon = 5.10^{-2}$. It must be noticed that the stepsize control improves the behaviour of the path following method. This can be seen at the irregular points of the curve for which the strategy of constant step size may have problems [2]. At those points the step size is automatically and drastically decreased.

6.1.2 Parametrized contour

We now consider the case of a parametrized contour. Let Γ be the circle defined by its center $z_c = 0.8 + 0i$ and its radius r . The parametrization is given by $\gamma(t) = z_c + r(\cos(2\pi t) + i\sin(2\pi t))$ with $t \in [0, 1]$. We compare the methods described in Section 4.

Let us first consider $r = 2.9$. In that case, the circle surrounds all the eigenvalues of G_{50} (see Figure 9). The function $u(t)$ which must be integrated

Method	# steps	# determinant eval.
Trap. rule - constant steps	S1 : 32 / S2 : 54	S1 : 64 / S2 : 108
Trap. rule - variable steps	45	120
Adams (order 4)	59	128

(S1 : strategy 1 ; S2 : strategy 2 for the stepsize definition)

Table 1: Counting the eigenvalues of G_{50} for a smooth situation

Method	nb of steps	nb of determinant eval.
Trap. rule with constant steps	S1 : 512 / S2 : 243	S1 : 1024 / S2 : 486
Trap. rule with variable steps	87	236
Adams (order 4)	80	176

(S1 : strategy 1 ; S2 : strategy 2 for the stepsize definition)

Table 2: Counting the eigenvalues of G_{50} for a stiff situation

is represented in Figure 10.

The table 1 gathers the number of steps for each methods. The most efficient method is the trapezoidal quadrature with constant stepsize for both strategies. This behaviour was easy to foresee because the shape is C^∞ and the function $u(t)$ to integrate is smooth. We notice that for the two variable stepsize methods, the number of computed determinants is larger than twice the number of steps, because it also takes into account the rejected steps.

To observe the behaviour of the methods for a stiff function, we consider the circle with radius $r = 1.93$. Figure 11 shows that the circle crosses some dense areas of eigenvalues which makes the function, $u(t)$, stiff to integrate (see Figure 12). In that case, small step sizes are needed when t varies in the neighborhood of 0.3 and 0.7. For the trapezoidal method with constant stepsize, small steps are needed on the whole interval $t \in [0, 1]$. With an adaptative stepsize, we obtain better results.

The Adams method appears to be the winner in the case of stiff problems. For the trapezoidal rule, Strategy 2 is better than Strategy 1 as soon as the number of refinement steps is large enough. When the number of refinement is small (as in the first case), then the two methods behave similarly and Strategy 1 may appear to be less expensive.

6.2 Example 2 : Matrix from a hydrodynamic problem

We consider now a matrix obtained from a physical problem. An incompressible fluid is enclosed between two coaxial cylinders, and two horizontal disks. The lower disk is rotating with a constant angular velocity. We assume that the flow is axisymmetric, and so, we only work in a two dimensional plane. We consider a discretization of this domain on an uniform grid 65×129 that leads to a linearized operator (L) of the Jacobian of size 16002×16002 . In order

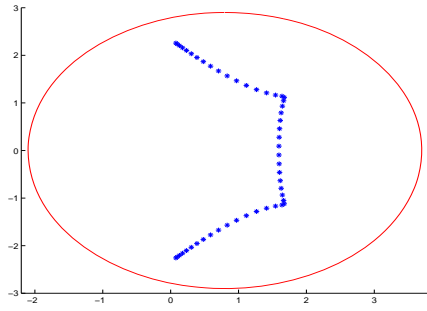


Figure 9: Circle surrounding the eigenvalues of G_{50}

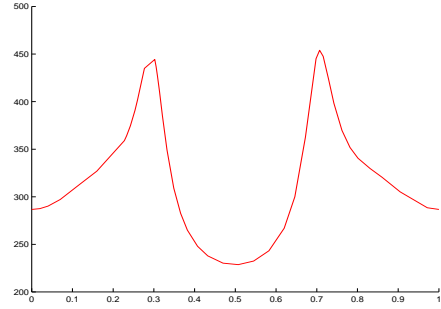


Figure 10: Evolution of $\Im\left(\frac{\frac{d}{dt}\det(\gamma(t)I-G_{50})}{\det(\gamma(t)I-G_{50})}\right)$ along the circle with radius $r = 2.9$

