Approximation numérique de racines isolées multiples de systèmes analytiques

Marc Giusti, Jean-Claude Yakoubsohn

APPROXIMATION NUMÉRIQUE DE RACINES ISOLÉES MULTIPLES DE SYSTÈMES ANALYTIQUES

par

Marc Giusti et Jean-Claude Yakoubsohn

Résumé. – L'approximation d'une racine isolée multiple est un problème difficile. En effet la racine peut même être répulsive pour une méthode de point fixe comme la méthode de Newton. La littérature sur le sujet est vaste mais les réponses proposées pour résoudre ce problème ne sont pas satisfaisantes. Des méthodes numériques qui permettent de faire une analyse locale de convergence sont souvent élaborées sous des hypothèses particulières. Ce point de vue privilégiant l'analyse numérique néglige la géométrie et la structure de l'algèbre locale. C'est ainsi qu'ont émergé des méthodes qualifiés de symboliques-numériques. Mais l'analyse numérique précise de ces méthodes pourtant riches d'enseignement n'a pas été faite.

Nous proposons dans cet article une méthode de type symbolique-numérique dont le traitement numérique est certifié. L'idée générale est de construire une suite finie de systèmes admettant la même racine, appelée *suite de déflation*, telle que la multiplicité de la racine chute strictement entre deux systèmes successifs. La racine devient ainsi régulière lors du dernier système. Il suffit alors d'en extraire un système carré régulier pour obtenir que nous appelons *système déflaté*. Nous avions déjà décrit la construction de cette suite de déflation quand la racine est connue. L'originalité de cette étude consiste d'une part à définir une suite de déflation à partir d'un point proche de la racine et d'autre part à donner une analyse numérique de cette méthode. Le cadre fonctionnel de cette analyse est celui des systèmes analytiques constitués de fonctions de carré intégrable. En utilisant le noyau de Bergman, noyau reproduisant de cet espace fonctionnel, nous donnons une α-*théorie à la Smale* de cette suite de déflation. De plus nous présentons des résultats nouveaux relatifs à la détermination du rang numérique d'une matrice et à celle de la proximité à zéro de l'application évaluation. Comme conséquence importante nous donnons un algorithme de calcul d'une suite de déflation qui est *libre de* ε, quantité-seuil qui mesure l'approximation numérique, dans le sens que les entrées de cet algorithme ne comportent pas la variable ε.

***Classification mathématique par sujets* (2010).** – 65F30, 65H10, 65Y20, 68Q25, 68W30.
Mots clefs. – systèmes d'équations, racines singulières, déflation, rang numérique, évaluation.
Key words. – systems of equations, singular roots, deflation, numerical rank, evaluation.

Abstract. – The approximation of a multiple isolated root is a difficult problem. In fact the root can even be a repulsive root for a fixed point method like the Newton method. However there exists a huge literature on this topic but the answers given are not satisfactory. Numerical methods allowing a local convergence analysis work often under specific hypotheses. This viewpoint favouring numerical analysis forgets the geometry and the structure of the local algebra. Thus appeared so-called symbolic-numeric methods, yet full of lessons, but their precise numerical analysis is still missing.
We propose in this paper a method of symbolic-numeric kind, whose numerical treatment is certified. The general idea is to construct a finite sequence of systems, admitting the same root, and called the *deflation sequence*, so that the multiplicity of the root drops strictly between two successive systems. So the root becomes regular. Then we can extract a regular square we call *deflated system*. We described already the construction of this deflated sequence when the singular root is known. The originality of this paper consists on one hand to construct a deflation sequence from a point close to the root and on the other hand to give a numerical analysis of this method. Analytic square integrable functions build the functional frame. Using the Bergman kernel, reproducing kernel of this functional frame, we are able to give a α-*theory à la Smale*. Furthermore we present new results on the determinacy of the numerical rank of a matrix and the closeness to zero of the evaluation map. As an important consequence we give an algorithm computing a deflation sequence *free of* ε, threshold quantity measuring the numerical approximation, meaning that the entry of this algorithm does not involve the variable ε.

Table des matières

Index des notations

1. Systèmes équivalents et multiplicité

L'article **Multiplicity hunting and approximating multiple roots of polynomial systems** [**15**] fut écrit par les deux auteurs de manière purement heuristique. Nous en présentons ici une analyse numérique, en simplifiant au passage l'algorithme initialement exhibé.

Définition 1. – *Une solution ζ d'un système analytique $f = 0$ (défini dans un voisinage de ζ) est dite* isolée *et* multiple *si :*

1– il existe un voisinage de ζ où ζ est la seule solution de $f = 0$;

2– la matrice Jacobienne $Df(\zeta)$ n'est pas de rang plein.

Une solution est dite isolée et régulière si le 1 est satisfait et $Df(\zeta)$ est de rang plein.

Nous utiliserons indifféremment les mots multiple et singulier. En particulier nous appellerons un système *singulier* s'il admet une solution singulière isolé. De même nous emploierons indistinctement les vocables *solution* ou *racine* d'un système. Afin de simplifier la lecture, la notation f désignera indistinctement une seule équation ou un système d'équations. Implicitement les boules seront toujours ouvertes.
Remarquons que la première hypothèse implique que le nombre d'équations s est plus grand ou égal au nombre n de variables. Notons aussi que ce cadre inclut le cas d'un système analytique obtenu par localisation d'un système algébrique.

Nous avions expliqué dans [**15**] comment dériver un système *régulier* (c'est-à-dire admettant ζ comme solution isolée avec la matrice jacobienne $Df(\zeta)$ de rang plein) d'un système singulier : évidemment ceci sous l'hypothèse que la solution ζ est exactement connue. Nous avions alors formalisé cette transformation par la notion de systèmes *équivalents* en un point ζ.

Notons que cette transformation est obtenue **sans ajout de nouvelles variables** (trait important que nous soulignons).

La *multiplicité* d'une racine est un important invariant. Dans le cas où il n'y a qu'une variable et qu'une équation, la multiplicité d'une racine est exactement le nombre de dérivées qui s'annulent en la racine, propriété qui malheureusement n'est plus valable dans le cas général. Il faut alors introduire une machinerie bien plus compliquée.

Appelons :

1– $\mathbf{C}\{x-\zeta\}$ l'algèbre des germes de fonctions analytiques en ζ, c'est-à-dire l'anneau local des séries convergentes dans un voisinage de ζ, d'idéal maximal engendré par $x_1 - \zeta_1, \ldots, x_n - \zeta_n$;

2– $I\mathbf{C}\{x-\zeta\}$ l'idéal induit engendré par l'idéal $I = I(f) :=< f_1, \ldots, f_s >$.

Définition 2. – *La multiplicité $\mu(\zeta)$ d'une racine isolée ζ est définie comme la dimension de l'espace quotient $\mathbf{C}\{x-\zeta\}/I\mathbf{C}\{x-\zeta\}$.*

Relativement à un ordre local $<$ compatible de $\mathbf{C}\{x-\zeta\}$, nous notons $LT_<(I\mathbf{C}(x-\zeta))$ l'idéal engendré par les termes dominants de tous les éléments de $I\mathbf{C}\{x-\zeta\}$.

Définition 3. – *Une base standard (minimale) de $I\mathbf{C}\{x-\zeta\}$ est un ensemble (fini) de séries de $I\mathbf{C}\{x-\zeta\}$ dont les termes dominants engendrent minimalement $LT_<(I\mathbf{C}(x-\zeta))$.*

Il existe alors un nombre fini de monômes, appelés monômes standard, qui n'appartiennent pas à I. Le théorème suivant est classique dans la littérature sur les bases standard, voir [**3**] page 178.

Théorème 1. – *Les assertions suivantes sont équivalentes :*

1– La racine ζ est isolée ;
2– $dim\,\mathbf{C}\{x-\zeta\}/I\mathbf{C}(x-\zeta)$ est fini ;
3– $dim\,\mathbf{C}\{x-\zeta\}/LT_<(I\mathbf{C}(x-\zeta))$ est fini ;
4– Il y a seulement un nombre fini de monômes standard ;

Qui plus est, quand n'importe laquelle de ces conditions est satisfaite, nous avons :

$$\mu(\zeta) = dim\,\mathbf{C}\{x-\zeta\}/LT_<(I\mathbf{C}(x-\zeta)) = \text{nombre de monômes standard.}$$

Dans le cas particulier d'un système polynomial localisé, dont quelque entier d borne supérieurement le degré total des équations, d^n constitue une borne supérieure pour la multiplicité de toute racine multiple.

2. Ce que contient cette étude

Approcher une racine isolée multiple est difficile car la racine peut être répulsive pour une méthode de point fixe comme la méthode de Newton (voir l'exemple donné par Griewank et Osborne [**20**], p. 752). Dans ce cas de racine isolée, l'arsenal des techniques développées si le rang de la matrice jacobienne est de rang constant ne s'applique pas. Le cas de rang constant, qui comprend respectivement les cas surjectif et injectif, est bien analysé dans l'ouvrage de J.-P. Dedieu [**9**]. Il existe une vaste littérature sur ce sujet, voir par exemple les articles [**47**], [**1**] et les références qu'ils contiennent.

Afin de pallier cet inconvénient, nous construisons une suite finie de systèmes équivalents, appelée *suite de déflation*, telle que la mutiplicité de la racine chute strictement entre deux systèmes successifs. La racine devient ainsi régulière lors du dernier système. Il suffit alors d'en extraire un système carré régulier pour obtenir ce que nous appelons *système déflaté*. L'opérateur de Newton *singulier* est juste l'opérateur de Newton classique associé au système déflaté.

Comment construisons-nous cette suite de déflation ? Commençons par expliquer l'idée générale de cette construction quand la racine ζ est connue. Tout d'abord on remplace les équations par leur gradients tant que ceux-ci s'annulent en ζ. On obtient ainsi un système équivalent au système initial avec la propriété suivante : chaque ligne de sa matrice jacobienne n'est pas identiquement nulle en ζ. On appelle *sélection* cette opération qui consiste à remplacer les équations par leurs dérivées à un certain ordre. Ensuite, tant que la matrice jacobienne n'est pas de rang plein en la racine, c'est qu'il existe des relations entre les lignes (respectivement entre les colonnes), qui sont données par un complément de Schur. Nous appelons *dénoyautage* l'opération qui consiste à ajouter à certaines équations initiales les éléments de ce complément de Schur. Nous montrerons en section 8 qu'après les

opérations de sélection et de dénoyautage effectuées sur un système singulier, nous obtenons un système équivalent où la multiplicité de la racine a strictement chuté. La méthode de déflation consiste à itérer cette construction, c'est à dire à faire suivre une opération de sélection par une opération de dénoyautage. Le nombre d'itérations nécessaires pour obtenir un système régulier est *la longueur* de la suite de déflation. Un sytème déflaté du système initial est un système régulier obtenu par la méthode de déflation. L'idée de la méthode est en fait assez naturelle. Pour s'en persuader, il suffit de se reporter à l'exemple illustré en sous-section 13.1. Pour conclure ce bref aperçu de la méthode de déflation, il faut souligner tout d'abord qu'il n'y a pas unicité de la suite de déflation. Ensuite que l'opération de sélection est en fait une succession d'opérations de dénoyautage dans le cas de nullité du rang de la matrice jacobienne. Ceci sera développé en section 7.

L'originalité de cette étude consiste d'une part à définir une méthode de déflation à partir d'un point x_0 proche de la racine ζ et d'autre part à donner une analyse numérique de cette méthode. Ceci constituera les sections 7 et 8. De plus nous estimerons le rayon d'une boule centrée en ζ dans laquelle le rang numérique de $Df(x_0)$ est égal au rang de $Df(\zeta)$ pour tout point x_0 de cette boule.

Le but étant d'effectuer l'analyse de cette suite de déflation, cette étude se place dans le contexte des systèmes analytiques constitués de fonctions de carré intégrable. Ainsi nous pouvons représenter une fonction et ses dérivées par un noyau reproduisant efficace : le noyau de Bergman. Ce cadre fonctionnel est décrit en section 5.

Qui plus est, notre étude est *libre de* ε (quantité-seuil qui mesure l'approximation numérique) dans le sens suivant :

Définition 4. – *Un algorithme numérique est dit libre de ε si les entrées de cet algorithme ne comportent pas la variable ε.*

La construction d'une suite de déflation présentée dans la table 3 est libre de ε sous l'hypothèse que la norme définie dans la section 5 (ou à tout le moins une borne supérieure) soit donnée. Pour cela nous présentons des résultats nouveaux afin de déterminer via des algorithmes libres de ε :

1– le rang numérique d'une matrice, dans la section 4 ;

2– la proximité à zéro de l'application évaluation, voir la section 6.

Expliquons les motivations de cette démarche. Si le rang de la matrice jacobienne d'un système de la suite de déflation n'est pas plein on ne dispose pas de critère d'existence d'une racine isolée de ce système. Pour y remédier on exhibe un critère au théorème 5 qui repose sur la surjectivité de l'application évaluation : il dit que si $f(x_0)$ est *petit* il existe un système g et un point y_0 *proches* de f et x_0 respectivement tel que $g(y_0) = 0$. Ce critère est utilisé de façon prédictive pour apppliquer les opérations de sélection et de dénoyautage. Par essence ce critère, qui formalise la notion de valeur suffisamment petite, montre l'existence d'une racine d'un système proche du système initial. Pour cette raison, une fois que la construction de la suite de déflation est terminée, nous devons disposer d'un autre critère pour tester l'existence d'une racine isolée sur le système initial augmenté du système déflaté. C'est pourquoi nous terminons cette étude en faisant l'α-théorie de Smale pour les systèmes constitués de fonctions analytiques de carré intégrables. L'analyse induite par le noyau de Bergman est au centre des résultats obtenus. Nous commençons par donner en section 9 un α-théorème, c'est à dire une condition de l'existence d'une racine

obtenue grâce au théorème de Rouché. Toujours dans le cas régulier, nous établissons ensuite en section 10, un γ-théorème, c'est-à-dire un résultat qui exhibe le rayon d'une boule de convergence quadratique pour l'opérateur de Newton. L'analyse du cas singulier est l'application au système déflaté des α-théorème et γ-théorème respectivement obtenus dans le cas régulier. Ceci est réalisé en section 12. Les résultats de cette section dépendent de l'estimation de la quantité γ du système déflaté. En notant par γ_0 (respectivement, γ_ℓ) la quantité γ en la racine du système initial (respectivement, du système déflaté)définie en (18), nous montrons, voir le théorème 12, que l'inégalité

$$\gamma_\ell \leq \ell + \gamma_0$$

subsiste dans une boule centrée en la racine dont le rayon est proportionnel à l'inverse du carré de la longueur de la suite de déflation.

Cet article peut être vu comme une généralisation de G-Lecerf-Salvy-Y [**16**]. Sous les hypothèses supplémentaires d'un système carré ($s = n$) et d'une racine multiple de dimension de plongement un (c'est-à-dire le rang de la matrice jacobienne chute numériquement de un), nous traitions le cas des grappes de racines en utilisant numériquement le théorème des fonctions implicites. Plus précisément il existe une fonction analytique $\varphi(x_1, \ldots, x_{n-1})$ telle que $\zeta_n = \varphi(\zeta_1, \ldots, \zeta_{n-1})$ et donc ζ_n est une racine de la fonction à une variable $h(x_n) = f_n(\varphi(x_1, \ldots, x_{n-1}), x_n)$. En appliquant [**17**] à la fonction $h(x_n)$, nous pouvons en déduire à la fois la multiplicité de ζ_n et un algorithme approximant rapidement la racine ζ_n. Remarquons que ce résultat inclut le cas des zéros "simples doubles" étudié précédemment par Dedieu et Shub [**10**].

3. Relation avec d'autres travaux

Le cas d'une variable et une équation a été étudié intensivement ; la généralisation de l'opérateur classique de Newton est due à Schröder ([**42**], page 324). L'α-théorie est faite dans [**17**], avec des citations sélectionnées.

Le cas général a été étudié soit d'un point de vue purement symbolique soit d'un point de vue numérique. Nous n'allons pas traiter le cas uniquement symbolique, en nous référant par exemple au livre de Cox, Little, O'Shea [**3**] pour les notions fondamentales et à l'article de Lecerf [**25**] pour la déflation (voir le paragraphe consacré à Ojika plus bas).

Un des pionniers de l'approche numérique est Rall [**38**]. Il traite le cas particulier où la racine multiple ζ satisfait l'hypothèse suivante : il existe un indice m, défini comme la multiplicité de ζ, tel que la suite d'espaces construits itérativement à partir de $N_1 = \mathsf{Ker}\, Df(\zeta)$ par

$$N_{k+1} = N_k \cap \mathsf{Ker}\, Df^{k+1}(\zeta), \quad k = 1 : m-1$$

aboutisse à $N_m = \{0\}$. Il est alors possible de construire itérativement un opérateur qui retrouve la convergence quadratique locale. L'idée consiste à projeter itérativement l'erreur $x_0 - \zeta$ sur les noyaux N_k et leur orthogonal $N_k^\perp$.

Au même moment, l'idée d'utiliser une variante de la méthode de Gauss-Newton afin d'approximer une racine multiple a été examiné par Shamanskii [**43**]. Mais l'algorithme

ne converge quadratiquement vers la racine singulière que sous des hypothèses très particulières.

D'autres techniques, dites d'extension, ont été étudiées, où quelques hypothèses sont faites sur la racine singulière. Par exemple si l'opérateur induit par la projection de $\mathsf{Ker}\, Df(\zeta)$ dans $\mathsf{Ker}\,(Df(\zeta)^*)^\perp$:

$$\pi_{(\mathsf{Ker}\, Df(\zeta)^*)^\perp} D^2 f(\zeta)(z, \pi_{\mathsf{Ker}\, Df(\zeta)})$$

est inversible, alors $(\zeta, 0)$ devient une racine régulière d'un nouveau système, dit étendu, possédant $2n - r$ variables. Le système étendu est bâti à partir du système initial et d'une décomposition en valeurs singulières de la matrice jacobienne. Cette voie est développée par Shen et Ypma [**44**] et généralise une technique d'extension utilisée par Griewank [**18**] dans le cas où la chute de rang n'est que de un. Au début des années 60 une série de papiers traitent purement numériquement de l'approximation des racines multiples à l'aide de techniques semblables, voir [**39**], [**40**], [**6**], [**7**], [**19**], [**8**], [**23**], [**49**]. Mais ni la géométrie du problème ni la notion de multiplicité n'y sont introduites.

Ojika dans [**35**] propose une méthode appelée de *déflation* pour dériver un système régulier d'un singulier, mêlant calculs symboliques et numériques. C'est une généralisation d'un algorithme précédemment développé dans [**36**]. Cette recherche d'un système régulier équivalent fait intervenir une élimination de Gauss mais aucune analyse n'en est donnée, en particulier il n'y a aucune détermination du rang numérique ni de relation avec le concept de multiplicité.

Dans le cas particulier important de la localisation d'un système polynomial et dans un esprit purement symbolique, Lecerf dans [**25**] reprend cet algorithme de déflation qui rend un système régulier *triangulaire*, avec une complexité arithmétique dans :

$$\mathcal{O}\left(n^3(nL + n^\Omega)\mu(\zeta)^2 \log(n\,\mu(\zeta))\right)$$

où n est le nombre de variables, $\mu(\zeta)$ la multiplicité, $3 \leq \Omega < 4$ et L est la longueur d'un calcul d'évaluation du système.

Leykin, Verschelde et Zhao exhibent dans [**26**] une méthode mêlant déflation et extension, fondée sur l'observation suivante : si le rang numérique est r, il existe une solution isolée $(\zeta, \delta) \in \mathbf{C}^n \times \mathbf{C}^{r+1}$ du système

$$Df(x)B\delta = 0, \quad \delta^* h - 1 = 0, \tag{1}$$

où $B \in \mathbf{C}^{n\times(r+1)}$ et $h \in \mathbf{C}^{r+1}$ sont choisis aléatoirement. La multiplicité de la racine (ζ, δ) du système déflaté et étendu chute strictement. Un pas de la méthode consiste alors à ajouter les équations (1). Ceci implique à chaque pas dans le pire des cas un doublement du nombre des variables et des équations. De plus la détermination du rang numérique, reposant sur un travail de Fierro-Hansen [**14**], n'est pas libre de ε. Leur théorème affirme alors qu'il suffit d'exécuter $\mu(\zeta)-1$ pas pour arriver à un système régulier.

Les papiers de Dayton-Zeng [**5**], Dayton-Li-Zeng [**4**] et Nan Li-Lihong Zhi [**29**] relèvent de la même veine et traitent le cas polynomial puis analytique. Plus récemment des cas

particuliers ont été étudiées par Nan Li et Lihong Zhi dans plusieurs travaux [**28**], [**27**]. Mais toutes ces contributions ne fournisent qu'une analyse numérique superficielle de leur algorithme.

La dualité et le rapport avec les matrices de Macaulay constituent le cœur théorique des travaux de Mourrain [**34**], Mantzaflaris et Mourrain [**30**], ou plus récemment de Hauenstein, Mourrain, Szanto [**21**]. Dans ce dernier travail ils proposent quand la racine est connue et dans un contexte purement symbolique, un nouvel algorithme pour déterminer un système régulier à partir du système initial. Le principe est de paramétrer les matrices de multiplication : le système régulier obtenu possède $N + \frac{n(+1)}{2}$ équations et $\frac{n\delta(\delta-1)}{2}$ variables, où N est le nombre d'équations du système initial, n le nombre de variables et δ le cardinal d'une base de l'anneau local. Par ce biais ils étudient également une méthode proche de la nôtre toujours en supposant connue la racine singulière : ils ajoutent les relations entre les colonnes des matrices jacobiennes de rang défectif.

