Yale University
Department of Computer Science

A New Iterative Projective Method for Solving Large Symmetric Eigenproblems

Nahid EMAD

YALEU/DCS/RR-872

This work has been supported in part by the MASI (Méthodologie et Architecture des Systèmes Informatiques) laboratory, Université Pierre et Marie Curie - CNRS UA 818, Paris, France and in part by the ONR grant N00014-91-J-1576 while the author was in residence at the Department of Computer Science, Yale University, New-haven, CT.

Abstract

We make use of Padé approximants and Krylov's sequence $(x, Ax, A^2x, \ldots, A^{m-1}x)$ in the projection subspace methods for computing a few eigenvalues of a hermitian matrix A of order n. This process consists of approximating the poles of $R_x(\lambda) = ((I - \lambda A)^{-1}x, x)$, the mean value of the resolvant of A, by those of $[m-1/m]_{R_x}(\lambda)$, where $[m-1/m]_{R_x}(\lambda)$ is the Padé approximant of order m of the function $R_x(\lambda)$. This is equivalent to approximating the eigenvalues of A by the roots of the polynomial of degree m of the denominator of $[m-1/m]_{R_x}(\lambda)$. This projection method, called the Padé-Rayleigh-Ritz (PRR) method, provides a simple way to determine the minimum polynomial of x in the Krylov's method for the symmetrical case. The numerical stability of the PRR method is ensured if there is not "considerable" variation in the matrix elements of A. The mainly expensive portion of this method is its projection phase, which is composed of the matrix-vector multiplications and, consequently, is well suited for parallel computing. This is also true when the matrices are sparse, as recently demonstrated, especially on massively parallel machines. This paper shows that a relationship between the PRR and Lanczos methods can be established by orthogonalizing the Krylov's vectors sequence. We then compare the PRR and Lanczos methods with regard to stability and natural parallelism.

Keywords : *projection method, large symmetrical eigenproblem, Padé approximant, Krylov's sequence, numerical stability, and parallelism.*

1 Introduction In numerical analysis one is often faced with the problem of computing a few eigenelements (λ, u) of a large hermitian matrix A :

$$Au = \lambda u \tag{1}$$

Projection methods provide a convenient approach to this problem. With these methods, instead of solving a problem whose size is n, one solves a problem restricted to a subspace F_m whose dimension is m with $m \ll n$. Recently a number of authors have significantly contributed to the theoretical and practical development of these methods [5, 8, 11, 13, 17].

One of the most-used projection methods is the orthogonal projection method, also called the Rayleigh-Ritz (RR) approximation method when A is a hermitian matrix. This method allows one to compute, for every couple (λ, u) of matrix A, a sequence of approximated eigenelements $(\lambda^{(m)}, u^{(m)})_m$. One can show[10] that $(\lambda^{(m)})_m$ is a monotonic sequence (with $\lambda^{(n)} = \lambda$). The drawback of a possible lack of strict monotonicity of the $(\lambda^{(m)})_m$ sequence is that one can choose the m parameter larger and larger ($m < n$) without significantly approaching the exact solution.

The Padé-Raleigh-Ritz (PRR) method, developed by D. Bessis and M. Villani [1] is a version of the RR method and is used in quantum mechanics and physics. The PRR method uses the Padé approximants and the Krylov's subspace as a projection subspace for solving the spectral problem of the Hamiltonian operator H, which has an important role in quantum mechanics. The authors carried out calculations for a *semibounded self-adjoint* operator H having the *discrete* (resp. *continuous*) part of the spectrum *positive* (resp. *nonpositive*).

This paper presents an adaption of the PRR method to the case of large matrices in the scope of numerical analysis ($n < \infty$), generalizing it to the case of hermitian matrices (*definite positive or not*). Also, a new PRR projection method for solving a hermitian eigenproblem is formulated and studied. Furthermore, we propose to use this method iteratively; that is, restarting a PRR step on a new Krylov's subspace with the same dimension until achieving satisfactory accuracy for the approximated eigenpairs.

The PRR method consists of approximating the poles of $R_x(\beta) = ((I-\beta A)^{-1}x, x)$, the mean value of the resolvant of A[1], by those of $[m-1/m]_{R_x}(\beta)$, where $[m-1/m]_{R_x}(\beta)$ is the Padé approximant of order m of the function $R_x(\beta)$. This is equivalent to approximating the eigenvalues of the matrix A by the roots of the polynomial of degree m of the denominator of $[m-1/m]_{R_x}(\beta)$.

