Image numérique, GMRES et polynômes de Faber
Analyse Numérique, Analyse Fonctionnelle

Bernhard Beckermann*

11 Avril 2005

Résumé

Soit F_n le polynôme de Faber de degré n associé à l'image numérique d'un opérateur linéaire continu A sur un espace de Hilbert. Nous montrons dans un premier temps que $\|F_n(A)\| \leq 2$. Nous en déduisons ensuite, en terme d'image numérique, de nouvelles estimations d'erreur pour la méthode GMRES, méthode itérative adaptée à la résolution des systèmes linéaires non-hermitiens.

Abstract

We first show that $||F_n(A)|| \leq 2$, where A is a linear continuous operator acting in a Hilbert space, and F_n is the Faber polynomial of degree n corresponding to the numerical range of A. Then we deduce several new error bounds based on the numerical range for GMRES, an iterative method for solving non-hermitian systems of linear equations.

Mots clefs : GMRES, non-symmetric systems, error estimates, numerical range, Faber polynomials, polynomials of a matrix.
Classification AMS : 65F10, 47A12

1 Abridged english version

The numerical range (or field of values) $W(A)$ of a linear continuous operator A in a Hilbert space $\mathcal{H}$ is defined by (2.1). Given a convex compact $E \subset \mathbb{C}$, we define as usual the nth Faber polynomial $F_n = F_n^E$ to be the polynomial part of the Laurent expansion at infinity of ϕ^n, where the Riemann map ϕ maps conformally the exterior of E onto the exterior of the closed unit disk $\mathbb{D}$, and $\phi(\infty) = \infty$, $\phi'(\infty) > 0$.

In the first part of this note we show in Theorem 1 that, for any convex and compact $E \subset \mathbb{C}$, for any A with $W(A) \subset E$ and for any $n \geq 1$ there holds $||F_n^E(A)|| \leq 2$, the constant 2 being optimal. Our proof of Theorem 1 is strongly inspired by some recent deep work of Crouzeix and his collaborators [7, 2, 4, 5, 6] on the norm of functions of operators $p(A)$, and uses in particular the Carl Neumann double layer potential representation of $p(A)$ introduced by Badea, Crouzeix and B. Delyon in [2, Section 4]. Theorem 1 has been known before for the case of an ellipse E, including the limiting cases of a disk and an interval [9, 13]. It is quite instructive to compare Theorem 1 to a result of Atzmon, Eremenko & Sodin [1] who show that $\limsup ||F_n^E(A)||^{1/n} \leq 1$ if and only if $\sigma(A) \subset E$, and to a result of Toh & Trefethen [17, Theorem 1.1] who give bounds of the form $||F_n^E(A)|| \leq 2e(n+1)\mathcal{K}$ with a constant $\mathcal{K}$ of Kreiss type depending on E, and in particular $\mathcal{K} = 1$ if $W(A) \subset E$. Also, in [17, Theorem 1.2] the authors give bounds for $||F_n^E(A)||$

*Laboratoire Paul Painlevé UMR 8524 (ANO-EDP), UFR Mathématiques – M3, UST Lille, F-59655 Villeneuve d'Ascq CEDEX, France, e-mail: bbecker@math.univ-lille1.fr, téléphone 03 20 43 45 62, Fax 03 20 43 43 02. Supported in part by INTAS network NeCCA 03-51-6637, and in part by the Ministry of Science and Technology (MCYT) of Spain and the European Regional Development Fund (ERDF) through the grant BFM2001-3878-C02-02.

which are independent of n but increasing linearly with the dimension of the Hilbert space $\mathcal{H}$ (supposed to be finite dimensional).

In the second part we apply Theorem 1 to the error analysis for the GMRES method [15], a popular iterative method for solving non-hermitian systems of equations $Ax = b$. For the kth residual $r_k = b - Ax_k$ of GMRES, we establish in Theorem 2 estimates of the form

$$||r_k|| \leq c\,||r_0||\,\gamma^k,$$

with an explicit constant $c \leq 3$, and an explicitly given asymptotic convergence factor $\gamma < 1$ depending only on the numerical range of A and its distance to the origin. GMRES error estimates in terms of the numerical range of the above type have been suggested first probably by Eiermann [9] and promoted in the book of Greenbaum (see [13, Section 2 and Section 3], where one also finds a nice comparision with other types of error estimates for Krylov subspace methods). To our knowledge, an explicit value for the constant c for arbitrary A has been only given recently in [3, Corollary 2.4], where it is shown that $c \leq 101.25$. This value is a consequence of a recent result of Crouzeix [5] saying that there is a universal contant $2 \leq C \leq 33.75$ such that for each polynomial and for each linear bounded operator A we have the bound (2.2). Both upper bounds for c and C have been considered as quite pessimistic, indeed the conjecture of Crouzeix says that C in (2.2) can be replaced by 2. Our proof of Theorem 2 via Theorem 1 is selfcontained and does not require the knowledge of the optimal constant in (2.2).