4. Détermination du rang numérique d'une matrice

Le but de cette section est de donner un critère de détermination du rang d'une matrice à partir d'une approximation de ses valeurs singulières sans l'introduction a priori d'un seuil de séparation sur ces valeurs singulières. Ceci conduit à un algorithme libre de ϵ de détermination du rang numérique d'une matrice. Soient $s \geq n$ deux entiers, et M une $s \times n$-matrice à coefficients complexes, $\tilde{U}\tilde{\Sigma}\tilde{V}^* := \tilde{M}$ une décomposition approchée de M en valeurs singulières $\tilde{\sigma_1} \geq \ldots \geq \tilde{\sigma_n}$, avec $\tilde{\Sigma} = diag(\tilde{\sigma_1} \ldots \tilde{\sigma_n})$.

On a le résultat de perturbation suivant :

Lemme 1. – *Soit $M = U\Sigma V^*$ une décomposition en valeurs singulières de M avec $\Sigma = diag(\sigma_1, \ldots, \sigma_n)$. On suppose que*

$$\max(\|\tilde{U}^*\tilde{U} - I_s\|, \|\tilde{V}^*\tilde{V} - I_n\|, \|\tilde{\Sigma} - \tilde{U}^*M\tilde{V}\|) \leq \delta < 1$$

où $\|.\|$ désigne la norme L2 des matrices. Alors on a

$$\frac{\tilde{\sigma_i} - \delta}{1+\delta} \leq \sigma_i \leq \frac{\tilde{\sigma_i} + \delta}{1-\delta}, \qquad 1 \leq i \leq n.$$

Démonstration. – Soient $\bar{\sigma}_i, 1 \leq i \leq n$, les valeurs singulières de la matrice $\tilde{U}^*M\tilde{V}$. Puisque $\tilde{U}^*M\tilde{V} = \tilde{\Sigma} + \tilde{U}^*M\tilde{V} - \tilde{\Sigma}$, on a en utilisant le théorème de Weyl [**46**] concernant la perturbation des valeurs singulières des matrices $\tilde{\Sigma}$ et $\tilde{U}^*M\tilde{V}$:

$$|\tilde{\sigma}_i - \bar{\sigma}_i| \leq \delta, \qquad 1 \leq i \leq n. \tag{2}$$

La matrice $\tilde{U}$ est inversible puisque $\tilde{U}^*\tilde{U} = I_s - (I_s - \tilde{U}^*\tilde{U})$ et $\|I_s - \tilde{U}^*\tilde{U}\| \leq \delta < 1$. Il en est de même pour $\tilde{V}$. On déduit du théorème 3.3 de [**12**] que les valeurs singulières de M et celles de $\tilde{U}^*M\tilde{V}$ satisfont les inégalités

$$|\bar{\sigma}_i - \sigma_i| \leq \delta\sigma_i, \qquad 1 \leq i \leq n. \tag{3}$$

Des inégalités (2) et (3) il vient

$$|\tilde{\sigma}_i - \sigma_i| \leq \delta(\sigma_i + 1), \qquad 1 \leq i \leq n.$$

On en déduit facilement que :

$$\frac{\tilde{\sigma}_i - \delta}{1+\delta} \leq \sigma_i \leq \frac{\tilde{\sigma}_i + \delta}{1-\delta}, \qquad 1 \leq i \leq n.$$

□

Nous considérons les fonctions symétriques élémentaires des $\tilde{\sigma}_i$:

$$s_k = \sum_{1 \leq i_1 < \ldots < i_k \leq n} \tilde{\sigma}_{i_1} \ldots \tilde{\sigma}_{i_k}, \quad k = 1 : n.$$

En d'autres termes, les valeurs singulières approchées sont les racines du polynôme $s(\lambda)$ de degré n :

$$s(\lambda) = \prod_{i=1}^{n} (\lambda - \tilde{\sigma}_i) = \lambda^n + \sum_{i=n-1}^{0} (-1)^{(n-i)} s_{n-i} \lambda^i.$$

Par convention $s_0 = 1$; remarquons que cette convention est naturelle, en ce qu'elle autorise le traitement du cas où toutes les valeurs singulières sont nulles, ce qui signifie que la matrice M est nulle et donc que le rang l'est aussi.

Notre analyse du rang numérique est fondée sur l'introduction des quantités définies ci-dessous.

Définition 5. –

1- $b_k(M) := \sup_{0 \leq i \leq k-1} \left(\frac{s_{n-i}}{s_{n-k}} \right)^{\frac{1}{k-i}}$, $k = 1 : n$;

2- $g_k(M) := \sup_{k+1 \leq i \leq n} \left(\frac{s_{n-i}}{s_{n-k}} \right)^{\frac{1}{i-k}}$, $k = 1 : n-1$ *et* $g_n(M) = 1$;

3- $a_k(M) := b_k(M)\, g_k(M)$, $k = 1 : n$.

Remarque 1. – *Les quantités* b_k, g_k *et* a_k *sont un cas particulier de celles introduites dans* [**17**] *page 261. Le choix de noter* b_k *plutôt que* b_{n-k} *est justifié par l'identité* $\dfrac{s_n}{s_{n-1}} = \dfrac{s(0)}{s'(0)}$ *obtenu pour* $k = 1$ *car ce rapport est noté* β_1 *dans* [**17**].

En fait la détermination des $b_k(M)$ et $g_k(M)$ est donnée par le résultat ci-dessous.

Proposition 1. – *On a :*

1- $b_k(M) = \dfrac{s_{n-k+1}}{s_{n-k}}$, $k = 1 : n$;

2- $g_k(M) = \dfrac{s_{n-k-1}}{s_{n-k}}$, $k = 1 : n-1$.

Démonstration. – C'est une conséquence du théorème 5.2 de [**48**] qui énonce que : *Soient r_0 et r_1 tels que $nr_1-(n-1)r_0\geq 0$. Nous considérons la suite $r_k=kr_1-(k-1)r_0$ pour $k\geq 2$. Tout polynôme $f(x)=\sum_{k=0}^{n}a_{n-k}x^k$ qui n'a que des racines réelles vérifie :*

$$\frac{r_{n-k}}{r_{n-k-1}}\frac{k}{k+1}a_{n-k}^2-a_{n-k-1}a_{n-k+1}\geq 0,\quad k=1:n-1.$$

Avec $r_0=n$ et $r_1=n-1$ nous avons $r_i=n-i$. Les coefficients de $s(\lambda)$ vérifient donc :

$$\frac{i^2}{(i+1)^2}s_{n-i}^2-s_{n-i-1}s_{n-i+1}\geq 0,\quad i=1:n-1.$$

Il s'ensuit que $s_{n-i}^2-s_{n-i-1}s_{n-i+1}\geq 0,\quad i=1:n-1$. C'est à dire

$$\frac{s_{n-i-1}}{s_{n-i}}\leq\frac{s_{n-i}}{s_{n-i+1}},\quad i=1:n-1.$$

Pour $k=1:n$ et $i=0:k-1$ il vient

$$\frac{s_{n-i}}{s_{n-k}}=\frac{s_{n-i}}{s_{n-i-1}}\frac{s_{n-i-1}}{s_{n-i-2}}\cdots\frac{s_{n-k+1}}{s_{n-k}}\leq\left(\frac{s_{n-k+1}}{s_{n-k}}\right)^{k-i}.$$

Donc $b_k(M)=\dfrac{s_{n-k+1}}{s_{n-k}}$. De la même façon nous obtenons la deuxième partie. □

Par simplicité nous noterons a_k, b_k, g_k les valeurs correspondantes $a_k(M), b_k(M), g_k(M)$.

Théorème 2. – *Considérons le polynôme $s(\lambda)$ défini précédemment.*

1– S'il existe un entier m, compris entre 1 et n, avec $a_m<1/9$, alors le polynôme $s(\lambda)$ possède m racines dans la boule $B(0,\varepsilon)$, où :

$$\varepsilon:=\frac{3a_m+1-\sqrt{(3a_m+1)^2-16a_m}}{4g_m}\ ;$$

2– Si $a_1>1/9$ alors $\sigma_n>\dfrac{1}{10g_m}$ où m est l'entier satisfaisant $s_{n-k}=0$, $k=1:m-1$ et $s_{n-m}\neq 0$.

Démonstration. – Prouvons la première des assertions. Comme $a_m<1/9$, la quantité s_{n-m} n'est pas nulle car elle est strictement positive. Considérons les polynômes

$$p(\lambda)=\frac{1}{s_{n-m}}s(\lambda)=\frac{1}{s_{n-m}}\prod_{i=1}^{n}(\lambda-\tilde{\sigma}_i)=\sum_{i=0}^{n}(-1)^{n-i}\frac{s_{n-i}}{s_{n-m}}\lambda^i$$

et

$$q(\lambda)=\sum_{i=m}^{n}(-1)^{n-i}\frac{s_{n-i}}{s_{n-m}}\lambda^i.$$

Lemme 2. – *Posons $t:=g_m|\lambda|$. Alors pour tout λ tel que $|\lambda|<1/g_m$, donc pour tout $t<1$:*

$$|q(\lambda)|\geq|\lambda|^m\frac{1-2t}{1-t}$$

Démonstration. –

$$\begin{aligned}
|q(\lambda)| &= \left|\lambda^m + \sum_{i=m+1}^{n} (-1)^{n-i} \frac{s_{n-i}}{s_{n-m}} \lambda^i\right| \\
&\geq |\lambda|^m - \sum_{i=m+1}^{n} \frac{s_{n-i}}{s_{n-m}} |\lambda|^i \\
&\geq |\lambda|^m \left(1 - \sum_{i=m+1}^{n} \frac{s_{n-i}}{s_{n-m}} |\lambda|^{i-m}\right) \\
&\geq |\lambda|^m \left(1 - \sum_{i\geq m+1} (g_m|\lambda|)^{i-m}\right) \\
&\geq |\lambda|^m \frac{1 - 2g_m|\lambda|}{1 - g_m|\lambda|}. \qquad (4)
\end{aligned}$$

□

Nous prouvons d'abord que 0 est la seule racine de $q(\lambda)$ dans la boule ouverte $B\left(0, \dfrac{1}{2g_m}\right)$. Soit $\nu \in B(0, \frac{1}{2g_m})$ une racine non nulle de $q(\lambda)$. Alors nous avons par le lemme 2

$$0 = q(\nu) = |q(\nu)| \geq |\nu|^m \frac{1 - 2g_m|\nu|}{1 - g_m|\nu|}.$$

Donc $|\nu| \geq \dfrac{1}{2g_m}$.
Considérons le trinôme

$$2t^2 - (3a_m + 1)t + 2a_m. \qquad (5)$$

Si $a_m < 1/9$, alors ce trinôme a deux racines réelles $t_1 < t_2$, car le discriminant

$$\Delta = (3a_m + 1)^2 - 16a_m = 9{a_m}^2 - 10a_m + 1 = (9a_m - 1)(a_m - 1)$$

est strictement positif. Nous pouvons vérifier explicitement que t_1 est strictement positif, puisque ceci se ramène à a_m strictement positif.

Nous prouvons que pour $|\lambda|$ satisfaisant $\dfrac{t_1}{g_m} \leq |\lambda| < \dfrac{1}{2g_m}$, $p(\lambda)$ a m racines, comptées avec multiplicité, dans la boule ouverte $B(0, |\lambda|)$ (notons que la longueur de l'intervalle où $|\lambda|$ vit est strictement positif, puisque $t_1 < 1/2$). Afin d'établir ce fait, nous allons vérifier que l'inégalité de Rouché

$$|p(\lambda) - q(\lambda)| < |q(\lambda)| \qquad (6)$$

est vérifiée sur la sphère de rayon $|\lambda|$. Nous avons

$$\begin{aligned} |p(\lambda) - q(\lambda)| &\leq \sum_{i=0}^{m-1} \frac{s_{n-i}}{s_{n-m}} |\lambda|^i \\ &\leq \sum_{i=0}^{m-1} b_m^{m-i} |\lambda|^i \\ &\leq |\lambda|^m \frac{b_m/|\lambda|}{1 - b_m/|\lambda|} \\ &\leq \frac{a_m}{g_m|\lambda| - a_m} |\lambda|^m. \end{aligned} \tag{7}$$

Nous vérifions que $t - a_m > t_1 - a_m = \dfrac{-a_m + 1 - \sqrt{\Delta}}{4}$ est strictement positif si $a_m < 1/9$.

De (7) et du lemme 2, nous voyons que l'inégalité de Rouché est satisfaite si

$$\frac{a_m}{t - a_m} |\lambda|^m < \frac{1 - 2t}{1 - t} |\lambda|^m.$$

Comme $|\lambda|$, $1 - t$ et $t - a_m$ sont strictement positifs, cette inégalité est équivalente au trinôme (5) négatif, ce qui est assuré par la condition $a_m < 1/9$.
Donc sous la condition $a_m < 1/9$ le polynôme $p(\lambda)$ a exactement m racines comptées avec multiplicité dans la boule ouverte $B(0, |\lambda|)$ où

$$\varepsilon := \frac{t_1}{g_m} \leq |\lambda| < \frac{1}{2g_m}.$$

Par conséquent nous avons

$$\tilde{\sigma}_1 \geq \ldots \geq \tilde{\sigma}_{n-m} > \varepsilon \geq \tilde{\sigma}_{n-m+1} \geq \ldots \geq \tilde{\sigma}_n.$$

Prouvons maintenant l'assertion 2. De l'hypothèse nous déduisons que $s_n \neq 0$ puisque $a_1 > 1/9$. Le polynôme $s(\lambda)$ s'écrit

$$s(\lambda) = s_n + (-1)^{n-m} s_{n-m} \lambda^m + \ldots - s_{n-1} \lambda^{n-1} + \lambda^n$$

avec $s_{n-m} \neq 0$. Nous avons :

$$\begin{aligned}
\left|\frac{s(\lambda)}{s_{n-m}}\right| &\geq \frac{s_n}{s_{n-m}} - \sum_{k=m}^{n} \frac{s_{n-k}}{s_{n-m}}|\lambda|^k \\
&\geq b_m^m - |\lambda|^m \sum_{k=m}^{n} (g_m|\lambda|)^{k-m} \\
&\geq b_m^m - \frac{|\lambda|^m}{1 - g_m|\lambda|} \\
&> \frac{1}{(9g_m)^m} - \frac{10}{(9(10g_m)^m} \qquad \text{puisque } 9b_m g_m \geq 1 \text{ et pour } |\lambda| \text{ tel que } 10|\lambda|g_m < 1 \\
&> \frac{10^m - 10 \times 9^{m-1}}{9 \times (90g_m)^m} \\
&> 0.
\end{aligned}$$

Donc le polynôme n'a pas de racine dans la boule $B\left(0, \dfrac{1}{10g_m}\right)$. Nous en concluons que $\sigma_n > \frac{1}{10g_m}$. □

Nous allons préciser la notion de ε-rang que nous utiliserons dans la suite.

Définition 6. – *Soit ε un nombre positif ou nul. Une matrice M a un ε-rang égal à r_ε si ses valeurs singulières vérifient :*

$$\sigma_1 \geq \ldots \geq \sigma_{r_\varepsilon} > \varepsilon \geq \sigma_{r_\varepsilon+1} \geq \ldots \geq \sigma_n. \tag{8}$$

Observons que le ε-rang est borné supérieurement par le rang r lui-même.

Soit Σ_ε la matrice obtenue à partir de Σ en mettant les $\sigma_{r+1}, \ldots, \sigma_n$ à 0. Définissons la matrice M_ε comme $U\Sigma_\varepsilon V^*$.

Remarque 2. – *Si le rang de M est au moins r, nous savons que M_ε est la matrice de rang r la plus proche de M.*

Remarque 3. – *La définition 6 est justifiée par le théorème de Eckart-Young-Mirsky* [**11**], [**33**] *qui possède une longue histoire en théorie de l'approximation de rang faible (voir Markovsky* [**31**] *pour des développements récents).*

Une conséquence du théorème 2 est :

Théorème 3. – *Soient M une matrice de taille $s \times n$ à coefficients complexes et un réel δ vérifiant les hypothèses du lemme 1.*

1– S'il existe un entier m' ($1 \leq m' \leq n$) avec $a_{m'} < 1/9$, soit m le plus petit des entiers compris entre 1 et n avec $a_m < 1/9$. Supposons de plus que

$$\frac{\tilde{\sigma}_{n-m+1} - \delta}{1+\delta} \leq \varepsilon = \frac{3a_m + 1 - \sqrt{(3a_m+1)^2 - 16a_m}}{4g_m} < \frac{\tilde{\sigma}_{n-m} + \delta}{1-\delta}.$$

Alors la matrice M a un ε-rang $n-m$.

Rang numérique

1- Entrée : une matrice $M \in \mathbf{C}^{s\times n}$, $s \geq n$

2- Calculer une approximation des valeurs singulières de M : $\tilde{\sigma_1} \geq \ldots \geq \tilde{\sigma_n}$

3- De ces $\tilde{\sigma_i}$, calculer les quantités a_k, $k = 1 : n$ et g_k définis dans la section 4

4- S'il existe un $m' \geq 1$ tel que $a_{m'} < 1/9$, soit m le plus petit des entiers compris entre 1 et n avec $a_m < 1/9$. Définissons

5- $$\varepsilon := \frac{3a_m + 1 - \sqrt{(3a_m+1)^2 - 16a_m}}{4g_m}$$

6- Avec les notations du lemme 1 supposons de plus que $\frac{\tilde{\sigma}_{n-m+1}-\delta}{1+\delta} \leq \varepsilon < \frac{\tilde{\sigma}_{n-m}+\delta}{1-\delta}$

7- Le ε-rang de la matrice M est $n-m$, de l'assertion 1 du théorème 3

8- sinon

9- $\varepsilon < \frac{\tilde{\sigma}_n+\delta}{1-\delta}$. Le ε-rang de la matrice M est n, où $\epsilon = \frac{1}{10g_m}$ comme dans l'assertion 2 du théorème 3

10- fin si

11- Sortie : le ε-rang de la matrice M

TABLE 1.

2– Si $a_1 > 1/9$ on considère $\varepsilon = \dfrac{1}{10g_m}$ où m est l'entier défini dans l'assertion 2 du théorème 2. Si $\varepsilon < \dfrac{\tilde{\sigma}_n + \delta}{1-\delta}$ alors le ϵ-rang de la matrice M est n.

Démonstration. – C'est une conséquence directe du lemme 1 et du théorème 2. □

Théorème 4. *– L'algorithme de la table 1 calcule le ε-rang d'une matrice grâce au théorème 3.*

Remarque 4. *– En fait cet algorithme est libre de ε et nous appellerons le ε-rang ainsi calculé le rang* numérique *de la matrice.*

5. Le cadre fonctionnel

Soient $n \geq 2$, $R_\omega \geq 0$ et $\omega \in \mathbf{C}^n$. Nous considérons l'ensemble $\mathbf{A}^2(\omega, R_\omega)$ des fonctions analytiques de carré intégrable dans la boule ouverte $B(\omega, R_\omega)$. C'est un espace de Hilbert équipé du produit intérieur

$$< f, g >= \frac{c_n}{R_\omega^{2n}} \int_{B(\omega,R_\omega)} f(z)\overline{g(z)}dz,$$

où $c_n = \dfrac{n!}{\pi^n}$. Nous normalisons ce produit hermitien en divisant l'intégrale par le volume de la boule $B(\omega, R_\omega) \subset \mathbf{C}^n$.
Ensuite nous munissons $(\mathbf{A}^2(\omega, R_\omega))^s$ d'une structure hermitienne via le produit intérieur

$$< f, g >= \sum_{i=1}^{s} < f_i, g_i > .$$

Par simplicité nous noterons $||f||$ la norme associée, indifféremment dans $\mathbf{A}^2(\omega, R_\omega)$ ou dans $(\mathbf{A}^2(\omega, R_\omega))^s$. De même, nous utilisons la même notation $||.||$ pour la norme euclidienne de $\mathbf{C}$ ou $\mathbf{C}^s$.

Observons que ce cadre inclut le cas d'un système analytique obtenu en localisant un système polynomial.

5.1. Le noyau de Bergman. – Pour des références de base nous renvoyons à W. Rudin [**41**] et S. G. Krantz [**24**].

Comme pour chaque $x \in B(\omega, R_\omega)$ et chaque $f \in \mathbf{A}^2(\omega, R_\omega)$ l'application évaluation $f \mapsto f(x)$ est une fonctionelle linéaire continue $eval_x$ sur $\mathbf{A}^2$, en appliquant le théorème de représentation de Riesz, il existe un élément $h_x \in \mathbf{A}^2$ tel que

$$f(x) = eval_x(f) =< f, h_x > .$$

Posons $\nu := x \mapsto \nu_x = \dfrac{\|x - \omega\|}{R_\omega}$.