Let $(\lambda^{(m)}, u^{(m)})_m$ be an approximated eigenelements sequence obtained by this method. We state that $(\lambda^{(m)})_m$ is a *strictly monotonic* sequence for $m \leq s$ where s is the number of poles of the mean value of the resolvent of A. Consequently, two different choices, m_1 and m_2, of the parameter m allow us to have two different approximated eigenvalues, $\lambda^{(m_1)}$ and $\lambda^{(m_2)}$, for the exact eigenvalue λ of A with :

$$|\lambda^{(m_2)} - \lambda| < |\lambda^{(m_1)} - \lambda| \tag{2}$$

[1]We posed $\beta = \lambda^{-1}$ in the original form of the mean value of the resolvant of A: $((A-\lambda I)^{-1}x, x)$ where x is any non null vector in $\mathcal{C}^n$.

for $m_2 > m_1$.

The linear dependence of a Krylov's vectors sequence $S_x^m = (x, Ax, \cdots, A^{m-1}x)$ is very important in the orthogonal projection methods which use the Krylov's subspace F_m spanned by these vectors. We thus give a *necessary and sufficient condition* for the linear dependence of S_x^m.

This paper suggests possible choices for the initial vector x and the projection subspace size m. We use this condition in the study of the PRR method, in which an assumption of linear dependence or independence takes place at every step.

The connection of the Padé approximants with the theory of orthogonal polynomials allows us to establish a relationship between the PRR and Lanczos methods[7, 2, 9]. In fact, it can be shown that by orthogonalizing the Krylov's vectors sequence S_x^m , we reproduce the results of the Lanczos method. A consequence of this equivalence is that the approximated eigenvalue sequence $(\lambda^{(m)})_m$ obtained by the Lanczos method is a strictly monotonic sequence.

We propose, in section 4, a simple iterative PRR algorithm, showing its important points. We shall see that for a fixed m, except for the projection phase, we must:

1. Solve a symmetric linear system of order m.
2. Compute the roots of a polynomial of degree m or compute the eigenvalues of a non-symmetric and sparse Hessenberg matrix of order m.

Then, in section 5, we present a comparison of the stability, complexity, and parallelism between the PRR and Lanczos methods.

Finally, in conclusion, we present the criteria for a reliable assessment of the PRR method.

2 Preliminary Study of the PRR method. Let x be any non null vector in $\mathcal{C}^n$, let $S_x^m = (x, Ax, \cdots, A^{m-1}x)$ be the sequence of Krylov's vectors, and let P_m be the projector onto the subspace F_m spanned by the sequence S_x^m. Consider the mean value of the resolvant of $A_m = P_m A P_m$, the projection matrix of A onto F_m:

$$R_x^m(\beta) = ((I - \beta A_m)^{-1}x, x) \tag{3}$$

The PRR method consists of approximating the poles of $R_x(\beta)$ by ones of $R_x^m(\beta)$. We will see that this is equivalent to approximating the poles of $R_x(\beta)$ by those of $[m-1/m]_{R_x}(\beta)$, the Padé approximant of order m of this function.

During the study of the PRR method, we show[7] that the sequence of approximated eigenvalues $(\lambda^{(m)})_m$ is strictly monotonic. In other words, $(\lambda^{(m)})_m$ (for $m < n$) is a strictly monotonic sequence of bounds for the exact eigenvalue λ of A.

For this, the function $\beta \mapsto \beta R_x(\beta)$ not being determinate in the position of the poles, one defines

the auxiliary function $f(\beta) = arctg(\beta R_x(\beta))$. This function has the remarkable properties that it is holomorphic and monotonic on the real axis. Consequently, the poles have been changed into points of holomorphy of $f(\beta)$. It is now possible to obtain bounds for $f(\beta)$. This enables us to build up a strictly monotonic sequence of bounds for the poles $\lambda_1, \lambda_2, \cdots, \lambda_s$ of $\beta R_x(\beta)$, and consequently for some of the eigenvalues of A [9].

Some properties of the resolvant of A.

The spectral decomposition of the resolvant of A allows us to redefine the mean value of the resolvant of A by:

$$\beta R_x(\beta) = (\beta(I - \beta A)^{-1}x, x) = \sum_{i=1}^{p} \frac{a_i \beta \beta_i}{(\beta_i - \beta)} \tag{4}$$

where $a_i = (q_i x, x) \geq 0$, with q_i the eigenprojection corresponding to the eigenvalue $\lambda_i = (\beta_i)^{-1}$ and p the number of the *distinct* eigenvalues of A.