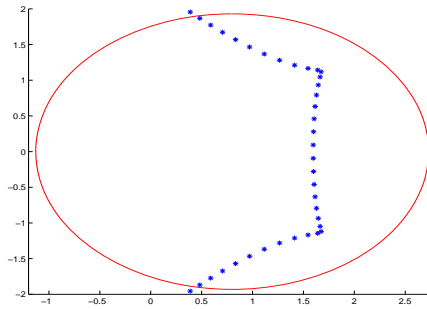


Figure 11: Parametrized contour through the eigenvalues

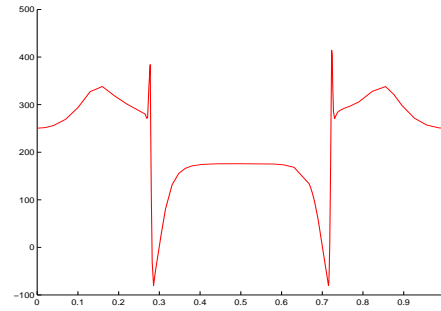


Figure 12: Evolution of $\Im\left(\frac{\frac{d}{dt}\det(\gamma(t)I-G_{50})}{\det(\gamma(t)I-G_{50})}\right)$ along the circle with radius $r = 1.93$

to study the stability of a stationary solution, it is useful to know where the eigenvalues of L lie.

Let \mathcal{D} be the domain of the complex plane defined by

$$\mathcal{D} = \{z \in \mathbb{C}, |\Im(z)| \leq 20, 0 \leq \Re(z) \leq 1\}. \quad (45)$$

We want to know how many eigenvalues lie in this area. We choose $\Re(z) \geq 0$ to take into account only the eigenvalues that lead to unstability of the stationary solution. Because the linearized operator $L \in \mathbb{R}^{16002 \times 16002}$ is real, the spectrum is symmetric and we restrict our domain to

$$\mathcal{D}_+ = \{z \in \mathbb{C}, 0 \leq \Im(z) \leq 20, 0 \leq \Re(z) \leq 1\}. \quad (46)$$

We obtain that 14 eigenvalues lie in \mathcal{D}_+ . Therefore 28 eigenvalues lie in \mathcal{D} . We report the performance of the methods in Table 3.

Method	nb of step	nb of determinant
Trapezoidal rule with variable steps	126	296
Adams (order 4)	168	396

Table 3: Results for the matrix $L \in \mathbb{R}^{16002 \times 16002}$

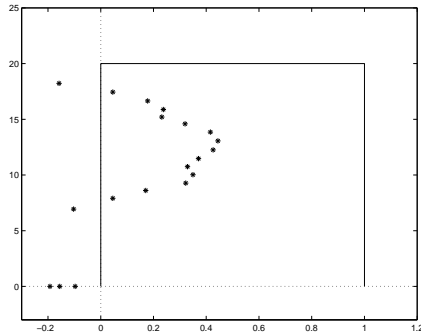


Figure 13: Eigenvalues of L and the domain \mathcal{D}_+

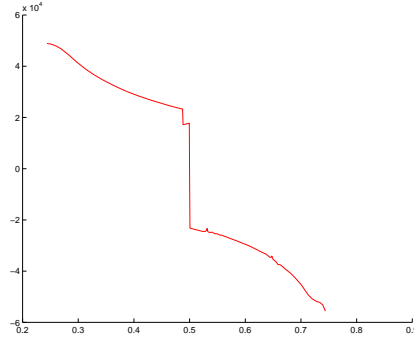


Figure 14: Evolution of $\Im \left(\frac{\frac{d}{dt} \det(\gamma(t)I-L)}{\det(\gamma(t)I-L)} \right)$ along \mathcal{D}_+

Figure 14 represents the function which is integrated. The parameter t can be read on the X-axis. Because the parametrization is done on the whole contour of \mathcal{D} , and because we only consider the contour of \mathcal{D}_+ (omitting the real axis), we use the parameter t in a range of width 0.5 (half of the whole closed contour). It is important to notice that the two breaks of the function at t around 0.5, correspond to the corners of the domain. It means that they can be predicted, and therefore have no effect on the accuracy of the algorithms. On the other hand, these two corners prevent from using the method of trapezoidal

quadrature with constant stepsize that take advantage of a smoother contourline. The results that are displayed in Table 3 show a comparable efficiency for the two methods, Trapezoidal rule and Adams method.