We finally give in Corollary 3 an estimate for GMRES which requires only the knowledge of the numerical radius of A and the distance of $W(A)$ to the origin. Estimates of this type have been given by Elman [12] (see also [11]), and recently in [3, Theorem 2.1], where a smaller asymptotic convergence factor was found. This latter result will be slightly improved in Corollary 3.

2 Introduction et résultats

Soit A un opérateur linéaire continu sur un espace de Hilbert complexe $\mathcal{H}$, muni de la norme $||\cdot||$ induite par le produit scalaire de $\mathcal{H}$. L'image numérique de A, définie par

$$W(A) = \{(Ay, y) : y \in \mathcal{H}, ||y|| = 1\}, \tag{2.1}$$

est convexe par le théorème de Toeplitz et Hausdorff, et contient le spectre $\sigma(A)$. Si A est normal, nous avons l'inégalité $||p(A)|| \leq ||p||_{L_\infty(\overline{W(A)})}$ pour tout polynôme p. Récemment, Michel Crouzeix [5] a montré qu'il existe une constante universelle $2 \leq C \leq 33.75$ telle que, pour tout polynôme p et pour tout opérateur linéaire borné A,

$$||p(A)|| \leq C\,||p||_{L_\infty(\overline{W(A)})}. \tag{2.2}$$

Il a aussi émis la conjecture que la constante optimale dans (2.2) est $C = 2$. On trouvera dans [7, 2, 4, 5, 6] des majorations plus précises pour des sous-classes d'opérateurs à image numérique contenue dans un fermé E donné, ainsi que des applications diverses (probabilités, edp, algèbre linéaire numérique,...).

Le but de cette note est d'expliciter une inégalité analogue à (2.2) pour certains polynômes particuliers, dits de Faber, au théorème 1 ci-dessous, et d'exploiter ce résultat pour en déduire au théorème 2 une nouvelle estimation d'erreur pour la méthode GMRES [15], méthode itérative très populaire pour la résolution des systèmes d'équations linéaires $Ax = b$, où A est une matrice carrée pas forcement normale, souvent creuse et de très grande taille. Etant donné un convexe compact $E \subset \mathbb{C}$, on considère l'application conforme de Riemann $\phi = \phi_E$ qui envoie le complémentaire de E sur le complémentaire du disque unité fermé $\mathbb{D}$, et qui vérifie $\phi(\infty) = \infty$ et $\phi'(\infty) > 0$. Le $n$ième polynôme de Faber $F_n = F_n^E$ associé à E est donné par la partie polynômiale de la série de Laurent de ϕ^n au voisinage de l'infini.

Pour un opérateur A linéaire continu, Atzmon, Eremenko et Sodin [1] ont montré que

$$\limsup_{n\to\infty} ||F_n^E(A)||^{1/n} \leq 1 \quad \Longleftrightarrow \quad \sigma(A) \subset E. \tag{2.3}$$

Pour un disque $E = r\mathbb{D}$ nous avons $F_n^E(z) = (z/r)^n$, et donc (2.3) est une généralisation naturelle de la formule de Gelfand sur le rayon spectral. Toh et Trefethen [17, Theorem 1.1] ont donné la borne $||F_n^E(A)|| \leq 2e(n+1)\mathcal{K}$ avec une constante $\mathcal{K}$ de type Kreiss qui dépend de E, en particulier $\mathcal{K} = 1$ si $W(A) \subset E$. Les auteurs donnent dans [17, Theorem 1.2] une autre borne pour $||F_n^E(A)||$ qui est indépendante de n mais qui croît linéairement avec la dimension de l'espace $\mathcal{H}$. Dans le cas $W(A) \subset E$, nous pouvons être plus précis.