- The function $\beta \mapsto \beta R_x(\beta)$ is well defined for all $\beta \in \mathcal{C}$, except for some poles. These poles $\beta_1, \beta_2, \cdots, \beta_s$ are given by the inverse of the eigenvalues for which x is not orthogonal to the corresponding eigenspace M_i (for $i = 1, .., s$):

$$(x, M_i) \neq 0 \tag{5}$$

- Since for all $\beta \in \mathcal{R}$ $(\beta \neq \beta_i)$:

$$\frac{d}{d\beta}(\beta R_x(\beta)) = \sum_{i=1}^{s} \frac{a_i \beta_i^2}{(\beta - \beta_i)^2} > 0 \tag{6}$$

  the function $\beta \mapsto \beta R_x(\beta)$ is *strictly increasing* in β.
- We note by $\tilde{\beta}_i$ (resp. $\overline{\beta}_i$) the positive (resp. negative) poles of $\beta R_x(\beta)$. Suppose that they are ordered following the scheme:

$$\overline{\beta}_{s^-} < \cdots < \overline{\beta}_2 < \overline{\beta}_1 < 0 < \tilde{\beta}_1 < \tilde{\beta}_2 < \cdots < \tilde{\beta}_{s^+} \tag{7}$$

  Where $s^+ + s^- = s$. The function $\beta \mapsto \beta R_x(\beta)$ is positive for $0 < \beta < \tilde{\beta}_1$, null in $\beta = 0$ and negative for $\overline{\beta}_1 < \beta < 0$. Therefore, it has *exactly* one zero between each two successive poles (Fig. 1).

The definition of a regular function from the resolvant.

1. We define the function $f(\beta)$, for $\beta \in \mathcal{R}$ by:

$$f(\beta) = \arctan(\beta R_x(\beta)) \tag{8}$$

   by normalizing it at $\beta = -\infty$ and prolonging it by continuity:

$$f(-\infty) = \arctan(\lim_{\beta \to -\infty} \beta R_x(\beta)) \tag{9}$$

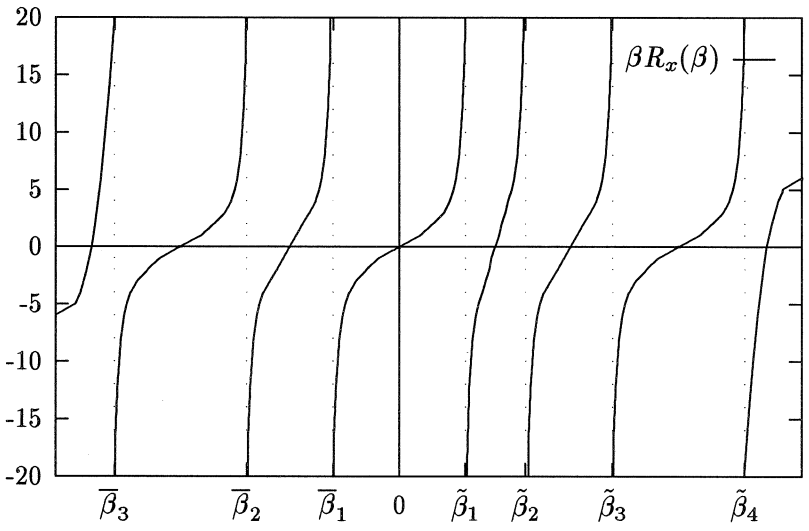

Figure 1: $s^+ = 4$ and $s^- = 3$

with

$$-(2s^- + 1)\frac{\pi}{2} < f(-\infty) < -(2s^- - 1)\frac{\pi}{2} \tag{10}$$

We have:

$$\frac{d}{d\beta}f(\beta) = \frac{\frac{d}{d\beta}(\beta R_x(\beta))}{1 + \beta^2 R_x^2(\beta)} \tag{11}$$