Définition 7. – *La fonction* $(z, x) \mapsto H(z, x) := \overline{h_x(z)}$ *est appelée le* noyau de Bergman *et possède la propriété reproduisante :*

$$f(x) = \frac{c_n}{R_\omega^{2n}} \int_{B(\omega, R_\omega)} f(z)\, H(z, x)\, dz, \quad \forall f \in \mathbf{A}^2(\omega, R_\omega).$$

Nous disons que le noyau de Bergman reproduit $\mathbf{A}^2(\omega, R_\omega)$ *; nous en énonçons quelques propriétés.*

5.2. Propriétés. –

Proposition 2. –

1- $H(z, x) = \dfrac{1}{(1 - \dfrac{< z - \omega, x - \omega >}{R_\omega^2})^{n+1}}$;

2- $H(x, x) = \|H(\bullet, x)\|^2 = \dfrac{1}{(1 - \nu_x^2)^{n+1}}$;

3- Pour tout $f \in \mathbf{A}^2(\omega, R_\omega)$ *nous avons*

$$|f(x)| = \frac{c_n}{R_\omega^{2n}} \left| \int_{B(\omega, R_\omega)} f(z) H(z, x) dz \right| \leq \frac{\|f\|}{(1 - \nu_x^2)^{\frac{n+1}{2}}}.$$

Démonstration. – Voir Theorem 3.1.3. page 37 dans [**41**]. □

La proposition antécédente se généralise aux dérivées d'ordre supérieur.

Proposition 3. – *Soient $k \geq 0$, $\omega \in \mathbf{C}^n$, $x \in B(\omega, R_\omega)$ et $u_i \in \mathbf{C}^n$, $i = 1 : k$. Introduisons*

$$H_k(z, x, u_1, \ldots, u_k) = \frac{(n+1)\cdots(n+k) < z-\omega, u_1 > \cdots < z-\omega, u_k >}{R_\omega^{2k}\left(1 - \dfrac{< z-\omega, x-\omega >}{R_\omega^2}\right)^k} H(z, x).$$

Nous avons

1- $D^k f(x)(u_1, \cdots, u_k) = \dfrac{c_n}{R_\omega^{2n}} \displaystyle\int_{B(\omega,R_\omega)} f(z)\, H_k(z, x, u_1, \cdots, u_k)\, dz$;

2- $\|D^k f(x)\| \leq ||f||\, \dfrac{(n+1)\cdots(n+k)}{R_\omega^k\,(1-\nu_x^2)^{\frac{n+1}{2}+k}}$.

(évidemment si $k = 0$ l'intervalle où vit i est vide, et les produits $(n+1)\cdots(n+k)$ et $< z-\omega, u_1 > \cdots < z-\omega, u_k >$ sont réduits à 1.)

Pour prouver ceci nous avons besoin du lemme suivant :

Lemme 3. –

$$\|H_k(\bullet, x, u_1, \ldots, u_n)\| \leq \frac{(n+1)\ldots(n+k)}{R_\omega^k(1-\nu_x^2)^{\frac{n+1}{2}+k}} \|u_1\| \ldots \|u_k\|.$$

Démonstration. – Nous devons calculer l'intégrale de $H_k \bar{H}_k$ sur la boule $B(\omega, R_\omega)$. Ceci se réduit à estimer

$$I_k = \frac{c_n}{R_\omega^{2n}} \int_{B(\omega,R_\omega)} \frac{1}{\left(1 - \dfrac{< z-\omega, x-\omega >}{R_\omega^2}\right)^{n+1+k} \left(1 - \dfrac{\overline{< z-\omega, x-\omega >}}{R_\omega^2}\right)^{n+1+k}} dz$$

puisque

$$\|H_k(z, x, u_1, \ldots, u_n)\| \leq \frac{(n+1)\ldots(n+k)}{R_\omega^k} \|u_1\| \ldots \|u_k\|\, I_k^{1/2}.$$

Nous avons

$$I_k = \frac{c_n}{R_\omega^{2n}} \int_{B(\omega,R_\omega)} H(z, x) \frac{1}{\left(1 - \dfrac{< z-\omega, x-\omega >}{R_\omega^2}\right)^{k} \left(1 - \dfrac{\overline{< z-\omega, x-\omega >}}{R_\omega^2}\right)^{n+1+k}} dz$$

$$= \frac{1}{(1-\nu_x^2)^{n+1+2k}}$$

en utilisant la formule du noyau de Bergman (Proposition 2) et sa propriété de reproduction appliquée à la fonction $z \mapsto \dfrac{1}{\left(1 - \dfrac{< z-\omega, x-\omega >}{R_\omega^2}\right)^{k} \left(1 - \dfrac{\overline{< z-\omega, x-\omega >}}{R_\omega^2}\right)^{n+1+k}}$.

Il s'ensuit la preuve du lemme. □

Nous démontrons maintenant la proposition 3.

Démonstration. – Nous procédons par récurrence. La proposition 2 règle le cas $k = 0$. Ensuite nous avons :

$$\begin{aligned}
D^{k+1}f(x)(u_1,\dots,u_k,u_{k+1}) &= \frac{d}{dt}D^k f(x+tu_{k+1})(u_1,\dots,u_k)\Big|_{t=0} \\
&= \frac{d}{dt}\frac{c_n}{R_\omega^{2n}}\int_{B(\omega,R_\omega)} f(z)H_k(z,x+tu_{k+1},u_1,\dots,u_k)dz\Big|_{t=0} \\
&= \frac{c_n}{R_\omega^{2n}}\int_{B(\omega,R_\omega)} f(z)\frac{H_k(z,x,u_1,\dots,u_k)(n+1+k)<z-\omega,u_{k+1}>}{R_\omega^2\left(1-\dfrac{<z-\omega,x-\omega>)}{R_\omega^2}\right)}dz \\
&= \frac{c_n}{R_\omega^{2n}}\int_{B(\omega,R_\omega)} f(z)H_{k+1}(z,x,u_1,\dots,u_{k+1})dz.
\end{aligned}$$

D'où la preuve de la première assertion. Pour la seconde nous écrivons

$$\|D^k f(x)(u_1,\dots,u_k)\| \le \|f\|\,\|H_k(\bullet,x,u_1,\dots,u_k)\|.$$

Nous concluons en utilisant le lemme 3. □

Des propositions 2 et 3 nous déduisons aisément que

Proposition 4. – *Pour tout $k \ge 0$, $x \in \mathbf{C}^n$ et $f \in (\mathbf{A}^2(\omega,R_\omega))^s$ nous avons*

$$\|D^k f(x)\| \le ||f||\,\frac{(n+1)\dots(n+k)}{R_\omega^k(1-\nu_x^2)^{\frac{n+1}{2}+k}}.$$

6. Analyse de l'application évaluation

L'application évaluation est définie par

$$eval\ :\ (f,x)\mapsto eval_x(f) = f(x)$$

de $(\mathbf{A}^2(\omega,R_\omega))^s \times B(\omega,R_\omega)$ dans $\mathbf{C}^s$.

Posons $c_0 := \sum_{k\ge 0}(1/2)^{2^k-1}$ ($\sim 1.63...$), et α_0 ($\sim 0.13...$) la première des racines positives du trinôme $(1-4u+2u^2)^2 - 2u$.

Quand la valeur $f(x)$ peut-elle être considérée comme petite ? Nous allons en donner un sens précis en calculant une valeur seuil.

Théorème 5. – *Soient $f = (f_1,\dots,f_s) \in \mathbf{A}^2(\omega,R_\omega)^s$ et $x \in B(\omega,R_\omega)$. Si*

$$c_0\,(1-\nu_x^2)^{\frac{n+1}{2}}\,\frac{\|f(x)\|}{R_\omega} + \nu_x < 1$$

et

$$\frac{(n+1)(n+2)}{2}(1-\nu_x^2)^{(n-1)/2}\left(\frac{\|f\|}{R_\omega}\frac{1}{1-\nu_x^2}+1\right)\frac{\|f(x)\|}{R_\omega} \le \alpha_0$$

alors $f(x)$ est petit dans le sens suivant : la suite de Newton définie par

$$(f^{(0)},x_0) = (f,x), \quad (f^{(k+1)},x_{k+1}) = ((f^{(k)},x_k) - D\,eval\,(f^{(k)},x_k)^\dagger\,eval\,(f^{(k)},x_k)), \quad k \ge 0,$$

en notant $D\,eval\,(f^{(k)},x_k)^\dagger$ *l'inverse généralisé de Moore-Penrose de* $D\,eval\,(f^{(k)},x_k)$, *converge quadratiquement vers un certain* $(g,y)\in(\mathbf{A}^2(\omega,R_\omega))^s\times B(\omega,R_\omega)$ *satisfaisant* $g(y)=0$. *Plus précisément nous avons*

$$(\|f-g\|+\|x-y\|^2)^{1/2}\leq c_0\,(1-\nu_x^2)^{\frac{n+1}{2}}\,\|f(x)\|.$$

Il s'ensuit immédiatement le corollaire :

Corollaire 1. – *Considérons le cas particulier* $x=\omega$ *dans le théorème 5. Si*

$$c_0\frac{\|f(x)\|}{R_\omega}<1$$

et

$$\frac{(n+1)(n+2)}{2}\left(\frac{\|f\|}{R_\omega}+1\right)\frac{\|f(x)\|}{R_\omega}\leq\alpha_0$$

alors $f(x)$ *est petit. Plus précisément il existe* $(g,y)\in(\mathbf{A}^2(x,R_\omega))^s\times B(x,R_\omega)$ *tel que* $g(y)=0$ *et*

$$(\|f-g\|+\|x-y\|^2)^{1/2}\leq c_0\|f(x)\|.$$

La suite de cette section consiste à établir le théorème 5.

6.1. Estimation des dérivées de l'application évaluation. –

Proposition 5. –

$$\|D\,eval\,(f,x)^\dagger\|\leq(1-\nu_x^2)^{\frac{n+1}{2}}.$$

Démonstration. – La dérivée de l'application évaluation est

$$D\,eval\,(f,x)(g,y)=g(x)+Df(x)y.$$

Donc $(g,y)\in\ker D\,eval\,(f,x)$ si et seulement si $g(x)+Df(x)y=0$, c'est-à-dire

$$<g_i,H(\bullet,x)>+<y,Df_i(x)^*>=0,\quad i=1:s.$$

Cette condition peut être exprimée à l'aide du produit intérieur de $(\mathbf{A}^2)^s\times\mathbf{C}^n$:

$$<g,(0,\ldots,0,H(\bullet,x),0,\ldots,0)>+<y,Df_i(x)^*>=0,\quad i=1:s.$$

Donc l'espace vectoriel $(\ker D\,eval\,(f,x))^\perp$ est engendré par l'ensemble

$$(H(\bullet,x)v,Df(x)^*v)$$

où $v\in\mathbf{C}^n$. La condition

$$D\,eval\,(f,x)(H(\bullet,x),\,Df(x)^*v)=u$$

devient

$$(H(x,x)I_s+Df(x)Df(x)^*)\,v=u.$$

La matrice $\mathcal{E}=H(x,x)I_s+Df(x)Df(x)^*$ est la somme d'une matrice diagonale positive et d'une matrice hermitienne. Appliquant le théorème de Weyl page 203 dans [**45**], les valeurs propres de la matrice $\mathcal{E}$ sont supérieures à celles de $H(x,x)I_s>0$. Donc la norme de la matrice inverse $\mathcal{E}^{-1}$ vérifie

$$\|\mathcal{E}^{-1}\|\leq\frac{1}{H(x,x)}.$$

Ceci permet de calculer $\|D\,eval\,(f,x)^\dagger\|$. En fait, soit $u, v \in \mathbf{C}^n$ tel que $\mathcal{E}v = u$. Nous avons

$$\begin{aligned}\|D\,eval\,(f,x)^\dagger u\|^2 &= \|H(\bullet,x)\|^2\,\|v\|^2 + \|Df(x)v\|^2\\ &= H(x,x)\,\|v\|^2 + \|Df(x)^*v\|^2.\end{aligned}$$

Comme la matrice $\mathcal{E}^{-1}$ est hermitienne, nous pouvons écrire

$$\begin{aligned}\|D\,eval\,(f,x)^\dagger u\|^2 &= v^*\mathcal{E}v\\ &= u^*\mathcal{E}^{-1}u\\ &\le \|\mathcal{E}^{-1}\|\,\|u\|^2.\end{aligned}$$

Finalement

$$\begin{aligned}\|D\,eval\,(f,x)^\dagger\|^2 &\le \|\mathcal{E}^{-1}\|\\ &\le \frac{1}{H(x,x)}\\ &\le (1-\nu_x^2)^{n+1}, \qquad \text{par la proposition 2 .}\end{aligned}$$

Ceci achève la preuve de la proposition. □

Proposition 6. –

$$\|D^k\,eval\,(f,x)\| \le \frac{(n+1)\dots(n+k)\,\|f\|}{R_\omega^k(1-\nu_x^2)^{\frac{n+1}{2}+k}} + \frac{k(n+1)\dots(n+k-1)}{R_\omega^{k-1}(1-\nu_x^2)^{\frac{n+1}{2}+k-1}}.$$

Démonstration. – Nous avons

$$\begin{aligned}&D^k\,eval\,(f,x)(g^{(1)},y^{(1)},\dots,g^{(k)},y^{(k)})\\ &\qquad = D^kf(x)(y^{(1)},\dots,y^{(k)}) + \sum_{j=1}^k D^{k-1}g^{(j)}(x)(y^{(1)},\dots,\widehat{y^{(j)}},\dots,y^{(k)}),\end{aligned}$$

où $\widehat{y^{(j)}}$ signifie que ce terme n'apparaît pas. Alors en utilisant la proposition 3 nous trouvons que

$$\begin{aligned}&\|D^k\,eval\,(f,x)(g^{(1)},y^{(1)},\dots,g^{(k)},y^{(k)})\|\\ &\quad\le \|D^kf(x)(y^{(1)},\dots,y^{(k)})\| + \sum_{j=1}^k \|D^{k-1}g^{(j)}(x)(y^{(1)},\dots,\widehat{y^{(j)}},\dots,y^{(k)})\|\\ &\quad\le \frac{(n+1)\dots(n+k)\,\|f\|}{R_\omega^k(1-\nu_x^2)^{\frac{n+1}{2}+k}}\|y^{(1)}\|\dots\|y^{(k)}\|\\ &\qquad + \sum_{j=1}^k \frac{(n+1)\dots(n+k-1)\,\|g^{(j)}\|}{R_\omega^{k-1}(1-\nu_x^2)^{\frac{n+1}{2}+k-1}}\|y^{(1)}\|\dots\widehat{\|y^{(j)}\|}\dots\|y^{(k)}\|.\end{aligned}$$

Nous bornons $\|y^{(j)}\|$ et $\|g^{(j)}\|$ par $\|(g^{(j)}, y^{(j)})\|$. Nous obtenons

$$\|D^k\, eval\,(f,x)(g^{(1)},y^{(1)},\dots,g^{(k)},y^{(k)})\| \leq \left(\frac{(n+1)\dots(n+k)\,\|f\|}{R_\omega^k(1-\nu_x^2)^{\frac{n+1}{2}+k}} + \frac{k(n+1)\dots(n+k-1)}{R_\omega^{k-1}(1-\nu_x^2)^{\frac{n+1}{2}+k-1}}\right) \|(g^{(1)},y^{(1)})\|\dots\|(g^{(k)},y^{(k)})\|.$$

Finalement

$$\|D^k\, eval\,(f,x)\| \leq \frac{(n+1)\dots(n+k)\,\|f\|}{R_\omega^k(1-\nu_x^2)^{\frac{n+1}{2}+k}} + \frac{k(n+1)\dots(n+k-1)}{R_\omega^{k-1}(1-\nu_x^2)^{\frac{n+1}{2}+k-1}}.$$

□

6.2. Démonstration du théorème 5. – La démonstration utilise le théorème 128 page 121 de J.-P. Dedieu, Points fixes, zéros et la méthode de Newton, Springer, 2006.

Théorème 6. – *Donnons-nous f une application analytique de $\mathbf{E}$ dans $\mathbf{F}$, deux espaces de Hilbert. Soit $x \in \mathbf{C}^n$. Nous supposons que la dérivée $Df(x)$ est surjective. En notant par $Df(x)^\dagger$ l'inverse généralisé de Moore-Penrose de $Df(x)$, nous introduisons les quantités*

1- $\beta(f,x) = \|Df(x)^\dagger f(x)\|$;

2- $\gamma(f,x) = \sup_{k\geq 2} \|\frac{1}{k!} Df(x)^\dagger D^k f(x)\|^{\frac{1}{k-1}}$;

3- $\alpha(f,x) = \beta(f,x)\gamma(f,x)$.

Rappelons que α_0 et c_0 sont les constantes introduites dans cette section.
Si $\alpha(f,x) \leq \alpha_0$ alors il existe un zéro ζ de f dans la boule $B(x_0, c_0\beta(f,x_0))$ et la suite de Newton

$$x_0 = x, \quad x_{k+1} = x_k - Df(x_k)^\dagger f(x_k), \quad k \geq 0,$$

converge quadratiquement vers ζ.

Nous sommes désormais prêts pour prouver le théorème 5.

Démonstration. – Elle consiste à vérifier la condition $\alpha(\,eval\,,(f,x)) \leq \alpha_0$. Utilisant les propositions 5 et 6, nous pouvons borner la quantité $\gamma(\,eval\,,(f,x))$. Nous obtenons

$$\gamma(\,eval\,,(f,x)) \leq \sup_{k\geq 2}\left(\frac{1}{k!}\,\|D\,eval\,(f,x)^\dagger\|\,\|D^k\,eval\,(f,x)\|\right)^{\frac{1}{k-1}}$$
$$\leq \sup_{k\geq 2}\left(\binom{n+k}{k}\frac{\|f\|}{R_\omega^k(1-\nu_x^2)^k} + \binom{n+k-1}{k-1}\frac{1}{R_\omega^{k-1}(1-\nu_x^2)^{k-1}}\right)^{\frac{1}{k-1}}.$$

Nous savons que $\binom{n+k}{k} = \frac{n+k}{k}\binom{n+k-1}{k-1}$. De plus la fonction $k \mapsto \binom{n+k}{k}^{\frac{1}{k-1}}$ est décroissante. Donc $\binom{n+k}{k}^{\frac{1}{k-1}} \leq \frac{(n+1)(n+2)}{2}$. Ainsi nous obtenons l'estimation au

point

$$\gamma(\,eval\,,(f,x)) \leq \frac{(n+1)(n+2)}{2R_\omega(1-\nu_x^2)}\left(\frac{\|f\|}{R_\omega(1-\nu_x^2)}+1\right). \tag{9}$$

De la même façon la quantité $\alpha(\,eval\,,(f,x))$ peut être bornée par

$$\begin{aligned}\alpha(\,eval\,,(f,x)) &\leq \gamma(\,eval\,,(f,x))\,\beta(\,eval\,,(f,x))\\ &\leq \gamma(\,eval\,,(f,x))\,\|D\,eval\,(f,x)^\dagger\|\,\|f(x)\|.\end{aligned}$$

En utilisant les inégalités de la proposition 5 et (9) il vient

$$\alpha(\,eval\,,(f,x)) \leq \frac{(n+1)(n+2)}{2R_\omega}\,(1-\nu_x^2)^{(n-1)/2}\left(\frac{\|f\|}{R_\omega(1-\nu_x^2)}+1\right)\|f(x)\|. \tag{10}$$

La condition

$$\frac{(n+1)(n+2)}{2R_\omega}\,(1-\nu_x^2)^{(n-1)/2}\left(\frac{\|f\|}{R_\omega(1-\nu_x^2)}+1\right)\|f(x)\| \leq \alpha_0$$

implique évidemment $\alpha(\,eval\,(f,x)) \leq \alpha_0$.
Donc le théorème 6 s'applique. La suite de Newton

$$(f^{(0)},x_0)=(f,x),\quad (f^{(k+1)},x_{k+1})=((f^{(k)},x_k)-D\,eval\,(f^{(k)},x_k)^\dagger\,eval\,(f^{(k)},x_k),\quad k\geq 0,$$

converge vers un certain $(g,y) \in B((f,x),c_0\beta(\,eval\,,(f,x))) \subset (\mathbf{A}^2(\omega,R_\omega)^s \times \mathbf{C}^n$. En d'autres termes

$$\begin{aligned}(\|f-g\|^2+\|x-y\|^2)^{\frac{1}{2}} &\leq c_0\beta(\,eval\,,(f,x))\\ &\leq c_0\|D\,eval\,(f,x)^\dagger\|\,\|f(x)\|\\ &\leq c_0\,(1-\nu_x^2)^{\frac{n+1}{2}}\|f(x)\|.\end{aligned}$$

Ceci implique que $y \in B(\omega,R_\omega)$ parce que

$$\begin{aligned}\|y-\omega\| &\leq \|y-x\|+\rho_x\\ &\leq c_0\,(1-\nu_x^2)^{\frac{n+1}{2}}\|f(x)\|+\rho_x\\ &< R_\omega. \qquad \text{par hypothèse.}\end{aligned}$$

Ceci achève la preuve. □

7. Déflation et opérateur de Newton singulier

Nous définissons dans cette section une suite de déflation en un point x_0 proche d'une racine ζ. Comme nous l'avons évoqué en section 2 celle-ci est la combinaison d'une opération de sélection et d'une opération de dénoyautage. Si $x_0=\zeta$ rappelons que nous commencons par remplacer les équations par les dérivées d'ordre la valuation moins un. Nous obtenons ainsi un système dont la jacobienne des équations est de rang plus grand que un. Puis, si ce rang est plus petit que n, nous préparons ce système en divisant les équations en deux familles. L'invariant qui préside à cette partition est le rang r de la matrice jacobienne $Df(\zeta)$. Sans perte de généralité nous pouvons supposer que les r premiers

générateurs possèdent des parties affines linéairement indépendantes. L'opération de dénoyautage consiste à ajouter aux r premières fonctions celles qui constituent le complément de Schur de $Df(x)$ associé à $D_{1:r}f_{1:r}(x)$. En section 8 nous montrons que la multiplicité de la racine du système obtenu après un cran de déflation a chuté strictement. Comme nous l'avons souligné en section 2, nous pouvons alors réitérer ce procédé.