Théorème 1 *Soit E un convexe compact et $W(A) \subset E$, alors, pour tout $n \geq 1$,*

$$||F_n^E(A)|| \leq 2. \tag{2.4}$$

La preuve de ce théorème, donnée en §3, est basée sur des techniques de Badea, Crouzeix et Delyon exposées dans [2, Section 4]. Notons que le théorème 1 est connu dans le cas d'un intervalle. Par exemple, si $E = [-1, 1]$ alors A est forcement auto-adjoint, et $F_n^E = 2T_n$, avec T_n le $n$ième polynôme de Chebyshev qui est borné par 1 sur E : par conséquent, (2.4) découle de la théorie spectrale. Plus généralement, le théorème 1 a été montré (implicitement) pour une ellipse E par Eiermann [9]. Notons l'exemple

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad ||A|| = 1, \qquad W(A) = E = \{|z| \leq 1/2\}$$

avec $F_1^E(z) = 2z$ qui montre que la quantité 2 dans (2.4) ne peut pas être remplacée par une constante plus petite. Pour comparer (2.4) avec (2.2), notons également que $||F_n^E||_{L_\infty(E)} \leq 2$ par convexité de E, voir Kővari et Pommerenke [14, Theorem 2].

Notre théorème 1 admet plusieurs applications dans l'estimation d'erreur des méthodes de Krylov en algèbre linéaire numérique. Nous allons discuter en détail de la méthode GMRES pour la résolution des systèmes d'équations linéaires $Ax = b$, où A est une matrice carrée pas forcément normale. La méthode GMRES peut être caractérisée par le fait que le résidu du $k$ième itéré, $r_k = b - Ax_k \in \mathrm{span}\{r_0, Ar_0, ..., A^{k-1}r_0\}$, vérifie la propriété

$$\frac{||r_k||}{||r_0||} = \min\{\frac{||p(A)r_0||}{|p(0)|\,||r_0||} : \quad p \text{ polynôme de degré au plus } k\}, \tag{2.5}$$

et donc $||r_k||$ est décroissant en k. Partant de l'expression obtenue en majorant $||p(A)r_0||$ dans (2.5) par $||p(A)||\,||r_0||$, une expression qui est indépendante du résidu initial r_0 et du second membre b, il existe un nombre important d'estimations pour la convergence de GMRES, voir [16, Chapter 6.11] ou [13, Chapter 3]. Ces bornes sont formulées en terme du conditionnement (potentiellement élevé) de la matrice des vecteurs propres [16], du pseudo-spectre de A, des angles entre sous-espaces de Krylov [10] ou de l'image numérique de A [9, 13], voir aussi [3, Corollary 2.4] pour une version plus explicite basée sur (2.2). Nous proposons ici le résultat suivant.

Théorème 2 *Soit E un convexe compact tel que $0 \notin E$ et $W(A) \subset E$. Nous avons alors, pour le $k$ième résidu relatif de GMRES, les estimations suivantes*

$$\frac{||r_k||}{||r_0||} \leq \frac{2}{1-\gamma_E^{k+1}}\,\gamma_E^k, \qquad \frac{||r_k||}{||r_0||} \leq (2+\gamma_E)\,\gamma_E^k, \tag{2.6}$$

où $\gamma_E := 1/|\phi_E(0)| < 1$.

L'estimation d'erreur du théorème 2 est typiquement intéressante pour un nombre d'itérations k pas trop élevé car elle ne permet pas de décrire un comportement super-linéaire de la convergence. Pour une matrice symétrique définie positive A, le choix $E = [\lambda_{\min}(A), \lambda_{\max}(A)]$ au théorème 2 donne lieu à $\gamma_E = (\sqrt{\kappa}-1)/(\sqrt{\kappa}+1)$, avec $\kappa = \lambda_{\max}(A)/\lambda_{\max}(A)$ le conditionnement de A ; on obtient ainsi le même taux de convergence que dans l'estimation classique du gradient conjugué (qui est aussi valable pour le résidu relatif dans GMRES). Le cas particulier d'un disque E est discuté dans [13, Section 3] ; pour une ellipse E, une combinaison de [13, Section 3] et [9] donne un résultat similaire au théorème 2.

Pour une matrice particulière A avec $0 \notin W(A)$, on peut expliciter numériquement la constante γ_E par exemple en déterminant un polygone convexe E contenant $W(A)$ (il suffit de calculer les valeurs propres extrêmes de $\mathrm{Re}(e^{it}A)$ pour certaines valeurs réelles de t), et en utilisant la boîte à outils Schwartz-Christoffel [8] pour le calcul approché de $\phi_E(0)$.

Nous proposons finalement comme conséquence du théorème 2 une borne qui est plus facilement utilisable car elle ne requiert que très peu d'informations sur la matrice A.