The function $\beta \mapsto \beta R_x(\beta)$ being meromorphic in $\mathcal{C}$, equation (11) shows that $\frac{d}{d\beta}f(\beta)$ is holomorphic for $\beta \in \mathcal{R}$. Therefore, f is also holomorphic in a vicinity of $\mathcal{R}$. On the other hand, equation (11) shows that we have always $\frac{d}{d\beta}f(\beta) > 0$, then, f is a strictly increasing function of β. Consequently, this function passes through the values $-(2k-1)\frac{\pi}{2}$ (resp. $+(2k-1)\frac{\pi}{2}$) for $\beta = \overline{\beta}_k$ (resp. $\beta = \tilde{\beta}_k$) and tends to a finite limit $f(+\infty)$ when $\beta \to +\infty$ (fig. 2) with:

$$(2s^+ - 1)\frac{\pi}{2} < f(+\infty) < (2s^+ + 1)\frac{\pi}{2} \tag{12}$$

2. Let P_m be the orthogonal projection onto subspace F_m of $\mathcal{C}^n$, and let $A_m = P_m A P_m$ be the projection matrix of A onto F_m. We define, in the same way as 1, the function F_m by:

$$f_m(\beta) = \arctan(\beta R_x^m(\beta)) \tag{13}$$

This function has the same properties as f.

2.1 Padé-Rayleigh-Ritz = (Rayleigh-Ritz) + Krylov + Padé. Suppose F_m is the subspace spanned by the Krylov's vectors sequence $S_x^m = (x, Ax, \cdots, A^{m-1}x)$. The subspace F_m so defined is called Krylov's subspace spanned by vector x with respect to matrix A.

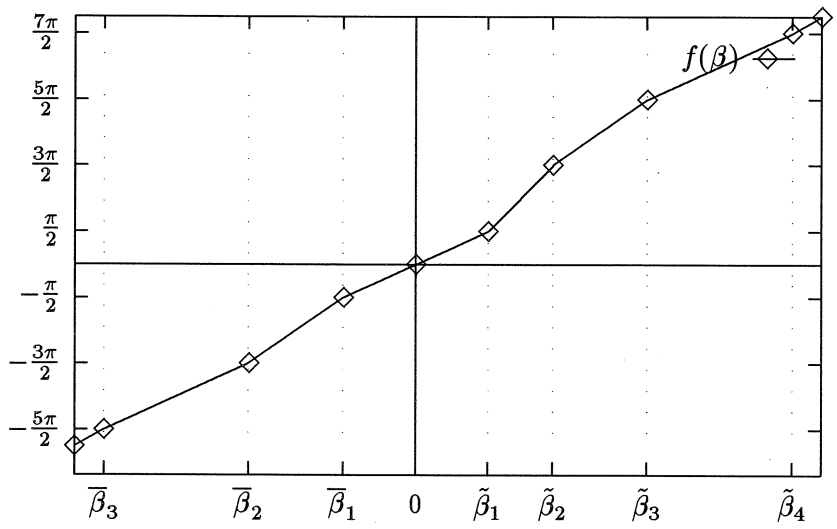

Figure 2: $s^+ = 4$ and $s^- = 3$

By expanding the functions $R_x(\beta)$ and $R_x^m(\beta)$, in power of β, we have formally:

$$R_x(\beta) = ((I - \beta A)^{-1}x, x) = \sum_{k=0}^{\infty} C_k \beta^k \tag{14}$$

$$R_x^m(\beta) = ((I - \beta A_m)^{-1}x, x) = \sum_{k=0}^{\infty} C_k^m \beta^k \tag{15}$$

with $C_k = (A^k x, x)$ and $C_k^m = (A_m^k x, x)$. We have the following fundamental result [1, 7] :

Theorem 1 *Suppose S_x^m is a linearly independent Krylov's vectors sequence. Then, the Padé approximation $[m-1/m]R_x(\beta)$ constructed out of the first $2m$ moments $C_k = (A^k x, x)$, $k = 0, 1, ..., 2m-1$, fulfills $[m-1/m]_{R_x}(\beta) = R_x^m(\beta)$. If the Krylov's vectors S_x^m are linearly dependent, then $[j-1/j]_{R_x}(\beta) = R_x(\beta)$ for all $j \lesseqgtr^{n}_{m}$.*

Proof 1 *The hypothesis of linear independence of the vectors $S_x^m = (x, Ax, \cdots, A^{m-1}x)$ spanning the subspace F_m implies:*

$$A_m^\ell x = (P_m A P_m)^\ell x = A^\ell x \qquad for \quad 0 \leq \ell \leq m-1 \tag{16}$$

and

$$A_m^{\ell'+1} x = (P_m A P_m) A^{\ell'} x = P_m A^{\ell'+1} x \qquad for \quad 0 \leq \ell' \leq m-1 \tag{17}$$

because $(P_m)^2 = P_m$. By using the inner product of (16) and (17) , for $1 \leq \ell + \ell' + 1 \leq 2m-1$, we have:

$$(A_m^{\ell'+\ell+1} x, x) = (P_m A^{\ell'+1} x, A^\ell x) = (A^{\ell'+1} x, P_m A^\ell x) = (A^{\ell'+\ell+1} x, x) \tag{18}$$

or

$$C_k^m = (A_m^k x, x) = (A^k x, x) = C_k \qquad for \quad 0 \leq k \leq 2m-1 \tag{19}$$