Détaillons le procédé de déflation décrit ci-dessus quand x_0 est proche de ζ. Celui-ci repose sur les deux propriétés suivantes. Premièrement l'évaluation en x_0 d'une fonction qui s'annule en ζ est petite au sens de l'analyse effectuée à la section 6. Deuxièmement le rang numérique de $Df(x_0)$ est celui de $DF(\zeta)$ pour un ϵ déterminé par l'analyse effectuée en section 4. Précisons ces idées en commençant par introduire la notion de valuation à ε près.

Définition 8. – *Soient $\varepsilon \geq 0$, $x_0 \in \mathbf{C}^n$ et $f \in \mathbf{C}\{x - x_0\}$. Nous disons que f a une ε-valuation p en x_0 si*

1- $\forall k < p, \forall \alpha \in \mathbf{N}^n$ *tel que* $|\alpha| = k \quad et \quad \left|\frac{\partial^k f(x_0)}{\partial x^\alpha}\right| \leq \varepsilon$;

2- $\exists\, \alpha \in \mathbf{N}^n$ *tel que* $|\alpha| = p \quad et \quad \left|\frac{\partial^p f(x_0)}{\partial x^\alpha}\right| > \varepsilon$.

Si $\varepsilon = 0$ la valuation est dite exacte.

Avertissement 1. – *Les définitions qui vont suivre nécessitent de considérer les systèmes tantôt comme des listes, tantôt comme des vecteurs, tantôt comme des ensembles. Par exemple, nous notons vec$(\bullet)$ l'opérateur qui concatène les éléments non nuls d'un ensemble ou d'une matrice en un vecteur ligne.*

Nous définissons ci-dessous un opérateur de sélection.

Définition 9. – *Soient $\varepsilon \geq 0$, $x_0 \in \mathbf{C}^n$ et $f = (f_1, \ldots, f_s) \in \mathbf{C}\{x - x_0\}^s$. Nous notons p_k l'ε-valuation en x_0 de f_k, $k = 1 : s$. Soit $\Delta_k = \left\{ \dfrac{\partial^{p_k - 1} f_k(x)}{\partial x^\alpha} \ : \ |\alpha| = p_k - 1 \right\}$. Nous définissons l'opérateur de sélection S par*

$$S \ : \ f \to vec\left(\bigcup_{k=1}^{s} \Delta_k \right).$$

Si $\varepsilon = 0$ l'opération de sélection est dite exacte.

Remarque 5. – *Le calcul de $S(f)$ s'effectue par l'algorithme récursif dit de sélection en faisant Sélection$(x_0, f, \emptyset, \emptyset)$. Il est facile de voir que cet algorithme est libre de ϵ en supposant que le calcul de la norme dans $\mathbf{A}^2(x_0, R_{x_0})$ le soit. De plus le nombre d'étapes de cet algorithme est fini et $S(f) := S_f$.*

Comme nous utiliserons souvent la notion de *complément de Schur* dans la suite, nous rappelons sa définition.

Définition 10. – *Le complément de Schur d'une matrice $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ de rang $r > 0$ associé à une sous-matrice inversible A de rang r est par définition Schur$(M) := D - CA^{-1}B$.*
Si $r = 0$ nous définissons Schur$(M) := M$.

Algorithme de sélection : Sélection(x_0, f, S_f, S_1)
avec : $x_0 \in \mathbf{C}^n$, $f \in \mathbf{A}^2(x_0, R_{x_0})^s$, $S_f \in \mathbf{A}^2(x_0, R_{x_0})^s$, $S_1 \in \mathbf{A}^2(x_0, R_{x_0})^s$

1- Pour $k = 1$ à $\sharp f$
2- $\quad \eta := \dfrac{2\alpha_0}{(n+1)(n+2)(R_{x_0} + \|f_k\|)\, R_{x_0}^{n-2}}$
3- $\quad$ Si $\|f_k(x_0)\| \leq \eta$ alors $\qquad$ test justifié par le corollaire 1
4- $\quad\quad S_1 := \{f_k\}$
5- $\quad\quad$ Sélection$(x_0, \{\nabla f_k\}\backslash\{0\}, S_f, S_1)$
6- $\quad$ Sinon
7- $\quad\quad S_f := S_f \cup S_1$
8- $\quad$ fin Si
9- fin Pour

TABLE 2.

Définition 11. – *Soient $\varepsilon \geq 0$, $0 \leq r < n$ et $f = (f_1, \ldots, f_s) \in \mathbf{C}\{x - x_0\}^s$. Supposons que $D_{1:r} f_{1:r}(x_0)$ a un ε-rang égal à r. Nous définissons l'opérateur de* dénoyautage

$$K \,:\, f \mapsto (f_1, \ldots, f_r, vec(\, Schur\,(Df(x)))) \in \mathbf{C}\{x - x_0\}^{r+(n-r)\,(s-r)}.$$

Nous disons que $K(f)$ est un ε-dénoyautage de f si nous avons

$$\|K(f)(x_0)\| \leq \varepsilon. \tag{11}$$

Le dénoyautage est exact *quand $\varepsilon = 0$.*

Définition 12. – (**Suite de déflation**). *Soient $\varepsilon \geq 0$, $x_0 \in \mathbf{C}^n$ et $f = (f_1, \ldots, f_s) \in \mathbf{C}\{x - x_0\}^s$. La suite*

$$\begin{aligned} F_0 &= S(f) \\ F_{k+1} &= S(\, K(F_k)\,), \quad k \geq 0, \end{aligned}$$

est appelée suite de déflation.
La longueur de la suite de déflation *est par définition l'indice ℓ où le ε-rang de $DF_\ell(x_0)$ est égal à n, et pas avant. Nous verrons dans la section 8 que ℓ est fini. Enfin nous appelons* système déflaté *de f un système de ϵ-rang égal à n extrait de F_ℓ. On le note par $\mathsf{dfl}(f)$.*

Remarque 6. – *Par souci de simplification nous avons noté les opérateurs de sélection et de dénoyautage S et K respectivement plutôt que $S_{x_0,\epsilon}$ et $K_{x_0,\epsilon}$.*

De même nous parlerons du rang d'un système au lieu du rang de la matrice jacobienne des équations.

Remarque 7. – *Par construction, le rang de chaque système d'une suite de déflation est non nul.*

Remarque 8. – *Quand le rang numérique de la matrice jacobienne $DF(x_0)$ est nul, on peut remarquer qu'une étape de l'opération de sélection correspond à une opération de*

Suite de déflation et système déflaté

1- Entrées : $x_0 \in \mathbf{C}^n$, $f \in \mathbf{A}^2(x_0, R_{x_0})^s$
2- $F := S(f)$.
3- $\eta := \dfrac{2\alpha_0}{(n+1)(n+2)(R_{x_0} + \|F\|)R_{x_0}^{n-2}}$
4- si $\|F(x_0)\| \leq \eta$ alors test justifié par le corollaire 1
5- $r :=$ **rang numérique**$(DF(x_0))$
6- extraction d'un système de rang r
7- si $r < n$ alors
8- $F := S(K(F))$
9- aller en 3
10- sinon
11- $\mathsf{dfl}\,(f)$ un système déflaté de f de ϵ-rang égal à n.
12- fin si
13- fin si
14- Sortie : $\mathsf{dfl}\,(f)$

TABLE 3.

dénoyautage. On verra qu'en section 11, l'analyse numérique de l'agorithme de calcul d'un système déflaté est simplifiée si le rang numérique de $DF(x_0)$ est strictement positif, ce qui est le cas après une opération de sélection.

Note historique. Remarquons qu'une borne supérieure pour ℓ est constitué par l'*épaisseur au sens* de la terminologie introduite par Emsalem dans [**13**]. Nous préférons utiliser cette terminologie plutôt que le terme *profondeur "depth"* utilisé plus récemment par Mourrain, Matzaflaris dans [**30**] ou Dayton, Li, Zeng [**5**], [**4**]. ∘

Théorème 7. – *Soient $x_0 \in \mathbf{C}^n$ et $f \in \mathbf{A}^2(x_0, R_\omega)^s$. Alors l'algorithme décrit dans la table 3 prouve l'existence d'une suite de déflation où les tests de vérification des inégalités (6) and (11) sont exécutés respectivement grâce au théorème 3 et au corollaire 1.*

Définition 13. – *L'opérateur de Newton* singulier *du système initial f est défini comme l'opérateur de Newton associé au système déflaté $dfl(f)$ de ε-rang égal à n.*

Plutôt que de calculer la suite de déflation introduite dans la définition 12, il est suffisant de la tronquer. Pour ce faire nous avons besoin de la définition suivante.

Définition 14. – *Soit $p \geq 1$. Nous notons $Tr_{x_0,p}(F)$ la série tronquée à l'ordre p de la fonction analytique F au point x_0.*

Newton singulier	
1-	Entrées : $x_0 \in \mathbf{C}^n$, $f \in \mathbf{A}^2(x_0, R_{x_0})^s$
2-	$\mathsf{dfl}\,(f) =$ système déflaté de f
3-	Sortie : Si $\mathsf{dfl}\,(f) \neq \emptyset$ alors $N_{\mathsf{dfl}\,(f)}(x_0)$ sinon x_0

TABLE 4.

Nous appelons alors suite de déflation tronquée *à l'ordre p au point x_0 la suite :*

$$\begin{aligned} T_0 &= Tr_{x_0,p}(S(f)) \\ T_{k+1} &= Tr_{x_0,p-k-1}\left(S\left(K(T_k)\right)\right), \quad 0 \leq k \leq p. \end{aligned}$$

Pour définir l'opérateur de Newton singulier il est alors suffisant de connaître la longueur de la suite de déflation.

Proposition 7. – *Soit ℓ la longueur de la suite de déflation. Considérons la suite de déflation tronquée $(T_k)_{k\geq 0}$ à l'ordre $\ell+1$ au point x_0 (définition 14). Alors l'opérateur de Newton singulier associé à f est égal à l'opérateur de Newton classique associé à T_ℓ.*

Démonstration. – Comme T_0 est la série tronquée à l'ordre ℓ de F_0, par construction il est aisé de voir que pour tout $k = 0 : \ell$, T_k est la série tronquée de F_k à l'ordre $p-k$. Il s'ensuit la conclusion. □

Remarque 9. – *Le calcul de $K(F)$ nécessite d'extraire un système d'ϵ-rang r. Celui-ci est déduit d'une élimination de Gauss de la matrice jacobienne $Df(x_0)$ avec pivot total afin d'obtenir un système le mieux conditionné possible. Le nombre d'étapes de l'algorithme GECP (Gaussan Elimination with Complete Pivoting) est égal au rang numérique calculé en Table 3. Dit autrement, ceci correspond au début d'une factorisation LU et les équations ainsi distinguées (12), (voir section suivante) aux générateurs dont les parties affines sont linéairement indépendantes.*

Pour une discussion du calcul du rang par factorisation LU et les relations avec la SVD, on peut se reporter aux travaux de Pan [**37**] *et Miranian et Gu* [**32**].

8. La multiplicité chute strictement lors de la déflation

Dans cette section nous démontrons que la suite de déflation stationne après un indice fini. Commençons par le montrer dans le cas de la déflation exacte. La proposition suivante montre que la multiplicité chute strictement lors d'une opération de dénoyautage.

Théorème 8. – *Supposons que le rang de $Df(\zeta)$ soit égal à r et que*

$$Df(x) := \begin{pmatrix} A(x) & B(x) \\ C(x) & D(x) \end{pmatrix}$$

où $A(\zeta) \in \mathbf{C}^{r\times r}$ est inversible. Alors la multiplicité de ζ comme racine de $K(f)$ est strictement plus petite que la multiplicité de ζ en tant que racine de f.

Démonstration. – Si $r = 0$ alors le système $K(f)$ est formé de toutes les dérivées partielles

$$\frac{\partial f_i(x)}{\partial x_j}, \quad 1 \leq j \leq n, \quad 1 \leq i \leq s.$$

Alors la conclusion découle du lemme 4.

Si $r > 0$ le système $K(f)$ est formé par $f_1, \ldots, f_r$ et les éléments du complément de Schur $D(x) - C(x)\, A(x)^{-1}\, B(x)$. De la proposition 8, les relations entre les lignes de la matrice jacobienne sont

$$(C(x), D(x)) - C(x)A(x)^{-1}(A(x), B(x)) = 0.$$

Il est facile de voir que le système $K(F) = 0$ est analytiquement équivalent en la racine ζ au système suivant

$$\left(f_1, \ldots, f_r, Df_i(x) - \sum_{j=1}^{r} \lambda_{ij}(x) Df_j(x) = 0, \quad i = r+1 : s \right) = 0, \tag{12}$$

avec $(\lambda_{ij}(x))$ la $(s-r) \times r$-matrice $\left(C(x)A(x)^{-1}\right)$.
En appliquant le théorème des fonctions implicites, nous savons qu'il existe un isomorphisme local Φ tel que

$$x_{1:r} - \zeta_{1:r} = f_{1:r} \circ \Phi.$$

En substituant $x_{1:r} - \zeta_{1:r}$ dans $f = 0$ nous obtenons le système

$$(x_1 - \zeta_1, \ldots, x_r - \zeta_r,\ f_{r+1:s} \circ \Phi) = 0. \tag{13}$$

L'idéal engendré par $f_{r+1:s} \circ \Phi$ contient seulement les monômes $x_i - \zeta_i$, $i = r+1 : n$. D'un autre côté remarquons que la multiplicité de la racine ζ du système (13) n'a pas changé : c'est aussi la multiplicité de $\zeta_{r+1:n}$ comme racine de $f_{r+1:s} \circ \Phi$. De plus, la multiplicité de ζ comme racine du système (12) est égale à la multiplicité de $\zeta_{r+1:n}$ comme racine du système $D(f_{r+1:s} \circ \Phi)$. Nous appliquons maintenant le lemme 4 au système $f_{r+1:s} \circ \Phi$ pour en déduire que la multiplicité chute. □

Proposition 8. – *Soit* $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathbf{C}^{s \times n}$ *of rang* r *où* $A \in \mathbf{C}^{r \times r}$ *est inversible. Alors les relations entre les lignes (respectivement les colonnes) de* M *sont données par*

$$(C, D) - CA^{-1}(A, B) = 0, \quad (\textit{respectivement} \begin{pmatrix} B \\ D \end{pmatrix} - \begin{pmatrix} A \\ C \end{pmatrix} A^{-1}B = 0).$$

Démonstration. – La proposition vient de l'équivalence :
$(C, D) - CA^{-1}(A, B) = 0$ et $\begin{pmatrix} B \\ D \end{pmatrix} - \begin{pmatrix} A \\ C \end{pmatrix} A^{-1}B = 0$ si et seulement si $D - CA^{-1}B = 0$.
Comme le rang de la matrice M est égal à r, c'est équivalent à $Schur\,(M) = 0$. □

Définition 15. – *La valuation d'un système analytique* $f = (f_1, \ldots, f_s)$ *en* ζ *est le minimum des valuations des* f_i *en* ζ.

Remarque 10. – *Un générateur de* $I\mathbf{C}\{x - \zeta\}$ *de valuation minimale peut toujours être pris comme élément d'une base standard (minimale) de* I.

C'est une conséquence d'une propriété fondamentale des ordres locaux : la valuation d'une somme est toujours plus grande que le minimum de la valuation de chacun de ses termes.

Dans le cas d'un localisé d'un système polynomial, la construction d'une base standard de $I\mathbf{C}\{x-\zeta\}$ à partir d'un ensemble de générateurs peut être réalisée par l'algorithme du cône tangent de Mora, par calcul successif de S-polynômes et des réductions qui en constituent un cas particulier. La valuation ne peut que croître lors des ces opérations, ce qui empêche de réduire $S(f,g)$ par f (ou g).

Lemme 4. – *Soit* $D^1 f(x) := \left(\dfrac{\partial f_i(x)}{\partial x_j}, \quad 1 \le j \le n, \quad 1 \le i \le s\right)$. *Supposons que* ζ *soit un zéro isolé de* f *et* $D^1 f$. *Alors la multiplicité de* ζ *comme racine de* $D^1 f$ *est strictement plus petite que la multiplicité de* ζ *comme racine de* f.

Démonstration. – Prenons un des f_k, disons f_i, de valuation minimale en ζ. Cette valuation est au moins 2. C'est donc qu'il existe un indice j tel que le terme dominant $\dfrac{\partial f_i(x)}{\partial x_j}$ n'est pas dans l'idéal engendré par f. D'ou la conclusion. □

L'opérateur de sélection fait chuter la multiplicité comme l'énonce la proposition ci-dessous.

Lemme 5. – *Soit* p *la valuation de* f *en* ζ. *Considérons le système*

$$D^{p-1} f(x) \ := \ \left(\frac{\partial^{|\alpha|} f_i(x)}{\partial x^\alpha}, \, |\alpha| = p-1, \, 1 \le i \le s\right).$$

Supposons que $p \ge 2$ *et que le rang de* $D^p f(\zeta)$ *soit égal à* r. *Alors la multiplicité de* ζ *comme racine de* $D^{p-1} f(x) = 0$ *est strictement plus petite que la multiplicité de* ζ *comme racine de* f. *Plus précisément la multiplicité de la racine* ζ *chute d'au moins* p^r.

Démonstration. – Comme la valuation est $p \ge 2$ alors $f(x) = \displaystyle\sum_{k \ge p} \frac{1}{k!} D^k f(\zeta)(x-\zeta)^k$ avec $D^p f(\zeta) \neq 0$. Les monômes de $LT(f)$ sont de type $(x-\zeta)^\alpha$ avec $|\alpha| \ge p \ge 2$. Comme le rang de la dérivée de $D^{p-1} f(x)$ en ζ est $r > 0$, nous pouvons supposer sans perte de généralité que $x_1 - \zeta_1, \ldots, x_r - \zeta_r$ sont dans l'idéal $LT(D^{p-1} f(x))$, et donc par conséquent le nombre de monômes standard chute d'au moins p^r. □

Lorsque la déflation est effectuée en un point x_0 suffisament proche de ζ l'évaluation en $f(x_0)$ sera petite. La section 6 quantifie la proximité de x_0 à ζ. Le résultat ci-dessous détermine le rayon d'une boule centrée en ζ dans laquelle le rang numérique de $Df(x_0)$ est identique à celui de $Df(\zeta)$.

Proposition 9. – *Supposons que le rang de* $Df(\zeta)$ *est égal à* r *et que ses valeurs singulières vérifient*

$$\sigma_1(\zeta) \ge \ldots \ge \sigma_r(\zeta) > \sigma_{r+1}(\zeta) = \ldots = \sigma_n(\zeta) = 0.$$

Notons par

$$\bar{\gamma}(f,\zeta) = \sup_{k\geq 2}\left(\frac{\|D^k f(\zeta)\|}{k!}\right)^{\frac{1}{k-1}}.$$

Soit $0 \leq \epsilon < \min\left(2-\sqrt{2}, \dfrac{\sigma_r(\zeta)}{2}\right)$. *Pour tout* $x_0 \in B\left(\zeta, \dfrac{\epsilon}{2\bar{\gamma}(f,\zeta)}\right)$ *le* ε*-rang de* $Df(x_0)$ *est égal au rang de* $Df(\zeta)$.