7 Conclusion

New tools are now considered to help the user in localizing eigenvalues with high reliability. There is a hope for building such toolboxes in a near future. The problem comes usually from the computational cost of the routines (eg. dichotomy of spectrum). Although the user might be willing to pay the price of a high reliability, it is necessary to decrease the amount of calculations. Progress have already been made for the situation of the ϵ -spectrum calculation but more is necessary to deal with matrices arising from large applications.

In this paper, procedures with feasible computational costs are proposed to count the number of eigenvalues of a given matrix that are surrounded by a curve (a priori given or determined step by step in the case of a level curve of the norm of the resolvent).

The method based on a series expansion of the logarithm assumes the computations of the smallest (at least) singular value and determinants. It is well adapted to the case of the boundary of the ϵ -spectrum. Its limit comes from the severe constraints on the step size that are imposed to guarantee that the series converges and that the branch of the logarithm is correctly followed.

The methods based on quadratures are well adapted to the case of an a priori known curve. The trapezoidal rule is considered under two versions depending on the step size which can be either constant or variable. When the curve is highly regular (as a circle for instance) the constant stepsize might be the most efficient solution since in that case there is an exponential convergence of the quadrature. However, when the norm of the resolvent is highly varying on the curve, it becomes more efficient to consider an adaptative stepsize. The schemes of higher order may bring some improvement but this is not always the situation.

The methods are limited by the factorization of the matrix. This is an expansive calculation that must be performed in each point on the discretized curve. In this paper, we report an example of order $n = 16002$.

In conclusion, the obtained routines are already acceptable solutions even if it is expected to obtain improvements in the future versions.

Acknowledgement

The authors are indebted to S. K. Godunov who suggested this work when he was visiting Irisa.

References

- [1] C. Bekas and E. Gallopoulos. Cobra : A hybrid method for computing the matrix pseudospectrum. In *Copper Mountain Conference on Iterative Methods*, 1998.
- [2] M. Brühl. A curve tracing algorithm for computing the pseudospectrum. *BIT*, 36:3:441–454, 1996.
- [3] M. Crouzeix and A.L. Mignot. *Analyse numérique des équations différentielles*. Masson, 1984.
- [4] S.K. Godunov. Spectral portrait of matrices and criteria of spectrum dichotomy. In eds L. Athanassova and J. Herzberger, editors, *3rd Internat. IMACS-CAMM Symposium on Computer Arithmetic and Enclosure Methods*, 1991.
- [5] P. Henrici. *Applied and computational complex analysis*, volume 1. Wiley-interscience, 1974.
- [6] V. Heuveline, B. Philippe, and M. Sadkane. Parallel computation of spectral portrait of large matrices by davidson type methods. *Numerical Algorithms*, (16):55–75, 1997.
- [7] N.J. Higham. *Accuracy and stability of numerical algorithms*. Siam, 1996.
- [8] B.J. Hoenders and C.H. Slump. On the calculation of the exact number of zeroes of a set of equations. *Computing*, 30:137–147, 1983.
- [9] R.A. Horn and C.R. Johnson. *Topics in matrix analysis*. Cambridge University Press, 1991.
- [10] P. Kravanja, R. Cools, and A. Haegemans. Computing zeros of analytic mappings: a logarithmic residue approach. *BIT*, 38(3):583–596, 1998.
- [11] J.N. Lyness and L.M. Delves. A numerical method for locating the zeros of an analytic function. *Math. Comp.*, 21:543–560, 1967.
- [12] Shampine and Gordon. *Computer solution of ordinary differential equations*. W.H. Freeman and company, 1975.
- [13] R. A. Silverman. *Introductory complex analysis*. Dover, 1972.
- [14] L. N. Trefethen. Pseudospectra of matrices. In D. F. Griffiths and G. A. Watson, editors, *14th Dundee Biennial Conference on Numerical Analysis*, 1991.
- [15] M.N. Vrahatis. Solving systems of nonlinear equations using the nonzero value of the topological degree. *ACM Trans. on Math. Software*, 14(4):312–329, 1988.

- [16] M.N. Vrahatis and D.J. Kavvadias. Locating and computing all the simple roots and extrema of a function. *SIAM J. Sci. Comp.*, 17(5):1232–1248, 1996.