Corollaire 3 *Soit* $0 \notin W(A)$. *Nous avons l'estimation suivante pour le kième résidu relatif de GMRES*

$$\frac{||r_k||}{||r_0||} \leq (2+\gamma)\,\gamma^k, \qquad \gamma := 2\sin\left(\frac{\beta}{4-2\beta/\pi}\right) < \sin(\beta), \tag{2.7}$$

ou l'angle $\beta \in (0, \pi/2)$ *est défini par* $\cos(\beta) = \mathrm{dist}(0, W(A))/\nu(A)$ *et* $\nu(A)$ *désigne le rayon numérique* $\nu(A) = \max\{|z| : z \in W(A)\}$.

L'inégalité (2.7) avec un facteur supplémentaire $2 + 3/\sqrt{2}$ et un angle généralement plus grand défini par $\cos(\beta') = \mathrm{dist}(0, W(A))/||A||$ a été démontrée récemment dans [3, Theorem 1]. Elle généralise l'inégalité d'Elman [11, 12] qui dit que $||r_k|| \leq \sin^k(\beta')\,||r_0||$.

3 Preuves

Preuve du théorème 1. Comme $\phi^n - F_n$ et ϕ^{-n} sont analytiques dans $\overline{\mathbb{C}} \setminus E$ et s'annulent en ∞ pour $n \geq 1$, nous avons d'après la formule de Cauchy pour tout $|r| > 1$ et $\zeta \in E$ ou $\zeta \notin E$ avec $|\phi(\zeta)| < r$,

$$\frac{1}{2\pi i}\int_{|\phi(z)|=r} \frac{\phi(z)^n}{z-\zeta}\,dz = \begin{cases} F_n(\zeta) & \text{si } n \geq 0, \\ 0 & \text{si } n < 0. \end{cases}$$

Sachant que $\sigma(A) \subset \overline{W(A)} \subset E$, nous déduisons en utilisant la représentation intégrale de Dunford-Taylor que

$$\frac{1}{2\pi i}\int_{|\phi(z)|=r} \phi(z)^n\,(zI-A)^{-1}\,dz = \begin{cases} F_n(A) & \text{si } n \geq 0, \\ 0 & \text{si } n < 0, \end{cases} \tag{3.1}$$

où nous remarquons qu'on peut faire tendre r vers 1, ce qui donne pour $n \geq 1$

$$0 = \frac{1}{2\pi i}\int_{\partial E} \phi(z)^{-n}\,(zI-A)^{-1}\,dz = \frac{1}{2\pi i}\int_{\partial E} \overline{\phi(z)}^n\,(zI-A)^{-1}\,dz. \tag{3.2}$$

L'ensemble E étant convexe, le vecteur unitaire $\nu(z)$ normal extérieur au point $z \in \partial E$ existe presque partout et on a $i\,\nu(z) = dz/|dz|$. En combinant (3.1) avec l'adjoint de (3.2), nous obtenons pour $n \geq 1$

$$F_n(A) = \int_{\partial E} \phi(z)^n \mu(z, A)\,|dz|, \quad \mu(z, A) := \frac{1}{2\pi}\big(\nu(z)(zI-A)^{-1} + \bar{\nu}(z)(\bar{z}I - A^*)^{-1}\big), \tag{3.3}$$

c'est-à-dire, nous obtenons pour $F_n(A)$ la représentation de Carl Neumann introduite dans [2, Section 4.1], avec une fonction ϕ^n facile à estimer. Pour conclure, nous suivons l'approche de

[7, 2, 4, 5, 6] : comme $\nu(z)$ est la normale extérieure de E en $z \in \partial E$ et $W(A) \subset E$, nous observons que $\mu(z, A)$ est auto-adjointe et semi-definie positive en presque tout $z \in \partial E$, et $\int_{\partial E} \mu(z, A)\, |dz| = 2I$ par (3.1). En conséquence,

$$||F_n(A)|| \leq \max_{\zeta \in \partial E} |\phi(\zeta)^n| \sup_{y \in \mathcal{H}, ||y||=1} \int_{\partial E} (\mu(z, A)y, y)\, |dz| = 2.$$

$\square$

Preuve du théorème 2. Montrons dans un premier temps la chaîne d'inégalités

$$\min\{\frac{||p(A)||}{|p(0)|} : \quad p \text{ polynôme de degré au plus } k\} \leq \frac{2}{|F_k^E(0)|} \leq \frac{2\gamma_E^k}{1 - \gamma_E^{k+1}}. \tag{3.4}$$