Démonstration. – Notons par $\sigma_k(x_0)$ les valeurs singulières de $Df(x_0)$. Nous avons successivement pour $k = 1 : n$:

$$\begin{aligned}
|\sigma_k(x_0) - \sigma_k(\zeta)| &\leq \|Df(x_0) - Df(\zeta)\| \qquad \text{par le théorème de Weyl [46]}\\
&\leq \sum_{k\geq 2}(k-1)\frac{\|D^k f(\zeta)\|}{k!}\|x_0-\zeta\|^{k-1}\\
&< \sum_{k\geq 2}(k-1)\left(\frac{\epsilon}{2}\right)^{k-1} \qquad \text{puisque } \|x_0-\zeta\| < \frac{\epsilon}{2\bar{\gamma}(f,\zeta)}\\
&< \frac{\epsilon/2}{(1-\epsilon/2)^2}\\
&< \epsilon \qquad \text{puisque } \epsilon < 2-\sqrt{2}
\end{aligned}$$

Puisque $\sigma_r(\zeta) > 2\epsilon$ nous en déduisons que pour $k \leq r$

$$\epsilon < \sigma_r(\zeta) - \epsilon < \sigma_k(\zeta) - \epsilon < \sigma_k(x_0).$$

D'autre part nous avons pour $k > r$:

$$\sigma_k(x_0) < \sigma_k(\zeta) + \epsilon = \epsilon.$$

Finalement nous avons pour $k \leq r$ et $j > r$

$$\sigma_j(x_0) < \epsilon < \sigma_k(x_0).$$

Il s'ensuit que le ε-rang de $Df(x_0)$ est égal à r. □

9. Un nouvel α-théorème fondé sur le noyau de Bergman

Dans cette partie nous considérons comme précédemment $\omega \in \mathbf{C}^n$ et $f \in (\mathbf{A}^2(\omega, R_\omega))^s$ avec $s \geq n$. Pour un système f donné, il est équivalent d'affirmer que $Df(x)$ est de rang plein ou que $Df(x)$ est injectif. L'inverse de Moore-Penrose $Df(x)^\dagger$ prend alors la forme $(Df(x)^*Df(x))^{-1}Df(x)^*$ où $Df(x)^*$ est l'adjoint de $Df(x)$. Nous avons $Df(x)Df(x)^\dagger = \pi_{im\,Df(x)}$ et $Df(x)^\dagger Df(x) = Id$ où Id est l'identité et $\pi_{im\,Df(x)}$ la projection sur l'image de $Df(x)$. Dans ce cas, l'opérateur de Newton s'écrit, voir par exemple [9] :

$$N_f^\dagger(x) := x - Df(x)^\dagger f(x).$$

Nous introduisons les quantités

$$\beta(f,x) = \|Df(x)^\dagger f(x)\| \tag{14}$$

$$\lambda(f,x) = \frac{\|f\|}{(1-\nu_x^2)^{\frac{n+1}{2}}} \tag{15}$$

$$\kappa_x = \max\left(1, \frac{(n+1)}{R_\omega(1-\nu_x^2)}\right) \tag{16}$$

$$\mu(f,x) = \|Df(x)^\dagger\| \tag{17}$$

$$\gamma(f,x) = \max\left(1, \lambda(f,x)\,\kappa_x\,\mu(f,x)\right) \tag{18}$$

$$\alpha(f,x) = \beta(f,x)\,\kappa_x. \tag{19}$$

Nous pouvons remarquer que les quantités $\gamma(f,x)$ et $\alpha(f,x)$ sont différentes de celles introduites dans la α-théorie de Shub-Smale. Ce parti pris d'utiliser les mêmes notations est justifié respectivement par les théorèmes 9 et 11. D'une part la quantité α du théorème 9 est relative à l'existence d'une racine comme dans le classique α-théorème de [**2**] page 164 . D'autre part la quantité γ du théorème 11 est relative au rayon d'une boule de convergence quadratique de la méthode de Newton comme dans le classique γ-théorème de [**2**] page 156. Nous pouvons aussi ajouter que la reproduction des fonctions analytiques de carré intégrables par le noyau de Bergman conduit naturellement à considérer respectivement les quantités $\gamma(f,x)$ et $\alpha(f,x)$.

Théorème 9. – *(α-théorème). Soient $R_\omega > 0$, $x_0 \in B(\omega, R_\omega)$, et $f = (f_1, \ldots, f_s) \in (\mathbf{A}^2(\omega, R_\omega))^s$. Nous notons α, β, λ, γ, μ, κ pour $\alpha(f, x_0)$, etc ... respectivement définis ci-dessus.*
Supposons que

$$\alpha < 2\gamma + 1 - \sqrt{(2\gamma+1)^2 - 1}.$$

Alors pour tout $\theta > 0$ tel que $B(x_0, \theta) \subset B(\omega, R_\omega)$ et

$$\frac{\alpha + 1 - \sqrt{(\alpha+1)^2 - 4\alpha(\gamma+1)}}{2(\gamma+1)} < u := \kappa\theta < \frac{1}{\gamma+1}$$

f possède une unique racine dans la boule $B(x_0, \theta)$.

Avant de prouver ce théorème nous aurons besoin de la proposition suivante :

Proposition 10. – *Pour tout $f \in (\mathbf{A}^2(\zeta, R_\omega))^s$ nous avons*

$$\forall k \geq 0, \quad \frac{1}{k!}\|D^k f(x_0)\| \leq ||f|| \frac{(n+1)^k}{R_\omega^k \left(1-\nu_{x_0}^2\right)^{\frac{n+1}{2}+k}}.$$

Démonstration. – Il suffit de tenir compte de l'inégalité

$$\frac{(n+1)\ldots(n+k)}{k!} \leq (n+1)^k$$

dans la proposition 4. □

Nous pouvons maintenant prouver le théorème 9.

Démonstration. – L'inégalité $\alpha < 2\gamma + 1 - \sqrt{(2\gamma+1)^2 - 1}$ implique $\alpha < 1$. Donc $Df(x_0)^\dagger$ est borné et $Df(x_0)$ est injectif. Nous avons donc $Df(x_0)^\dagger f(x) = Df(x_0)^\dagger f(x_0) + g(x)$ avec

$$g(x) = x - x_0 + \sum_{k\geq 2} \frac{1}{k!} Df(x_0)^\dagger D^k f(x_0)(x - x_0)^k.$$

Nous remarquons premièrement que pour tout $x \in \mathbf{C}^n$ nous avons

$$\begin{aligned}
\|g(x)\| &\geq \|x - x_0\| - \sum_{k\geq 2} \frac{1}{k!} \|Df(x_0)^\dagger D^k f(x_0)\| \, \|x - x_0\|^k \\
&\geq \|x - x_0\| - \frac{\|f\| \, \|Df(x_0)^\dagger\|}{(1-\nu_{x_0}^2)^{\frac{n+1}{2}}} \sum_{k\geq 2} \left(\frac{(n+1)\,\|x - x_0\|}{R_\omega (1 - \nu_{x_0}^2)} \right)^k \quad \text{de la proposition 10} \\
&\geq \frac{u}{\kappa} - \frac{\gamma}{\kappa} \sum_{k\geq 2} u^k \quad \text{de la définition de } \gamma = \lambda\kappa\mu \text{ et } u = \kappa\,\|x - x_0\| \\
&\geq \frac{1}{\kappa} \left(u - \gamma \frac{u^2}{1-u} \right). \qquad (20)
\end{aligned}$$

Soit $\theta > 0$. Le théorème de Rouché énonce que les applications analytiques $Df(x_0)^\dagger f(x)$ et $g(x)$ ont le même nombre de racines, chacune d'elles comptées avec leurs multiplicités respectives, dans la boule $B(x_0, \theta)$ si l'inégalité

$$\|Df(x_0)^\dagger f(x) - g(x)\| < \|g(x)\|$$

est satisfaite pour tout $x \in \partial B(x_0, \theta)$. Tout d'abord montrons que x_0 est l'unique racine de $g(x)$ dans la boule $B\left(x_0, \dfrac{1}{\kappa(\gamma+1)}\right)$. En effet considérons y une racine de $g(x)$ distincte de x_0 dans la boule $B(\omega, R_\omega)$. Nous posons $v = \kappa\|y - x_0\|$. Si $v \geq 1$ alors $\|y - x_0\| \geq 1/\kappa > \dfrac{1}{\kappa(\gamma+1)}$. Par hypothèse nous savons que $\dfrac{1}{\kappa(\gamma+1)} > \theta$. Donc dans le cas où $v \geq 1$ nous concluons que $y \notin B(x_0, \theta)$. Sinon $v < 1$. Nous déduisons de l'inégalité (20) que

$$\|g(y)\| = 0 \geq \frac{1}{\kappa} \left(v - \frac{\gamma v^2}{1-v} \right).$$

Donc $\dfrac{1}{\gamma+1} \leq v$. Il s'ensuit que la distance entre les deux racines x_0 et y de $g(x)$ est minorée par

$$\|y - x_0\| \geq \frac{1}{\kappa(\gamma+1)} > \theta.$$

Ceci montre que x_0 est la seule racine de $g(x)$ dans la boule $B\left(x_0, \dfrac{1}{\kappa(\gamma+1)}\right)$.

Maintenant nous considérons $x \in B(\omega, R_\omega)$ tel que $\|x - x_0\| = \theta = \dfrac{u}{\kappa}$. Alors $B(x_0, \theta) \subset B(\omega, R_\omega)$. De l'inégalité (20) nous déduisons que l'inégalité

$$\beta := \|Df(x_0)^\dagger f(x_0)\| < \frac{1}{\kappa} \left(u - \frac{\gamma u^2}{1-u} \right) \qquad (21)$$

implique $\|Df(x_0)^\dagger f(x) - g(x)\| < \|g(x)\|$ sur la frontière de la boule $B(x_0,\theta)$. Puisque $\alpha = \beta\kappa$, ceci est satisfait si le numérateur

$$(\gamma+1)u^2 - (\alpha+1)u + \alpha$$

de l'expression précédente (21) est strictement négative. Alors il est facile de voir que sous la condition

$$\alpha := \beta\kappa < 2\gamma + 1 - \sqrt{(2\gamma+1)^2 - 1}$$

le trinôme $(\gamma+1)u^2 - (\alpha+1)u + \alpha$ possède deux racines égales à $\dfrac{\alpha+1 \pm \sqrt{(\alpha+1)^2 - 4\alpha(\gamma+1)}}{2(\gamma+1)}$. Donc pour tout θ tel que

$$\frac{\alpha+1-\sqrt{(\alpha+1)^2-4\alpha(\gamma+1)}}{2(\gamma+1)} < u := \kappa\theta < \frac{1}{\gamma+1}$$

nous avons $(\gamma+1)u^2 - (\alpha+1)u + \alpha < 0$. Alors l'inégalité (21) est satisfaite et le système f possède une unique racine dans la boule $B(x_0,\theta)$. Le théorème est démontré. □

10. Un nouveau γ-théorème fondé sur le noyau de Bergman

Soit $f = (f_1, \dots, f_n)$ un système analytique régulier en une de ces racines ζ. Le rayon de la boule dans lequel la suite de Newton converge quadratiquement est contrôlé par la quantité

$$\gamma(f,\zeta) = \sup_{k\geq 2} \left(\frac{1}{k!} \|Df(\zeta)^{-1} D^k f(\zeta)\| \right)^{\frac{1}{k-1}}$$

introduite par M. Shub et S. Smale. Plus précisément nous avons le résultat suivant appelé γ-théorème.

Théorème 10. – *(γ-theorem de [**2**]). Soit $f(x)$ un système analytique et ζ une racine régulière de $f(x)$. Soit $R_\zeta = \dfrac{3-\sqrt{7}}{2\gamma(f,\zeta)}$. Alors pour tout $x_0 \in B(\zeta, R_\zeta)$ la suite de Newton*

$$x_{k+1} = x_k - Df(x_k)^{-1} f(x_k), \quad k \geq 0,$$

converge quadratiquement vers ζ.

Nous donnons ici une version d'un γ-théorème qui prend en compte le noyau de Bergman pour reproduire les fonctions analytiques de carré intégrables.

Théorème 11. – *(γ-théorème). Soit ζ une racine régulière isolée d'un système analytique $f = (f_1, \dots, f_s) \in \mathbf{A}^2(\omega, R_\omega)^s$. Soit $\theta \geq 0$ tel que $B(\zeta,\theta) \subset B(\omega, R_\omega)$. Nous notons γ, μ et κ pour $\gamma(f,\zeta)$, $\mu(f,\zeta)$ et κ_ζ respectivement définis en (18), (17) et (16). Supposons que*

$$u := \kappa\,\theta < \frac{2\gamma+1-\sqrt{4\gamma^2+3\gamma}}{\gamma+1}.$$

Alors pour tout $x \in B(\zeta,\theta)$ la suite de Newton

$$x_0 = x, \quad x_{k+1} = N_f^\dagger(x_k) := x_k - Df(x_k)^\dagger f(x_k), \quad k \geq 0,$$

converge quadratiquement vers ζ. Plus précisément

$$\|x_k - \zeta| \leq \left(\frac{1}{2}\right)^{2^k-1} \|x - \zeta\|, \quad k \geq 0.$$

Démonstration. – Nous utilisons la proposition 11 ci-dessous pour montrer par récurrence le résultat. Le schéma de la preuve est classique et peut être trouvé dans [**2**] page 158. L'hypothèse $u < \dfrac{2\gamma + 1 - \sqrt{4\gamma^2 + 3\gamma}}{\gamma + 1}$ implique que $\dfrac{\gamma u}{(1+\gamma)(1-u)^2 - \gamma} \leq \dfrac{1}{2}$. C'est une condition suffisante pour la convergence quadratique de la suite de Newton avec une raison de $\dfrac{1}{2}$. □

Proposition 11. – *Avec les notations du théorème 11 nous avons :*

1– Pour tout x satisfaisant $u < 1 - \sqrt{\dfrac{\gamma}{1+\gamma}}$, $Df(x)$ est injective. De plus nous avons

$$\|Df(x)^\dagger\| \leq \frac{(1-u)^2}{(1+\gamma)\,(1-u)^2 - \gamma} \|Df(\zeta)^\dagger\|.$$

2– $\|Df(\zeta)^\dagger\|\,\|Df(x)(x-\zeta) - f(x)\| \leq \dfrac{\gamma u^2}{(1-u)^2}$;

3– $\|N_f(x) - \zeta\| \leq \dfrac{\gamma u^2}{(1+\gamma)\,(1-u)^2 - \gamma}$.

Démonstration. – La racine ζ est régulière donc le rang de $Df(\zeta)$ est plein et $Df(\zeta)^\dagger$ existe.

1– Nous écrivons

$$Df(x) - Df(\zeta) = \sum_{k \geq 2} \frac{D^k f(\zeta)}{(k)!}(x - \zeta)^{k-1}.$$

De la proposition 10, il vient

$$\frac{1}{k!}\|D^k f(\zeta)\|\,\|Df(\zeta)^\dagger\| \leq \frac{\|f\|\,\|Df(\zeta)^\dagger\|\,(n+1)^k}{R_\omega^k \left(1 - \nu_\zeta^2\right)^{\frac{n+1}{2}+k}}$$

$$\leq \lambda\mu\kappa^k = \gamma\kappa^{k-1}.$$

D'où

$$\|Df(\zeta)^\dagger(Df(x) - Df(\zeta))\| \leq \gamma \sum_{k \geq 2} k\,(\kappa\,\|x - \zeta\|)^{k-1}$$

$$\leq \gamma\left(\frac{1}{(1-u)^2} - 1\right) < 1, \qquad \text{puisque par hypothèse } u < 1 - \sqrt{\frac{\gamma}{1+\gamma}},$$

avec $u = \kappa\|x-\zeta\|$. Alors grâce au lemme de Von Neumann, voir par exemple [**22**] page 30, l'application $Id - Df(\zeta)^\dagger(Df(x) - Df(\zeta)) = -Df(\zeta)^\dagger Df(x)$ est inversible. Puisque

$$(Df(\zeta)Df(\zeta)^\dagger Df(x))^\dagger Df(\zeta)Df(\zeta)^\dagger Df(x) = Id$$

il s'ensuit que $(Df(\zeta)^\dagger Df(x))^{-1} = (Df(\zeta)Df(\zeta)^\dagger Df(x))^\dagger Df(\zeta)$. Alors $Df(\zeta)Df(\zeta)^\dagger Df(x)$ est injectif ainsi que $Df(x)$. Maintenant grâce au lemme 158 page 148 de [**9**] on peut écrire

$$|\mu(f,\zeta)\mu(f,x)^{-1} - 1| \leq ||Df(\zeta)^\dagger(Df(x) - Df(\zeta))||.$$

Donc

$$\begin{aligned}\mu(f,\zeta)\mu(f,x)^{-1} &\geq 1 - |\mu(f,\zeta)\mu(f,x)^{-1} - 1| \\ &\geq 1 - ||Df(\zeta)^\dagger(Df(x) - Df(\zeta))|| \\ &\geq 1 - \gamma\left(\frac{1}{(1-u)^2} - 1\right).\end{aligned}$$

On en déduit

$$\mu(f,x) \leq \frac{(1-u)^2}{(1+\gamma)(1-u)^2 - \gamma}\mu(f,\zeta).$$

L'assertion 1 est démontrée.

2– Puisque $f(\zeta) = 0$ nous avons $Df(x)(x-\zeta) - f(x) = \sum_{k\geq 2}(k-1)\frac{1}{k!}D^k f(\zeta)(x-\zeta)^k$.

Donc, utilisant de nouveau la proposition 10 un calcul direct conduit à

$$\begin{aligned}\|Df(x)(x-\zeta) - f(x)\| &\leq \frac{\gamma}{\mu}\sum_{k\geq 2}(k-1)\left(\kappa\|x-\zeta\|\right)^k \\ &\leq \frac{\gamma u^2}{\mu(1-u)^2}.\end{aligned}$$

Ceci prouve l'assertion 2.

3– Nous avons

$$N_f^\dagger(x) - \zeta = Df(x)^\dagger(Df(x)(x-\zeta) - f(x)).$$

Des items 1 et 2, nous déduisons le résultat.

□

La condition de convergence quadratique du γ-théorème 11 est exprimée en une racine ζ. Le corollaire 2 donne une version de ce résultat avec une condition dépendante d'un point x de la boule $B(\zeta,\theta)$. Il est basé sur le lemme ci-dessous qui établit des estimations des quantités κ_ζ, $\lambda(f,zeta)$ et $\gamma(f,\zeta)$ à l'aide des quantités κ_x, $\lambda(f,x)$ et $\gamma(f,x)$ respectivement.

Lemme 6. – *Soient $x,\zeta \in B(\omega,R_\omega)$ et $\theta \geq 0$ tels que $\|x-\zeta\| \leq \theta$. On note $r_x = \frac{\nu_x}{R_\omega(1-\nu_x^2)}$ et on suppose que $3\theta r_x < \min\left(\frac{4}{n}, \frac{-1-5\nu_x^2+\sqrt{1+22\nu_x^2+13\nu_x^4}}{2(1-\nu_x^2)}\right)$. Nous avons :*

1– $\dfrac{1}{1-\nu_\zeta^2} \leq \dfrac{1}{1-\nu_x^2}\left(1+3\theta r_x\right).$

2– $\kappa_\zeta \leq \kappa_x(1+3\theta r_x)$.

3– $\lambda(f,\zeta) \le \lambda(f,x)\left(1+\dfrac{3(n+1)\theta r_x}{2-\frac{3}{2}n\theta r_x}\right)$.

4– $\gamma(f,\zeta) \le (1+3\theta r_x))\left(1+\dfrac{3(n+1)\theta r_x}{2-\frac{3}{2}n\theta r_x}\right)\dfrac{(1-\kappa_x\theta)^2}{(1+\gamma(f,x))(1-\kappa_x\theta)^2-\gamma(f,x)}\gamma(f,x)$.

Démonstration. – 1– On écrit que $\dfrac{1}{1-\nu_\zeta^2} \le \dfrac{1}{1-(\nu_x+\theta/R_\omega)^2}$. On vérifie par un calcul direct pour que l'inégalité $\dfrac{1}{1-(\nu_x+\theta/R_\omega)^2} \le \dfrac{1}{1-\nu_x^2}\left(1+\dfrac{3\theta\nu_x}{R_\omega(1-\nu_x^2)}\right)$ ait lieu, il suffit que

$$-3(\theta/R_\omega)^2\mu_x - (5\mu_x^2+1)\theta/R_\omega + (1-\nu_x^2)\nu_x \ge 0.$$

Ceci a lieu pour $\theta/R_\omega \le \dfrac{-1-5\nu_x^2+\sqrt{1+22\nu_x^2+13\nu_x^4}}{6\nu_x}$.

2– C'est une conséquence de l'item 1.

3– Pour $0 \le v \le 1$ on a $(1+v)^{1/2} \le 1+v/2$. De plus, un simple raisonnement par récurrence montre que, pour $0 \le v \le 2/n$, on a

$$(1+v)^{n+1} \le 1+\frac{(n+1)v}{1-\frac{n}{2}v}.$$

Donc avec $v = 3\theta r_x$ et $\dfrac{3n}{4}\theta r_x < 1$ le résultat suit.

□

Corollaire 2. – *Soit ζ une racine régulière isolée d'un système analytique $f = (f_1,\dots,f_s) \in \mathbf{A}^2(\omega,R_\omega)^s$. Soient $\theta \ge 0$ et x tels que $x \in B(\zeta,\theta) \subset B(\omega,R_\omega)$. Nous notons γ, μ et κ pour $\gamma(f,x)$, $\mu(f,x)$ et κ_x respectivement définis en (18), (17) et (16). On note $r_x = \dfrac{\nu_x}{R_\omega(1-\nu_x^2)}$ et*

$$g_x = (1+3\theta r_x))\left(1+\frac{3(n+1)\theta r_x}{2-\frac{3}{2}n\theta r_x}\right)\frac{(1-\kappa_x\theta)^2}{(1+\gamma(f,x))(1-\kappa_x\theta)^2-\gamma(f,x)}\gamma(f,x).$$

Supposons que

$$3\theta r_x < \min\left(\frac{4}{n}, \frac{-1-5\nu_x^2+\sqrt{1+22\nu_x^2+13\nu_x^4}}{2(1-\nu_x^2)}\right)$$

et

$$\kappa_x(1+3\theta r_x)\,\theta < \frac{2g_x+1-\sqrt{4g_x^2+3g_x}}{g_x+1}.$$

Alors la suite de Newton

$$x_0 = x, \quad x_{k+1} = N_f^\dagger(x_k) := x_k - Df(x_k)^\dagger f(x_k), \quad k \ge 0,$$

converge quadratiquement vers ζ. Plus précisément

$$\|x_k-\zeta| \le \left(\frac{1}{2}\right)^{2^k-1}\|x-\zeta\|, \quad k \ge 0.$$

Démonstration. – Il suffit d'appliquer le lemme 6 pour majorer κ_ζ et minorer la fonction décroissante $\gamma(f,\zeta) \to \dfrac{2\gamma(f,\zeta)+1-\sqrt{4\gamma(f,\zeta)^2+3\gamma(f,\zeta)}}{\gamma(f,\zeta)+1}$. □

11. Estimation de la quantité γ du système déflaté

Nous considérons les notations introduites précédemment où $\zeta \in B(\omega, R_\omega)$ est une racine du système $F \in \mathbf{A}^2(\omega, R_\omega)^s$. Nous notons $[F]_\zeta = \sum_{k\geq 0} \frac{1}{k!}\|D^k F(\zeta)\|\,\|x-\zeta\|^k$.

Lemme 7. – *Soient* $\kappa := \kappa_\zeta$, $\lambda = \dfrac{\|F\|}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}$ *et* $F \in \mathbf{A}^2(w, R_\omega)^s$ *tel que* $F(x) = \sum_{k\geq p} \frac{1}{k!} D^k F(\zeta)(x-\zeta)^k$ *avec* $p \geq 1$. *Nous notons* $u = \kappa\|x-\zeta\|$. *Alors*

$$[F]_\zeta \leq \frac{\lambda u^p}{1-u}.$$

Démonstration. – De la majoration $\frac{1}{k!}\|D^k F(\zeta)\| \leq ||F|| \dfrac{(n+1)^k}{R_\omega^k \left(1-\nu_\zeta^2\right)^{\frac{n+1}{2}+k}} = \lambda\kappa^k$ donnée par la proposition 10, nous avons successivement :

$$\begin{aligned}[F]_\zeta &\leq \lambda u^p \sum_{k\geq 0} u^k \\ &\leq \frac{\lambda u^p}{1-u}.\end{aligned}$$

□

Lemme 8. – *Soient* $t \in [0,1[$ *et* $p \geq 1$.

$$\sum_{k\geq 0} \binom{p-1+k}{k} t^k = \frac{1}{(1-t)^p}.$$

Démonstration. – Par récurrence. C'est vrai pour $p=1$. Supposons-le au cran p. Alors

$$\sum_{k\geq 1} k\binom{p-1+k}{k} t^{k-1} = \frac{p}{(1-t)^{p+1}}$$

et

$$\sum_{k\geq 0} \frac{k+1}{p}\binom{p+k}{k+1} t^k = \frac{1}{(1-t)^{p+1}}.$$

Donc

$$\sum_{k\geq 0} \binom{p+k}{k} t^k = \frac{1}{(1-t)^{p+1}}.$$

□

Lemme 9. – *Pour tout $p \geq 1$ et $u \in [0, 2/(p+1)[$ nous avons*

$$\frac{1}{(1-u)^p} - 1 \leq \frac{p\,u}{1 - \frac{p+1}{2}u}.$$

Démonstration. – L'inégalité est vraie pour $p = 1$. Supposons-la pour p donné. Puisque $(1-pu)(1-u) = 1-(p+1)u+pu^2 \geq 1-(p+1)u$ nous avons successivement :

$$\begin{aligned}\frac{1}{(1-u)^{p+1}} - 1 &\leq \left(\frac{pu}{1-\frac{p+1}{2}u} + 1\right)\frac{1}{1-u} - 1 \\ &\leq \frac{(p+1)u(1-u/2)}{(1-\frac{p+1}{2}u)(1-u)}\end{aligned}$$

De plus, pour $u \in [0, 2/(p+1)]$ nous avons :

$$\frac{1}{1-\frac{p+2}{2}u} - \frac{1-u/2}{(1-\frac{p+1}{2}u)(1-u)} = \frac{pu^2}{4(1-\frac{p+1}{2}u)(1-\frac{p+2}{2}u)(1-u)} \geq 0.$$

Il s'ensuit :

$$\frac{1}{(1-u)^{p+1}} - 1 \leq \frac{(p+1)u}{1-\frac{p+2}{2}u}.$$

L'inégalité est vraie au cran $p+1$. Le lemme est démontré. □

Lemme 10. – *Soit $p \geq 1$. Avec les hypothèses du lemme 7 nous avons :*

$$\left[\frac{1}{(p-1)!}\left(D^{p-1}F - D^{p-1}F(\zeta)\right)\right]_\zeta \leq \lambda\kappa^{p-1}\left(\frac{1}{(1-u)^p} - 1\right) \leq \lambda\kappa^{p-1}\frac{p\,u}{1-\frac{p+1}{2}u}.$$

Démonstration. – En procédant comme dans la preuve du lemme 7 nous obtenons successivement :

$$\begin{aligned}\left[\frac{1}{p-1)!}\left(D^{p-1}F - D^{p-1}F(\zeta)\right)\right]_\zeta &\leq \sum_{k\geq 1}\frac{(p-1+k)!}{(p-1)!\,k!}\frac{\|D^{p-1+k}F(\zeta)\|}{(p-1+k)!}\|x-\zeta\|^k \\ &\leq \lambda\kappa^{p-1}\sum_{k\geq 1}\binom{p-1+k}{k}u^k \\ &\leq \lambda\kappa^{p-1}\left(\frac{1}{(1-u)^p} - 1\right) \qquad \text{(par le lemme 8)} \\ &\leq \lambda\kappa^{p-1}\frac{p\,u}{1-\frac{p+1}{2}u} \qquad \text{(par le lemme 9).}\end{aligned}$$

□

Lemme 11. – *Soient $r > 0$ et $F = (F_{1:r}, F_{r+1:s}) \in \mathbf{A}^2(\omega, R_\omega)^s$ tels que $DF(\zeta)$ soit de rang r et $D_{1:r}F_{1:r}(\zeta)$ soit inversible.*

Nous notons $\kappa := \kappa_\zeta$, $\lambda = \dfrac{\|F\|}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}$, $\mu = \|D_{1:r}F_{1:r}^{-1}(\zeta)\|$ et $\gamma = \lambda\kappa\mu$. Nous introduisons

aussi $u = \kappa\|x-\zeta\|$. *Pour tout* $x \in B(\omega, R_\omega)$ *tel que* $u < 1 - \sqrt{\dfrac{\gamma}{1+\gamma}}$ *il s'ensuit que* $D_{1:r}F_{1:r}(x)$ *est inversible. De plus nous avons la majoration :*

$$[D_{1:r}F_{1:r}^{-1} - D_{1:r}F_{1:r}(\zeta)^{-1}]_\zeta \leq \frac{\gamma\mu\, v\, u}{1-\gamma\, v\, u}$$

où $v = \dfrac{2-u}{(1-u)^2}$.

Démonstration. – Nous avons $D_{1:r}F_{1:r}(x) = D_{1:r}F_{1:r}(\zeta) + \sum_{k\geq 1} \dfrac{1}{k!} D^k\left(\left(D_{1:r}F_{1:r}\right)(\zeta)\right)(x-\zeta)^k$. Puisque $D_{1:r}F_{1:r}(\zeta)$ est inversible et que $\|D^k\left(\left(D_{1:r}F_{1:r}\right)(\zeta)\right)\| \leq \|D_{1:r}^{k+1}F_{1:r}(\zeta)\|$ nous pouvons écrire en procédant comme dans la preuve du lemme 10 :

$$\begin{aligned}
||E|| := \|D_{1:r}F_{1:r}(\zeta)^{-1}D_{1:r}F_{1:r}(x) - I|| &\leq \sum_{k\geq 1} \frac{1}{k!}\|D^k\left(\left(D_{1:r}F_{1:r}\right)(\zeta)\right)\|\,\|x-\zeta)^k\| \\
&\leq \mu \sum_{k\geq 1} \binom{k+1}{k} \frac{\|\,D^{k+1}F(w)\|}{(k+1)!}|x-\zeta\|^k \\
&\leq \lambda\kappa\mu \sum_{k\geq 1} \binom{k+1}{k} u^k \\
&\leq \gamma\left(\frac{1}{(1-u)^2} - 1\right) \\
&\leq \frac{\gamma(2-u)u}{(1-u)^2} = \gamma\, v\, u.
\end{aligned}$$

La condition $u < 1 - \sqrt{\frac{\gamma}{1+\gamma}}$ implique $\|E\| \leq \gamma v < 1$. Donc $D_{1:r}F_{1:r}(x)$ est inversible. De plus $D_{1:r}F_{1:r}(x)^{-1} = (I+E)^{-1}D_{1:r}F_{1:r}(\zeta)^{-1}$. Alors nous avons

$$D_{1:r}F_{1:r}(x)^{-1} - D_{1:r}F_{1:r}^{-1}(\zeta) = \left(\sum_{k\geq 1} E^k\right) D_{1:r}F_{1:r}(\zeta)^{-1}.$$

Finalement

$$\begin{aligned}
\|D_{1:r}F_{1:r}^{-1} - D_{1:r}F_{1:r}^{-1}(\zeta)\| &\leq \frac{\|D_{1:r}F_{1:r}(\zeta)^{-1}\|\,\|E\|}{1-\|E\|} \\
&\leq \frac{\gamma\mu\, v\, u}{1-\gamma\, v\, u}.
\end{aligned}$$

□

Proposition 12. – *Soient* $\kappa = \kappa_\zeta$, $\lambda = \dfrac{\|F\|}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}$ *et* $u = \kappa\|x-\zeta\|$. *Soit* r *le rang de* $DF(\zeta)$. *Nous supposons que* $r > 0$ *et que* $D_{1:r}F_{1:r}(\zeta)$ *est inversible. Nous notons*

$\mu = \|D_{1:r}F_{1:r}(\zeta)^{-1}\|$ *et* $\gamma = \lambda\kappa\mu$. *Alors nous avons :*

$$[K(F)]_\zeta \leq \frac{\lambda u}{1-u} + \frac{\lambda\kappa\,(1+\gamma)^2\,v\,u}{1-\gamma\,v\,u} \tag{22}$$

où $v = \dfrac{2-u}{(1-u)^2}$.

Démonstration. – Nous pouvons écrire :

$$DF(x) = \begin{pmatrix} D_{1:r}F_{1:r}(x) & D_{r+1:n}F_{1:r}(x) \\ D_{1:r}F_{r+1:m}(x) & D_{r+1:n}F_{r+1:m}(x) \end{pmatrix} := \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

Nous avons $(D - CA^{-1}B)(\zeta) = 0$. Un cran de déflation conduit à $K(F) = (F_{1:r},\, vec\,(D - CA^{-1}B))$. Nous avons :

$$\begin{aligned} D - CA^{-1}B = {} & D - D(\zeta) + (C(\zeta) - C)A^{-1}(\zeta)B(\zeta) \\ & + C(A^{-1}(\zeta) - A^{-1})B(\zeta) + CA^{-1}(B(\zeta) - B). \quad (\text{puisque } (D - CA^{-1}B)(\zeta) = 0.) \end{aligned}$$

Il s'ensuit :

$$\begin{aligned} [D - CA^{-1}B]_\zeta \leq {} & [D - D(\zeta)]_\zeta \\ & + [C - C(\zeta)]_\zeta \|A^{-1}(\zeta)\|\,\|B(\zeta)\| \\ & + [C]_\zeta\,[A^{-1} - A^{-1}(\zeta)]_\zeta\,\|B(\zeta)\| \\ & + [C]_\zeta [A^{-1}]_\zeta\,[B(\zeta) - B]_\zeta. \end{aligned}$$

Rappelons les notations : $\lambda = \dfrac{\|F\|}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}$, $\kappa = \max\left(1,\, \dfrac{n+1}{R_\omega(1-\nu_\zeta^2)}\right)$, $\mu = \|A^{-1}(\zeta)\|$ et $v = \dfrac{2-u}{(1-u)^2}$.
Nous avons les estimations successives :

$[D - D(\zeta)]_\zeta \leq [DF - DF(\zeta)]_\zeta \leq \lambda\kappa\,v\,u$, du lemme 10 avec $p = 2$,

$[A^{-1}(\zeta)(B - B(\zeta))]_\zeta$, $[(C - C(\zeta))A^{-1}(\zeta)]_\zeta \leq \lambda\kappa\mu\,v\,u = \gamma\,v\,u$, du lemme 10 avec $p = 2$

$\|B(\zeta)\| \leq \|DF(\zeta)\| \leq \lambda\kappa$, de la proposition 4 avec $k = 1$

$[C]_\zeta \leq ||DF(\zeta)|| + [DF - DF(\zeta)]_\zeta \leq \lambda\kappa(1 + v\,u)$,

$[A^{-1} - A(\zeta)^{-1}]_\zeta \leq \dfrac{\gamma\mu\,v\,u}{1-\gamma\,v\,u}$, du lemme 11

$$[A^{-1}]_\zeta \leq \|A(\zeta)^{-1}\| + [A^{-1} - A(\zeta)^{-1}]_\zeta \leq \mu + \frac{\gamma\mu\,v\,u}{1-\gamma\,v\,u} = \frac{\mu}{1-\gamma\,v\,u}.$$

Alors nous obtenons en tenant compte de ces majorations :

$$\begin{aligned}[D - CA^{-1}B]_\zeta &\leq \lambda\kappa\, v\, u + \lambda^2\,\kappa^2\,\mu v\, u + \lambda\kappa\,(1+v\,u)\frac{\gamma^2\, v\, u}{1-\gamma\, v\, u} \\ &\quad + \lambda\kappa\,(1+v\,u)\frac{\mu}{1-\gamma\, v\, u}\lambda\kappa\, v\, u \\ &\leq \frac{(1+\gamma)^2\lambda\kappa\, v\, u}{1-\gamma\, v\, u}.\end{aligned}$$

D'un autre côté, le lemme 7 avec $p = 1$ implique

$$[F_{1:r}]_\zeta \leq \frac{\lambda u}{1-u}.$$

Nous en concluons que

$$[K(F)]_\zeta \leq \frac{\lambda u}{1-u} + \frac{\lambda\kappa\,(1+\gamma)^2\, v\, u}{1-\gamma\, v\, u}.$$

□

Proposition 13. – *Soient* $\kappa = \kappa_\zeta$, $\lambda = \dfrac{\|F\|}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}$ *et* $u = \kappa\|x-\zeta\|$. *Soit* p *la valuation de* F *en* ζ. *Nous avons :*

$$[S(F)]_\zeta \leq \lambda\kappa^{p-1}\frac{p\, u}{1-\frac{p+1}{2}u}.$$

Si $u \leq \dfrac{1}{p+1}$ *alors*

$$[S(F)]_\zeta \leq \lambda\kappa^{p-1}\frac{2p}{p+1} \leq 2\lambda\kappa^{p-1}. \tag{23}$$

Démonstration. – Par construction de $S(F)$, c'est une conséquence directe du lemme 10. □

Dorénavant nous supposerons que chaque élément F_k de la suite de déflation

$$\begin{aligned}F_0 &= S(f) \\ F_{k+1} &= S(\,K(F_k)\,), \quad k \geq 0.\end{aligned}$$

est de rang r_k. On sait par la remarque 7 que $r_k > 0$. Sans perte de généralité nous pouvons dire que $D_{1:r_k}F_{k,1:r_k}$ est inversible.

Théorème 12. – *Soient* $f = (f_1, \ldots, f_s) \in \mathbf{A}^2(\omega, R_\omega)^s$ *et* ζ *une racine isolée de* f. *Nous considérons la suite de déflation de longueur* ℓ *de la définition 12 :*

$$\begin{aligned}F_0 &= S(f) \\ F_{k+1} &= S(\,K(F_k)\,), \quad k \geq 0.\end{aligned}$$

Nous notons p_0 *(respectivement,* p_k*) le maximum des valuations des équations du système* f *(respectivement,* $K(F_k)$, $k = 0 : \ell-1$*). Nous considérons* $p = \max\limits_{k=0:\ell-1} p_k$. *Soient* $\kappa = \kappa_\zeta$ *et*

$\lambda_k = \dfrac{\|F_k\|}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}$. *Soit* r_k *le rang de* $DF_k(\zeta)$ *et* $\mu = \max_{k=0:\ell} ||D_{1:r_k}F_{k,1:r_k}(\zeta)^{-1}||$. *Nous notons* $\gamma_0 = \frac{2p_0}{p_0+1}\lambda_0\kappa^{p_0}\mu$ *et* $\gamma_k = \gamma(F_k,\zeta)$ *pour* $k \geq 1$. *Nous considérons* $R > 0$ *tel que*

$$u := \kappa R \leq \min\left(\frac{1}{p+1}, \quad \frac{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}{6(\ell+\gamma_0)\left(4\kappa^p(1+\ell+\gamma_0)+(1-\nu_\zeta^2)^{\frac{n+1}{2}}\right)}\right).$$

Alors la majoration

$$\gamma_\ell \leq \ell + \gamma_0 \tag{24}$$

est vraie dans la boule $B(\zeta, R)$.

Démonstration. – La proposition 13 implique

$$\gamma(F_0,\zeta) := \lambda(F_0,\zeta)\kappa\mu \leq \frac{2p_0}{p_0+1}\lambda_0\kappa^{p_0-1}\kappa\mu := \gamma_0.$$

De la remarque 7 nous savons que $r_k \geq 1$. Nous procédons par récurrence sur k pour démontrer l'inégalité (24) qui est trivialement vraie pour $k = 0$. Supposons $\gamma_k \leq k + \gamma_0$ et montrons que $\gamma_{k+1} \leq 1 + k + \gamma_0$. Nous utilisons simultanément les propositions 12 et 13 pour écrire

$$\begin{aligned}
\gamma_{k+1} &\leq \lambda_{k+1}\kappa\mu \\
&\leq \frac{||F_{k+1}||}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}\kappa\mu \\
&\leq \frac{||S(K(F_k))||}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}\kappa\mu \\
&\leq \left(\frac{\lambda_k u}{1-u} + \frac{\lambda_k\kappa(1+\gamma_k)^2 vu}{1-\gamma_k vu}\right)\frac{2\kappa^p\,\mu}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}} \qquad \text{des inégalités (22), (23) et } v = \frac{2-u}{(1-u)^2} \\
&\leq \left(\frac{1}{1-u} + \frac{\kappa(1+\gamma_k)^2 v}{1-\gamma_k vu}\right)\frac{2\kappa^{p-1}\gamma_k u}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}} \qquad \text{puisque } \gamma_k = \max(1, \lambda_k\kappa\mu) \\
&\leq \left(\frac{1}{1-u} + \frac{\kappa(1+k+\gamma_0)^2 v}{1-(k+\gamma_0)vu}\right)\frac{2\kappa^{p-1}(k+\gamma_0)u}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}} \qquad \text{de l'hypothèse de récurrence } \gamma_k \leq k+\gamma_0 \\
&\leq \left(\frac{1}{1-u} + \frac{6\kappa(1+k+\gamma_0)^2}{(1-6(k+\gamma_0)u)}\right)\frac{2\kappa^{p-1}(k+\gamma_0)u}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}} \qquad \text{car } u \leq \frac{1}{p+1} \leq \frac{1}{2} \text{ implique } v \leq 6 \\
&\underset{\text{def}}{\leq} U + V.
\end{aligned} \tag{25}$$

Le reste de la preuve consiste à montrer que les termes U et V sont plus petits que $(1+k+\gamma_0)/2$. Ainsi nous aurons $\gamma_{k+1} \leq 1 + k + \gamma_0$. Montrons tout d'abord que

$$u \leq \frac{(1+k+\gamma_0)(1-\nu_\zeta^2)^{\frac{n+1}{2}}}{(1+k+\gamma_0)(1-\nu_\zeta^2)^{\frac{n+1}{2}} + 4\kappa^{p-1}(k+\gamma_0)} \quad \text{implique } U := \frac{1}{1-u}\frac{2\kappa^{p-1}(k+\gamma_0)u}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}} \leq \frac{1}{2}(1+k+\gamma_0).$$

Puisque $\kappa, \gamma_0, k \geq 1$, nous montrons par des majorations élémentaires que :

$$\frac{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}{6(k+\gamma_0)\left(4\kappa^p(1+k+\gamma_0)+(1-\nu_\zeta^2)^{\frac{n+1}{2}}\right)} \leq \frac{(1+k+\gamma_0)(1-\nu_\zeta^2)^{\frac{n+1}{2}}}{(1+k+\gamma_0)(1-\nu_\zeta^2)^{\frac{n+1}{2}}+4\kappa^{p-1}(k+\gamma_0)}.$$

Il s'ensuit que pour $\ell \geq k$ nous avons :

$$u \leq \frac{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}{6(\ell+\gamma_0)\left(4\kappa^p(1+\ell+\gamma_0)+(1-\nu_\zeta^2)^{\frac{n+1}{2}}\right)} \quad \text{implique } U := \frac{1}{1-u}\frac{2\kappa^{p-1}(k+\gamma_0)u}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}} \leq \frac{1}{2}(1+k+\gamma_0). \tag{26}$$

D'autre part un calcul direct établit que :

$$u \leq \frac{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}{6(\ell+\gamma_0)(4\kappa^p(1+\ell+\gamma_0)+(1-\nu_\zeta^2)^{\frac{n+1}{2}})} \quad \text{implique } V := \frac{12\kappa^p(1+k+\gamma_0)^2}{(1-6(k+\gamma_0)u)}\frac{(k+\gamma_0)u}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}} \leq \frac{1}{2}(1+k+\gamma_0) \tag{27}$$

En effet il est facile de voir que pour $u \leq \dfrac{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}{6(k+\gamma_0)(4\kappa^p(1+k+\gamma_0)+(1-\nu_\zeta^2)^{\frac{n+1}{2}})}$ nous avons

$$\frac{2V}{1+k+\gamma_0} := \frac{24\kappa^p(1+k+\gamma_0)}{(1-6(k+\gamma_0)u)}\frac{(k+\gamma_0)u}{(1-\nu_\zeta^2)^{\frac{n+1}{2}}} \leq 1.$$

En tenant compte des inégalités (26) et (27) dans (25) il s'ensuit que $\gamma_{k+1} \leq 1+k+\gamma_0$. □

12. γ-théorème et α-théorème pour un système déflaté

Nous énonçons un $\gamma-$théorème pour un système déflaté.