La première inégalité dans (3.4) est une conséquence simple du théorème 1. Pour estimer inférieurement $|F_k(0)|$ pour le polynôme de Faber $F_k = F_k^E$, nous allons procéder comme par exemple dans [3, Lemma 2.2] : d'après l'hypothèse $0 \notin E$ et le principe du maximum appliqué à la fonction $\phi_E(\phi_E^k - F_k)$ analytique dans $\overline{\mathbb{C}} \setminus E$,

$$\begin{aligned}
|\phi_E(0)|\, |F_k(0)| &\geq |\phi_E(0)|^{k+1} - |\phi_E(0)(\phi_E(0)^k - F_k(0))| \\
&\geq |\phi_E(0)|^{k+1} - \max_{z \in \partial E} |\phi_E(z)(\phi_E^k(z) - F_k(z))| \\
&= |\phi_E(0)|^{k+1} - \max_{z \in \partial E} |\phi_E^k(z) - F_k(z)| \geq |\phi_E(0)|^{k+1} - 1,
\end{aligned}$$

la dernière estimation utilisant la convexité de E ayant été montrée par Kővari et Pommerenke [14, Theorem 2]. Ceci démontre (3.4). Finalement, une combinaison de (2.5) et (3.4) donne la première estimation du théorème 2, et la deuxième découle du fait que $\min\{1, 2\gamma_E^k/(1-\gamma_E^{k+1})\} \leq (2 + \gamma_E)\gamma_E^k$. $\square$

Preuve du corollaire 3. En multipliant si nécessaire A par un facteur e^{it} de module 1, nous pouvons supposer que $d = \text{dist}(0, W(A))$ coïncide avec la plus petite valeur propre de $(A+A^*)/2$. Il suffit alors d'appliquer le théorème 2 avec la lentille

$$E = \{z \in \mathbb{C} : \text{Re}(z) \geq d, |z| \leq \nu(A)\}$$

qui contient clairement $W(A)$. Le calcul de ϕ_E est élémentaire, les formules correspondantes sont données dans [3, preuve du lemme 2.2]. $\square$

Remerciements : L'auteur remercie Michel Crouzeix et Catalin Badea pour leurs remarques constructives et pour les références [1, 17].

Références

[1] Atzmon, A, Eremenko, A., Sodin, M., Spectral inclusions and analytic continuation, *Bull. London Math. Soc.* **31** (1999) 722-728.

[2] Badea, C., Crouzeix, M., Delyon, B., Convex domains and K-spectral sets. Manuscript (2004).

[3] Beckermann, B., Goreinov, S.A., Tyrtyshnikov, E.E., Some remarks on the Elman estimate for GMRES, Manuscript (2004).

[4] Crouzeix, M., Delyon, B., Some estimates for analytic functions of strip or sectorial operators, *Arch. Math.* **81** (2003) 553-566.

[5] Crouzeix, M., Numerical range, holomorphic calculus and applications. Manuscript (2005).

[6] Crouzeix, M., Operators with numerical range in a parabola, *Arch. Math.* **82** (2004) 517-527.

[7] Delyon, B., Delyon, F., Generalization of Von Neumann's spectral sets and integral representation of operators, *Bull. Soc. Math. France* **1** (1999) 25-42.

[8] Driscoll, T., A MATLAB toolbox for Schwartz-Christoffel mapping, *ACM Trans. Math. Softw.* **8** (1996) 168186.

[9] Eiermann, M., Fields of Values and Iterative Methods, *Lin. Alg. Applics* **180** (1993) 167-197.

[10] Eiermann, M., Ernst, O.G., Geometric aspects in the theory of Krylov subspace methods, *Acta Numerica* **10** (2001) 251-312.

[11] Eisenstat, S.C., Elman, H.C., Schultz, M.H., Variational Iterative Methods for Nonsymmetric Systems of Linear Equations, *SIAM J. Numer. Anal.* **20** (1983) 345-357.

[12] Elman, H.C., Iterative Methods for Sparse Nonsymmetric Systems of Linear Equations, PhD Thesis, Yale University, Department of Computer Science, 1982.

[13] Greenbaum, A., Iterative Methods for Solving Linear Systems, Frontiers in Applied Mathematics, **17**, SIAM (1997).

[14] Kövari T., Pommerenke, Ch., On Faber Polynomials and Faber Expansions, *Math. Zeitschrift* **99** (1967) 193–206.

[15] Saad, Y., Schultz, M.H., GMRES : A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Comput.* **7** (1986) 856-869.

[16] Saad, Y., Iterative methods for sparse linear systems, PWS Publishing, Boston, MA, 1996.

[17] Toh, K.C., Trefethen, L.N., The Kreiss matrix theorem on a general complex domain, *SIAM J. Matrix Anal. Appl.* **21** (1999) 145-165.