Théorème 13. – *(γ-théorème). Soient $f \in \mathbf{A}^2(\omega, R_\omega)^s$ et $\zeta \in B(\omega, R_\omega)$ une racine isolée de f. Soit ℓ la longueur d'une suite de déflation telle que pour tout $0 \leq k < \ell$, chaque élément de la suite $F_0 = S(f)$, $F_{k+1} = S(K(F_k))$, satisfait $F_k(\zeta) = 0$ et $r_k := \operatorname{rang}(DF_k(\zeta)) < n$. Soient p_k pour $k = 0 : \ell - 1$ et $p = \max_{k=0:\ell-1} p_k$. Nous notons $\kappa = \kappa_\zeta$, $\mu = \max_{k=0:\ell} ||D_{1:r_k}F_{k,1:r_k}(\zeta)^{-1}||$, $\gamma_0 := \dfrac{2p_0}{p_0+1}\lambda(f,\zeta)\,\kappa^{p_0-1}\,\mu$ et $\gamma_\ell := \gamma_0 + \ell$. Soit R tel que*

$$\kappa R := \min\left(\frac{1}{p+1}, \quad \frac{(1-\nu_\zeta^2)^{\frac{n+1}{2}}}{6\gamma_\ell\left(4\kappa^p(\gamma_\ell+1)+(1-\nu_\zeta^2)^{\frac{n+1}{2}}\right)}, \quad \frac{2\gamma_\ell+1-\sqrt{4\gamma_\ell^2+3\gamma_\ell}}{\gamma_\ell+1}\right).$$

Alors pour tout $x \in B(\zeta, R)$ la suite de Newton définie par la table 4,

$$x_0 = x, \quad x_{k+1} = N_{\mathit{dfl}(f)}(x_k), \quad k \geq 0,$$

converge quadratiquement vers ζ.

Démonstration. – Le théorème 12 montre que $\gamma(F_\ell, \zeta) \leq \gamma_\ell$ dans la boule $B(\zeta, R)$. Nous appliquons alors le théorème 11 au système F_ℓ avec γ_ℓ. □

Nous donnons également un résultat d'existence d'une racine singulière reposant sur le théorème 9.

Théorème 14. – *Soit $f \in \mathbf{A}^2(\omega, R_\omega)^s$ and $x_0 \in B(\omega, R_\omega)$. Supposons que :*

1– Il existe une suite de déflation $(F_k)_{0\le k\le \ell}$ de longueur ℓ en x_0. Plus précisément, si pour tout $0 \le k < \ell$ chaque élément $F_0 = S(f)$, $F_{k+1} = S(K(F_k))$ satisfait

1.1– $\|F_k(x_0)\| \le \eta_k := \dfrac{2\alpha_0}{(n+1)(n+2)(R_{x_0} + \|F_k\|)R_{x_0}^{n-2}}$;

1.2– $DF_k(x_0)$ possède un ε_k-rang numérique strictement inférieur à n où ε_k est le ε donné en ligne 5 de la table 1.

2– Le système $\{f, \mathsf{dfl}(f)\}$ satisfait les hypothèses de l'α-théorème 9 en x_0.

Alors le système $\{f, \mathsf{dfl}(f)\}$ a une seule racine ζ dans la boule $B(x_0, \theta)$ où θ est défini dans l'α-théorème 9.

Si de plus le système $\mathsf{dfl}(f)$ satisfait les hypothèses du corollaire 2, la suite de Newton associé au système $\mathsf{dfl}(f)$ et initialisé en x_0 converge quadratiquement vers la racine ζ.

Démonstration. – C'est une conséquence directe de la définition d'une suite de déflation, du théorème 9 et du corollaire 2. □

13. Exemple

Donnons un exemple afin d'illustrer les algorithmes exact et numérique, en considérant $f(x,y) = (f_1(x,y), f_2(x,y))$ avec

$$f_1(x,y) = x^3/3 + y^2x + x^2 + 2yx + y^2, \quad f_2(x,y) = x^2y - y^2x + x^2 + 2yx + y^2.$$

Le zéro $(0,0)$ est de multiplicité 6.

13.1. Calculs exacts. – Nous avons

$$Df(x,y) = \begin{pmatrix} x^2 + y^2 + 2x + 2y & 2xy + 2x + 2y \\ 2xy - y^2 + 2x + 2y & x^2 - 2xy + 2x + 2y \end{pmatrix}.$$

Le rang en $(0,0)$ de la matrice jacobienne est 0. Donc le premier cran de la suite de déflation consiste juste à remplacer chaque équation du système initial par son gradient :

$$F_0 := S(f) = (x^2 + y^2 + 2x + 2y, 2xy + 2x + 2y, 2xy - y^2 + 2x + 2y, x^2 - 2xy + 2x + 2y).$$

Les 4 lignes de la matrice jacobienne de F_0 sont :

$$\begin{pmatrix} 2x+2 & 2y+2 \\ 2y+2 & 2x+2 \\ 2y+2 & 2x-2y+2 \\ 2x-2y+2 & -2x+2 \end{pmatrix}.$$

Le rang en $(0,0)$ de la matrice $DF_0(0,0) = \begin{pmatrix} 2 & 2 \\ 2 & 2 \\ 2 & 2 \\ 2 & 2 \end{pmatrix}$ est 1.

Le complément de Schur de $DF_0(x,y)$ associé à la sous-matrice $2x+2$ est

$$Schur\,(DF_1(x,y)) = \frac{2}{x+1}\begin{pmatrix} 2x-2y+x^2-y^2 \\ 2x-3y+x^2-xy-y^2 \\ -x-x^2-xy+y^2. \end{pmatrix}$$

Nous pouvons facilement vérifier que le système $F_1 = (f_1, vec(\,Schur\,(DF_0(x,y))$ est régulier et équivalent en $(0,0)$ à f. Remarquons que le système tronqué de F_1 à l'ordre 1

$$2(x+y,\, 2x-2y,\, 2x-3y,\, -x)$$

est un système régulier équivalent en $(0,0)$ à f.

13.2. Calculs numériques. – Un code Maple reproduisant les calculs ci-dessous est téléchargeable à https://perso.math.univ-toulouse.fr/yak/curriculum-vitae/. Nous donnons le comportement de la suite de déflation.

1– Le point initial $(x_0, y_0) = (-0.0005, 0.0006)$.

2– Le système

$$f = \begin{pmatrix} 1/3\,x^3 + y^2x + x^2 + 2\,xy + y^2 \\ x^2y - y^2x + x^2 + 2\,xy + y^2 \end{pmatrix}.$$

3– La boule $B(x_0, y_0, R_{x_0,y_0}) := B(x_0, y_0, 1)$.

4– Série tronquée $Tr(f) := Tr_{(x_0,y_0),3}(f)$ du système $f(x+x_0, y+y_0)$.

$$Tr(f) = \begin{pmatrix} 0.00000000978 + 0.000201\,x + 0.000199\,y + 1.0\,x^2 + 2.0\,xy + 1.0\,y^2 + 0.333\,x^3 + y^2x \\ 0.0000000103 + 0.000201\,y + 0.000199\,x + 2.0\,xy + 1.0\,x^2 + 1.0\,y^2 + x^2y - 1.0\,y^2x \end{pmatrix}$$

5– Calcul de $F_0 = S(Tr(f))$.

$$F_0 = \begin{pmatrix} 0.00019940 + 2.0012\,x + 1.9990\,y + 2.0\,xy \\ 0.00019904 + 1.9978\,y + 2.0012\,x + 2.0\,xy - 1.0\,y^2 \\ 0.00020061 + 1.9990\,x + 2.0012\,y + 1.0\,x^2 + y^2 \\ 0.00020085 + 1.9978\,x + 2.0010\,y + x^2 - 2.0\,xy \end{pmatrix}$$

Le tableau ci-dessous donne le détail des tests effectués par l'algorithme de sélection qui conduisent à ce système.

Fonction	Évaluation en $(0,0)$	η	F_0
$Tr(f)_1$	10^{-8}	0.014	
$\partial_y Tr(f)_1$	2×10^{-4}	0.01	F_{01}
$\partial^2_{yy} Tr(f)_1$	1.99	0.0078	
$\partial^2_{xy} Tr(f)_1$	2.012	0.0078	
$\partial_x Tr(f)_1$	2×10^{-4}	0.0098	F_{02}
$Tr(f)_2$	10^{-8}	0.014	
$\partial_x Tr(f)_2$	2×10^{-4}	0.0098	F_{03}
$\partial^2_{xx} Tr(f)_2$	2.01	0.0078	
$\partial^2_{xy} Tr(f)_2$	1.99	0.0074	
$\partial_x Tr(f)_2$	2×10^{-4}	0.0098	F_{04}

6– Nous avons successivement $\|F_0\| = 2.4045$,
$\eta = \dfrac{2\alpha_0}{12(R_{x_0} + \|F_0\|)R_{x_0}^{n-2}} = 0.0064 \ > \ \|F_0(0,0)\| \ = \ 2.8 \times 10^{-4}$.

7– Jacobienne de F_0 en $(0,0)$: $DF_0(0,0) = \begin{pmatrix} 2.00012 & 1.999 \\ 2.00012 & 1.9978 \\ 1.999 & 2.00012 \\ 1.9978 & 2.0001 \end{pmatrix}$. Les valeurs singulières de cette jacobienne sont 0.0039 et 5.6562. Cette jacobienne a un $\varepsilon_0 = 0.008$-rang égal à 1.

8– Dénoyautage de F_0 en $(0,0)$:

$$K(F_0) = \begin{pmatrix} 0.00019940 + 2.0012\,x + 1.9990\,y \\ -0.0012 - 2.0\,y \\ 0.0043976 + 3.9956\,y - 3.9956\,x \\ 0.0053963 - 5.9944\,x + 3.9922\,y \end{pmatrix}.$$

9– $F_1 := S(K(F_0)) = K(F_0)$.

9– Évaluation de F_1 en $(0,0)$: $F_1(0,0) = (-0.12e-2, 0.19940e-3, 0.43976e-2, 0.53963e-2)$. Nous avons $\|F_1\| = 3.9048$ et
$\eta = \dfrac{2\alpha_0}{12(R_{x_0} + \|F_1\|)R_{x_0}^{n-2}} = 0.0044 > \|F_1(0,0)\| = 0.0012165$.

10– Matrice jacobienne de F_1 et son évaluation en $(0,0)$:

$$DF_1(x,y) = \begin{pmatrix} 0 & -2 \\ 2.0012 & 1.999 \\ -3.9956 & 3.9956 \\ -5.9944 & 3.9922 \end{pmatrix}$$

Les valeurs singulières de $DF_1(0,0)$ sont 9.2 et 3.34 et son $\varepsilon_1 = 3.34$-rang est 2.

11– Extraction d'un système régulier de F_1 en $(0,0)$.

$$dfl(f) = \begin{pmatrix} 0.00019940 + 2.0012\,x + 1.9990\,y \\ 0.0043976 + 3.9956\,y - 3.9956\,x \end{pmatrix}$$

Nous trouvons que l'itéré de $(x_0, y_0) = (-0.0005, 0.0006)$ par l'opérateur de Newton est $(1.5\times10^{-7}, -4.5\times10^{-7})$. Ceci illustre la propriété d'une convergence quadratique au premier pas de l'itération.

Celle-ci est confirmée par le comportement des itérés succesifs déterminés par l'algorithme Newton singulier.

$$[-0.0005, 0.0006]$$

$$[-1.5 \times 10^{-7}, -4.5 \times 10^{-7}]$$

$$[10^{-13}, -1.7 \times 10^{-13}]$$

$$[9.6 \times 10^{-27}, -2.6 \times 10^{-26}]$$

$$[3.5 \times 10^{-52}, -6.13 \times 10^{-52}]$$

$$[1.2 \times 10^{-103}, -3.4 \times 10^{-103}]$$

$$[5.9 \times 10^{-206}, -1.02 \times 10^{-206}]$$

13.3. Illustration des théorèmes 9 et 11. – Donnée par la table ci-dessous :

	β	κ	γ	α	$2\gamma+1-\sqrt{(2\gamma+1)^2-1}$	$\theta = \frac{\alpha+1-\sqrt{(\alpha+1)^2-4\alpha(\gamma+1)}}{2\kappa(\gamma+1)}$	$\frac{2\gamma+1-\sqrt{4\gamma^2+3\gamma}}{\kappa(\gamma+1)}$
(x_0, y_0)	0.00078	3	2.68	0.00234	0.079	0.000786	
$\zeta = (0,0)$	0	3	2.67				0.0269

Appendice A. Feuille de calcul Maple de la sous-section 13.2

Restart

```
restart:
with(LinearAlgebra):
print_f:=proc(f)
local k;
for k to nops(f) do print(f[k]);od;
end:
```

Constantes et procédure de la norme L2

```
(1-4*u+2*u^2)^2-2*u;
alpha0:=fsolve(%,u=0..1);
c0:=evalf(sum((1/2)^(2^k-1),k=0..infinity),20);
```

$$\left(2\,u^2-4\,u+1\right)^2-2\,u$$

$$\alpha_0 := 0.1307169444$$

$$c_0 := 1.63284301804378 62874$$

```
Norm_L2:=proc(f,x0,r)
local i,N,c,k; global n,x,y;
N:=0;  c:=2/Pi^n/r^(2*n);
for k to nops(f) do
   N:=N+evalf(int(f[k]^2,[y=x0[2]-sqrt(r^2-(x-x0[1])^2)..
   x0[2]+sqrt(r^2-(x-x0[1])^2),x=x0[1]-r..x0[1]+r]),20);
od:
evalf(sqrt(c*N));
end:
```

Procédure S de sélection

```
S:=proc(f,x0,r)
local eta,i; global Sf,Sf1,alpha0; global x,y;
for i to nops(f) do
  eta:=2*alpha0/12/(r+Norm_L2({f[i]},x0,r))/r;
  if eta>=evalf(abs(subs(x=x0[1],y=x0[2],f[i]))) then
     Sf1:=f[i];
     S({diff(f[i],x),diff(f[i],y)} minus {0},x0,r);
  else
     Sf:={op(Sf),Sf1};
  fi;
od;
end:
```

Procédure S de sélection détaillée

```
S_print:=proc(f,x0,r)
local eta_,ei,si,i,j,nopsSf; global Sf,Sf1,alpha0; global x,y,n;
for i to nops(f) do
  print();
  ei:=[seq(f[i][j],j=2..3)]; si:=ei[1]+ei[2];
```

```
  if si=0 then n:=i;fi;
  eta_:=2*alpha0/12/(r+Norm_L2({f[i][1]},x0,r))/r;
  if si=0  then printf("%s%g%s%v",‘f[‘,n,‘]=‘,evalf(f[i][1],5));
  else printf("%s%g%s%g%s%g%s%v",‘d??riv??e de f[‘,n,‘] ?? l’ordre
       [‘,ei[1],‘,‘,ei[2],‘]=‘,evalf(f[i][1],5)); fi;
  if eta_>=evalf(abs(subs(x=x0[1],y=x0[2],f[i][1]))) then
     print();print(‘??valuation en (0,0) =‘,evalf(abs(subs(x=x0[1],y=x0[2],f[i][1])),5),
     ‘<  eta=‘,evalf(eta_,5));
     Sf1:=f[i][1];
     S_print({[diff(Sf1,x),ei[1]+1,ei[2]],[diff(Sf1,y),ei[1],ei[2]+1]}
              minus {[0,ei[1]+1,ei[2]],[0,ei[1],ei[2]+1]},x0,r);
  else
     print();print(‘evaluation en (0,0) =‘,evalf(abs(subs(x=x0[1],y=x0[2],f[i][1])),5),
          ‘>  eta=‘,evalf(eta_,5));
     nopsSf:=nops(Sf);
     Sf:={op(Sf),Sf1};
     if nops(Sf)>nopsSf then  print();print(‘on retient la fonction ‘,evalf(Sf1,5));
     else print();print(‘la fonction ‘,evalf(Sf1,5), ‘  est d??j\‘a  retenue‘);
     fi:
  fi;
od;
end:
```

Procédure de détermination du rang numérique

```
P:=proc(s)
local i; global lambda;
expand(product(lambda-s[i],i=1..nops(s)));
end:

numerical_rank:=proc(A)
local s,p,bg,k,B,i,m,r,epsilon,e,n,beta,gama,ar,alpha0,gr; global lambda;
m,n:=Dimension(A);
B:=A;
if n<m then B:=Transpose(A);fi;
s:=SingularValues(B,output='list');
p:=P(s);
print(p);
bg:=[]:
n:=nops(s):
for k to n do
   if coeff(p,lambda,n-k)<>0 then
      e:=seq((abs(coeff(p,lambda,i)/coeff(p,lambda,k)))(1/(k-i)),i=0..k-1);
      beta:=max(%);
      gama:=1;
      if k< n then
        seq((abs(coeff(p,lambda,i)/coeff(p,lambda,k)))^(1/(i-k)),i=k+1..n);
        gama:=max(%);
      fi;
   else beta:=1;gama:=1; fi;
```

```
   bg:=[op(bg),[beta,gama]];
od;
print(bg);
r:=0;
for k to nops(bg) do
  if bg[k][1]*bg[k][2]<= 1.0/9 then
      ar:=bg[k][1]*bg[k][2];  gr:=bg[k][2];   r:=k;
  fi;
od;
if r=0 then epsilon:=s[n];
else
   epsilon:=(3*ar+1-sqrt((3*ar+1)^2-16*ar))/4/gr;
fi;
print(valeurs_singuli??res,evalf(s,5));
print(epsilon_,evalf(epsilon,5));
print(rang_num??rique,n-r);
[n-r,epsilon,s];
end:
```

Point initial, système, rayon

```
Digits:=250:
x0y0:=[-0.0005,0.0006];
n:=2:
f:=[x^3/3+y^2*x+x^2+2*x*y+y^2, x^2*y-x*y^2+x^2+2*x*y+y^2]:
print_f(%);
Rx0y0:=1.0;
```

$$x0y0 := [-0.0005, 0.0006]$$
$$f := [1/3\,x^3 + y^2x + x^2 + 2\,xy + y^2, x^2y - y^2x + x^2 + 2\,xy + y^2]$$
$$Rx0y0 := 1.0$$

Série tronquée à l'ordre 3 de f(x+x0,y+y0)

```
x0:=x0y0[1]:  y0:=x0y0[2]:
subs(x=x+x0,y=y+y0,f):
f:=[seq(mtaylor(%[i],[x,y],4),i in {1,2})]:
print_f(evalf(%,5));
```

$$0.00000000978 + 0.000201\,x + 0.000199\,y + 1.0\,x^2 + 2.0\,xy + 1.0\,y^2 + 0.333\,x^3 + y^2x$$
$$0.0000000103 + 0.000201\,y + 0.000199\,x + 2.0\,xy + 1.0\,x^2 + 1.0\,y^2 + x^2y - 1.0\,y^2x$$

Calcul de F0 :=S(f)

```
Sf:={}:
S({op(f)},[0.0,0.0],1.0):
F0:=[op(Sf)]:
print_f(evalf(%,5));
```

$$0.00019940 + 2.0012\,x + 1.9990\,y + 2.0\,xy$$
$$0.00019904 + 1.9978\,y + 2.0012\,x + 2.0\,xy - 1.0\,y^2$$
$$0.00020061 + 1.9990\,x + 2.0012\,y + 1.0\,x^2 + y^2$$
$$0.00020085 + 1.9978\,x + 2.0010\,y + x^2 - 2.0\,xy$$

Détail des calculs de F0 :=S(f)

```
Sf:={}:
g:=[seq([f[i],0,0],i=1..nops(f))]:
S_print({op(g)},[0.0,0.0],1.0):
f[1]=.97783e-8+.20061e-3*x+.19940e-3*y+.99950*x^2+2.0012*x*y+.9995*y^2+.33333*x^3+y^2*x
     \'evaluation en (0,0) = 9.7783 10^{-9}   <  eta= 0.010711
d\'eriv\'ee de f[1] \`a  l'ordre [0,1]=.19940e-3+2.0012*x+1.9990*y+2.*x*y
     \'evaluation en (0,0) = 0.00019940 <  eta= 0.0070694
d\'eriv\'ee de f[1] \`a  l'ordre [0,2]= 1.9990+2.*x
     \'evaluation en (0,0) = 1.9990 >  eta= 0.0052358
on retient la fonction 0.00019940 + 2.0012 x + 1.9990 y + 2. x y
d\'eriv\'ee de f[1] \`a  l'ordre [1,1]= 2.0012+2.*y
     \'evaluation en (0,0) = 2.0012 >  eta= 0.0052323
la fonction  0.00019940 + 2.0012 x + 1.9990 y + 2. x y  est d\'ej\`a retenue
d\'eriv\'ee de f[1] \`a  l'ordre [1,0]= .20061e-3+1.9990*x+2.0012*y+1.0000*x^2+y^2
     \'evaluation en (0,0) = 0.00020061 <  eta= 0.0068934
d\'eriv\'ee de f[1] \`a  l'ordre [2,0]= 1.9990+2.0000*x
     \'evaluation en (0,0) = 1.9990 >  eta= 0.0052358
on retient la fonction 0.00020061 + 1.9990 x + 2.0012 y + 1.0000 x^2  + y^2
d\'eriv\'ee de f[1] \`a l'ordre [2,0]= 1.9990+2.0000*x
     \'evaluation en (0,0) = 1.9990 >  eta= 0.0052358
on retient la fonction 0.00020061 + 1.9990 x + 2.0012 y + 1.0000 x^2  + y^2
d\'eriv\'ee de f[1] \`a l'ordre [1,1]= 2.0012+2.*y
     \'evaluation en (0,0) = 2.0012 >  eta= 0.0052323
la fonction  0.00020061 + 1.9990 x + 2.0012 y + 1.0000 x^2  + y^2
est d\'ej\`a retenue
f[2]= .10330e-7+.20085e-3*y+.19904e-3*x+1.9978*x*y+1.0006*x^2+1.0005*y^2+x^2*y-1.*y^2*x
     \'evaluation en (0,0) = 1.0330 10^{-8}   <  eta= 0.013775
d\'eriv\'ee de f[2] \`a l'ordre [1,0]= .19904e-3+1.9978*y+2.0012*x+2.*x*y-1.*y^2
     \'evaluation en (0,0) = 0.00019904 <  eta= 0.0098688
d\'eriv\'ee de f[2] \`a  l'ordre [2,0]= 2.0012+2.*y
     \'evaluation en (0,0) = 2.0012 >  eta= 0.0078227
on retient la fonction 0.00019904 + 1.9978 y + 2.0012 x + 2. x y - 1. y^2
```

```
d\'eriv\'ee de f[2] \‘a  l'ordre [1,1]= 1.9978+2.*x-2.*y
     \'evaluation en (0,0) = 1.9978 >  eta= 0.0073777
la fonction  0.00019904 + 1.9978 y + 2.0012 x + 2. x y - 1. y^2   est d\'ej\‘a retenue
d\'eriv\'ee de f[2] \‘a l'ordre [0,1]= .20085e-3+1.9978*x+2.0010*y+x^2-2.*x*y
     \'evaluation en (0,0) = 0.00020085 <  eta= 0.0098688
d\'eriv\'ee de f[2] \‘a l'ordre [0,2]= 2.0010-2.*x
     \'evaluation en (0,0) = 2.0010 >  eta= 0.0078231
on retient la fonction  0.00020085 + 1.9978 x + 2.0010 y + x^2  - 2. x y
d\'eriv\'ee de f[2] \‘a  l'ordre [1,1]= 1.9978+2.*x-2.*y
     \'evaluation en (0,0) = 1.9978 >  eta= 0.0073777
la fonction  0.00020085 + 1.9978 x + 2.0010 y + x^2  - 2. x y est d\'ej\‘a retenue
```

Évaluation en (0,0) et norme L2 de F0

```
eval_F0:=subs(x=0,y=0,F0): print(evalf(%,5));
NF0:=Norm(Vector(2,%),2):  print(evalf(%,5));
NL2F0:=Norm_L2(F0,[0,0],Rx0y0): print(evalf(%,5));
```

$$[0.00019940, 0.00019904, 0.00020061, 0.00020085]$$

$$0.00028174$$

$$2.4045$$

Test ||F0(0,0)|| <eta

```
eta:=2*alpha0/12/(Rx0y0+NL2F0)/Rx0y0^(n-2):
NF0<eta:  print(evalf(%,5));
```

$$0.28174e-3 < 0.63992e-2$$

Valeurs singulières et rang numérique de la jacobienne de F0 en (0,0)

```
J:=VectorCalculus[Jacobian](F0,[x,y]): print(evalf(%,5));
J0:=subs(x=0,y=0,J):  print(evalf(%,5));
numerical_rank(J0):
```

$$\begin{bmatrix} 2.0012+2\,y & 1.9990+2\,x \\ 2.0012+2\,y & 1.9978+2\,x-2.0\,y \\ 1.999+2.0\,x & 2.0012+2\,y \\ 1.9978+2\,x-2.0\,y & 2.0010-2.0\,x \end{bmatrix}$$

$$\begin{bmatrix} 2.0012 & 1.9990 \\ 2.0012 & 1.9978 \\ 1.999 & 2.0012 \\ 1.9978 & 2.0010 \end{bmatrix}$$

$$\text{valeurs singulières} = [5.6562, 0.0039667]$$

$$\epsilon = 0.0079335$$

$$\text{rang numérique} = 1$$

Dénoyautage de F0 en (0,0)

```
J(2..4,2)-J(2..4,1)*J(1,2)/J(1,1):
F1:=map(mtaylor,[F0[1],%[1],%[2],%[3]],[x,y],2):
print_f(evalf(%,5));
```

$$0.00019940 + 2.0012\,x + 1.9990\,y$$
$$-0.0012 - 2.0\,y$$
$$0.0043976 + 3.9956\,y - 3.9956\,x$$
$$0.0053963 - 5.9944\,x + 3.9922\,y$$

Calcul de F1 :=S(F1)

```
Sf:={}:
S({op(F1)},[0.0,0.0],1.0):
F1:=[op(Sf)]:
print_f(evalf(%,5));
```

$$-0.0012 - 2.0\,y$$
$$0.00019940 + 2.0012\,x + 1.9990\,y$$
$$0.0043976 + 3.9956\,y - 3.9956\,x$$
$$0.0053963 - 5.9944\,x + 3.9922\,y$$

Évaluation en (0,0) et norme L2 de F1

```
eval_F1:=subs(x=0,y=0,F1): print(evalf(%,5));
NF1:=Norm(Vector(2,%),2):  print(evalf(%,5));
NL2F1:=Norm_L2(F1,[0,0],Rx0y0): print(evalf(%,5));
```

$$[-0.0012, 0.00019940, 0.0043976, 0.0053963]$$
$$0.0012165$$
$$3.9048$$

Test ||F1(0,0)|| <eta

```
eta:=2*alpha0/12/(Rx0y0+NL2F1)/Rx0y0^(n-2):
NF1<eta:  print(evalf(%,5));
```

$$0.12165e-2 < 0.44418e-2$$

Valeurs singulières et rang numérique de la jacobienne de F1 en (0,0)

```
eta:=2*alpha0/12/(Rx0y0+NL2F0)/Rx0y0^(n-2):
NF0<eta:  print(evalf(%,5));
```

$$\begin{bmatrix} 0.0 & -2.0 \\ 2.0012 & 1.9990 \\ -3.9956 & 3.9956 \\ -5.9944 & 3.9922 \end{bmatrix}$$
$$\text{valeurs singulières} = [9.2020, 3.3353]$$
$$\epsilon = 3.3353$$
$$\text{rang numérique} = 2$$

Extraction d'un système régulier

```
R:=[seq(F1[i],i={2,3})]: print_f(evalf(%,5));
```

$$0.00019940 + 2.0012\,x + 1.9990\,y$$
$$0.0043976 + 3.9956\,y - 3.9956\,x$$

Itéré de (x0,y0) par l'opérateur de Newton singulier

```
eta:=2*alpha0/12/(Rx0y0+NL2F0)/Rx0y0^(n-2):
NF0<eta:  print(evalf(%,5));
```

$$0.28174e-3 < 0.63992e-2$$

Déflation et itérations successives

```
deflation:=proc(X0)
local f, X, x0, y0, J, F0,F1,i; global Sf,x,y;
f:=[x^3/3+y^2*x+x^2+2*x*y+y^2, x^2*y-x*y^2+x^2+2*x*y+y^2];
x0:=X0[1]:  y0:=X0[2]:
subs(x=x+x0,y=y+y0,f): #print(f);
F0:=[seq(mtaylor(%[i],[x,y],4),i in {1,2})]; #print(%);
Sf:={}:
S({op(F0)},[0.0,0.0],1.0):
F0:=[op(Sf)]:
J:=VectorCalculus[Jacobian](F0,[x,y]):
J(2..4,2)-J(2..4,1)*J(1,2)/J(1,1):
F1:=map(mtaylor,[F0[1],%[2]],[x,y],2);
op(solve(F1,[x,y])):
[rhs(%[1])+x0,rhs(%[2])+y0];
end:

Digits:=250:
X0:=[-0.0005,0.0006];
for i to 6 do
  X0:=deflation(X0): print(evalf(X0,5));
od:
```

$$[-0.0005, 0.0006]$$
$$[0.00000015231, -0.00000045263]$$
$$[1.0038 \times 10^{-13}, -1.6932 \times 10^{-13}]$$
$$[9.6859 \times 10^{-27}, -2.6681 \times 10^{-26}]$$
$$[3.5521 \times 10^{-52}, -6.1365 \times 10^{-52}]$$
$$[1.2568 \times 10^{-103}, -3.4366 \times 10^{-103}]$$
$$[5.9038 \times 10^{-206}, -1.0223 \times 10^{-205}]$$

Théorème-alpha 9

```
alpha_theorem:=proc(f,x0,rho)
local Nf,kappa,J,beta,gama,alpha,theta,nu;  global n,x,y;
nu:=Norm(Vector(2,x0))/rho;
kappa:=max(1,(n+1)/rho/(1-nu^2));
VectorCalculus[Jacobian](f,[x,y]);
J:=MatrixInverse(subs(x=0,y=0,%));
```

```
%.Vector(2,subs(x=0,y=0,f));
beta:=Norm(%,2);
Nf:=Norm_L2(f,[0,0],rho);
gama:=max(1,Nf*kappa*Norm(J,2)/(1-nu^2)^((n+1)/2));
alpha:=beta*kappa;
theta:=0;
if alpha < 2*gama+1-sqrt((2*gama+1)^2-1) then
   theta:=(alpha+1-sqrt((alpha+1)^2-4*alpha*(gama+1)))/2/kappa/(gama+1);
fi;
print(beta_,evalf(beta,5));
print(gamma_,evalf(gama,5));
print(kappa_,evalf(kappa,5));
alpha<2*gama+1-sqrt((2*gama+1)^2-1); print(test_alpha,evalf(%,5));
print(theta_,evalf(theta,5));
end:

Digits:=50:
X0:=[-0.0005,0.0006];  print(evalf(F1,5));
alpha_theorem([seq(F1[i],i in {2,3})],X0,Rx0y0):
```

$$\beta = 0.00078147$$
$$\gamma = 2.6737$$
$$\kappa = 3.0000$$
$$\text{testalpha} : 0.0023444 < 0.079267$$
$$\theta = 0.00078644$$

Théorème-gamma 11

```
gamma_theorem:=proc(f,x0,rho,r)
local Nf,NormJ,kappa,J,beta,gama,alpha,nu;  global  x,y;
nu:=Norm(Vector(2,x0))/rho;
kappa:=max(1,(n+1)/rho/(1-nu^2));
VectorCalculus[Jacobian](f,[x,y]);
if r>0 then
  J:=MatrixInverse(subs(x=0,y=0,%[1..r,1..r]));
  NormJ:=Norm(J,2);
elseyyygggxc
fi;
Nf:=Norm_L2(f,x0,rho);
gama:=max(1,Nf*kappa*NormJ/(1-nu^2)^((n+1)/2));
print(kappa_,evalf(kappa,5));
print(gamma_,evalf(gama,5));
(2*gama+1-sqrt(4*gama^2+3*gama))/kappa/(gama+1);
print(rayon,evalf(%,5));
end:

gamma_theorem([2*(x+y),4*(x-y)], [0,0],1,2);
```

$$\kappa = 3.$$
$$\gamma = 2.6761$$
$$rayon = 0.026858$$

Remerciements. Les remarques et les suggestions des rapporteuses et\ou rapporteurs ont conduit à des améliorations substantielles de ce texte. Qu'elles ou ils en soient vivement remercié(e)s !

Références

[1] Argyros, I. K., and Hilout, S. Semilocal convergence of Newton's method for singular systems with constant rank derivatives. *J. Korean Society of Mathematical Education Ser. B : Pure and Applied Mathematics 18*, 2 (2011), 97–111.

[2] Blum, L., Cucker, F., Shub, M., and Smale, S. *Complexity and Real Computation.* Springer-Verlag, New York-Berlin, 1998.

[3] Cox, D.A. and Little, J. and O'shea, D. *Using algebraic geometry*, vol. 185. Springer, 2005.

[4] Dayton, B. and Li, T.Y. and Zeng, Z. Multiple zeros of nonlinear systems. *Mathematics of Computation 80* (2011), 2143–2168.

[5] Dayton, B.H. and Zeng, Z. Computing the multiplicity structure in solving polynomial systems. In *Proceedings of the 2005 international symposium on Symbolic and algebraic computation* (2005), ACM, pp. 116–123.

[6] Decker, D.W and Kelley, C.T. Newton's method at singular points. i. *SIAM Journ.al on Numerical Analysis 17*, 1 (1980), 66–70.

[7] Decker, D.W. and Kelley, C.T. Newton's method at singular points. ii. *SIAM Journal on Numerical Analysis 17*, 3 (1980), 465–471.

[8] Decker, D.W., Keller, H.B. and Kelley, C.T. Convergence rates for Newton's method at singular points. *SIAM Journal on Numerical Analysis 20*, 2 (1983), 296–314.

[9] Dedieu, J.-P. *Points fixes, zéros et la méthode de Newton.* Springer, 2006.

[10] Dedieu, J.-P. and Shub, M. On simple double zeros and badly conditioned zeros of analytic functions of n variables. *Mathematics of computation* (2001), 319–327.

[11] Eckart, C., and Young, G. The approximation of one matrix by another of lower rank. *Psychometrika 1*, 3 (1936), 211–218.

[12] Eisenstat, S.C. and Ipsen, Ilse C.F. Relative perturbation techniques for singular value problems. *SIAM Journal on Numerical Analysis 32*, 6 (1995), 1972–1988.

[13] Emsalem, J. Géométrie des points épais. *Bull. Soc. math. France 106* (1978), 399–416.

[14] Fierro, R. D., and Hansen, P. C. UTV expansion pack : Special-purpose rank-revealing algorithms. *Numerical Algorithms 40*, 1 (2005), 47–66.

[15] Giusti, M., and Yakoubsohn, J.-C. Multiplicity hunting and approximating multiple roots of polynomial systems. *Recent Advances in Real Complexity and Computation 604* (2014), 105–128.

[16] Giusti, M. and Lecerf, G. and Salvy, B. and Yakoubsohn, J.-C. On location and approximation of clusters of zeros : Case of embedding dimension one. *Foundations of Computational Mathematics 7*, 1 (2007), 1–58.

[17] Giusti, M. and Lecerf, G. and Salvy, B. and Yakoubsohn, J.C. On location and approximation of clusters of zeros of analytic functions. *Foundations of Computational Mathematics 5*, 3 (2005), 257–311.

[18] Griewank, A. On solving nonlinear equations with simple singularities or nearly singular solutions. *SIAM review 27*, 4 (1985), 537–563.

[19] Griewank, A. and Osborne, M.R. Newton's method for singular problems when the dimension of the null space is > 1. *SIAM Journal on Numerical Analysis 18*, 1 (1981), 145–149.

[20] Griewank, A. and Osborne, M.R. Analysis of Newton's method at irregular singularities. *SIAM Journal on Numerical Analysis 20*, 4 (1983), 747–773.

[21] Hauenstein, J. D., Mourrain, B., and Szanto, A. On deflation and multiplicity structure. *Journal of Symbolic Computation 83*, (2017), 228–253.

[22] Kato, T. *Perturbation theory for linear operators*, vol. 132. Springer Science & Business Media, 2013.

[23] Kelley, C., and Suresh, R. A new acceleration method for Newton's method at singular points. *SIAM journal on numerical analysis 20*, 5 (1983), 1001–1009.

[24] Krantz, S. G. *Geometric analysis of the Bergman kernel and metric*. Springer, 2013.

[25] Lecerf, G. Quadratic Newton iteration for systems with multiplicity. *Foundations of Computational Mathematics 2*, 3 (2002), 247–293.

[26] Leykin, A., Verschelde, J., and Zhao, A. Newton's method with deflation for isolated singularities of polynomial systems. *Theoretical Computer Science 359*, 1 (2006), 111–122.

[27] Li, N., and Zhi, L. Computing isolated singular solutions of polynomial systems : case of breadth one. *SIAM Journal on Numerical Analysis 50*, 1 (2012), 354–372.

[28] Li, N., and Zhi, L. Computing the multiplicity structure of an isolated singular solution : Case of breadth one. *Journal of Symbolic Computation 47*, 6 (2012), 700–710.

[29] Li, N., and Zhi, L. Verified error bounds for isolated singular solutions of polynomial systems. *SIAM Journal on Numerical Analysis 52*, 4 (2014), 1623–1640.

[30] Mantzaflaris, A., and Mourrain, B. Deflation and certified isolation of singular zeros of polynomial systems. In *Proceedings of the 36th international symposium on Symbolic and algebraic computation* (2011), ACM, pp. 249–256.

[31] Markovsky, I. Low rank approximation : algorithms, implementation, applications. *Springer Science & Business Media*, 2011.

[32] Miranian, L. and Gu, M. Strong rank revealing LU factorizations. Linear algebra and its applications, 367, (2003), 1–16.

[33] Mirsky, L. Symmetric gauge functions and unitarily invariant norms. *The quarterly journal of mathematics 11*, 1 (1960), 50–59.

[34] Mourrain, B. Isolated points, duality and residues. *Journal of Pure and Applied Algebra 117* (1997), 469–493.

[35] Ojika, T. Modified deflation algorithm for the solution of singular problems. I. A system of nonlinear algebraic equations. *Journal of mathematical analysis and applications 123*, 1 (1987), 199–221.

[36] Ojika, T., Watanabe, S., and Mitsui, T. Deflation algorithm for the multiple roots of a system of nonlinear equations. *Journal of mathematical analysis and applications 96*, 2 (1983), 463–479.

[37] Pan, C.-T. On the existence and computation of rank-revealing LU factorizations. *Linear Algebra and its Applications*, vol. 316, 1-3, (2000), 199–222.

[38] Rall, L.B. Convergence of the Newton process to multiple solutions. *Numerische Mathematik 9*, 1 (1966), 23–37.

[39] Reddien, G. On Newton's method for singular problems. *SIAM Journal on Numerical Analysis 15*, 5 (1978), 993–996.

[40] Reddien, G. Newton's method and high order singularities. *Computers & Mathematics with Applications 5*, 2 (1979), 79–86.

[41] Rudin, W. *Function theory in the unit ball of* $\mathbf{C}^n$. Springer, 2008.

[42] Schröder, E. Ueber unendliche viele Algorithmen zur Auflösung der Gleichungen. *Mathematische Annalen 2* (1870), 317–365.

[43] Shamanskii, V. On the application of Newton's method in a singular case. *USSR Computational Mathematics and Mathematical Physics 7*, 4 (1967), 72–85.

[44] Shen, Y.-Q., and Ypma, T. J. Newton's method for singular nonlinear equations using approximate left and right nullspaces of the jacobian. *Applied numerical mathematics 54*, 2 (2005), 256–265.

[45] Stewart, G. W. and Sun, J. G. *Matrix Perturbation Theory.* Academic press New York, 1990.

[46] Weyl, H. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Mathematische Annalen 71*, 4 (1912), 441–479.

[47] Xu, X., and Li, C. Convergence criterion of Newton's method for singular systems with constant rank derivatives. *Journal of Mathematical Analysis and Applications 345*, 2 (2008), 689–701.

[48] Yakoubsohn, J.-C. On Newton's rule and Sylvester's theorems. *Journal of Pure and Applied Algebra 65*, 3 (1990), 293–309.

[49] Yamamoto, N. Newton's method for singular problems and its application to boundary value problems. *Journal of mathematics, Tokushima University 17* (1983), 27–88.

Version du 18 septembre 2019

Marc Giusti, Laboratoire LIX, Campus de l'École Polytechnique, 1 rue Honoré d'Estienne d'Orves, Bâtiment Alan Turing, CS35003, 91120 Palaiseau, France.
E-mail : Marc.Giusti@Polytechnique.fr

Jean-Claude Yakoubsohn, Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 9, France. • *E-mail :* yak@mip.ups-tlse.fr