Then, by using equations (14,15) and the equalities (19), we have:

$$R_x(\beta) - R_x^m(\beta) = \sum_{k=0}^{\infty}(C_k - C_k^m)\beta^k = O(\beta^{2m}) \tag{20}$$

Consider the spectral decomposition of the matrix A_m (which is of rank m):

$$A_m = \sum_{j=1}^{m} \lambda_j^{(m)} h_j^m \tag{21}$$

where h_j^m is the eigenprojection corresponding to the eigenvalue $\lambda_j^{(m)}$ of A_m. Since:

$$R_x^m(\beta) = ((I - \beta A_m)^{-1}x, x) = (\frac{I}{I - \beta\sum_{j=1}^{m}\lambda_j^{(m)}h_j^m}x, x) = \sum_{j=1}^{m}(\frac{(h_j^m)^2}{h_j^m(1-\beta\lambda_j^{(m)})}x, x) \tag{22}$$

and $(h_j^m)^2 = h_j^m$. We have, then:

$$R_x^m(\beta) = \sum_{j=1}^{m}\frac{(h_j^m x, x)}{(1-\beta\lambda_j^m)} \tag{23}$$

This equation shows that $R_x^m(\beta)$ is a rational fraction with a denominator of degree m and a numerator of degree $m-1$ in β. On the other hand, equation (20) shows that it differs from $R_x(\beta)$ by a factor of order β^{2m}. Consequently, according to the definition of the Padé approximations, we have:

$$[m-1/m]_{R_x}(\beta) = R_x^m(\beta) \tag{24}$$

If the vectors $(x, Ax, A^2x, ..., A^{m-1}x)$ are linearly dependent, we have:

$$A_m^\ell x = (P_m A P_m)^\ell x = A^\ell x \qquad for\ all \quad \ell \geq 0 \tag{25}$$

Hence:

$$[m-1/m]_{R_x}(\beta) = R_x^m(\beta) = R_x(\beta) \tag{26}$$

□

According to the above result and some properties of the Padé approximants [2, 9, 12], the function $f_m(\beta)$ can also be define by:

$$f_m(\beta) = \arctan\beta[m-1/m]_{R_x}(\beta) = \arctan[m/m]_{\beta R_x}(\beta) \tag{27}$$

2.2 Strict Monotonicity of the Approximated Eigenvalues. The definition of f_m and some properties [2, 9, 12] of Padé approximations allow us to show[1]:

Theorem 2 (Bessis-Villani, Emad) *For $\beta > 0$, we have always $f_{m+1}(\beta) > f_m(\beta)$. If at one point $\beta^* \neq 0$, we have $f_{m+1}(\beta^*) = f_m(\beta^*)$, then, for all k $(0 \leq k \leq n-m)$, we have $f_{m+k}(\beta) = f_m(\beta) = f(\beta)$, and for $\beta < 0$ we always have $f_{m+1}(\beta) < f_m(\beta)$.*

Proof 2 *Suppose:*

$$[m-1/m]_{R_x}(\beta) = \frac{P_{m-1}(\beta)}{Q_m(\beta)} \tag{28}$$

where $P_{m-1}(\beta)$ and $Q_m(\beta)$ are polynomials in β of degree $m-1$ and m respectively, and $Q_m(0) = 1$. Let be $D_j = det(\Delta_j)$ (cf: section 4). We recall the following identity [9]:

$$[m/m+1]_{R_x}(\beta) - [m-1/m]_{R_x}(\beta) = \frac{\beta^{2m} D_m^2}{Q_m(\beta) Q_{m+1}(\beta)} \tag{29}$$

For a hermitian matrix, we can show that D_m is always positive. In fact, we have the following relationship:

$$D_m = \gamma_m D_{m-1} \tag{30}$$

where $(\gamma_m)_m$ is some sequence of the real positive values and $D_{-1} = 1$. The above equality shows that if for some m_0, $D_{m_0} = 0$, we have then:

$$D_{m_0+k} = 0 \qquad \text{for all } k \text{ such that} \quad 0 \leq k \leq (n-1) - m_0 \tag{31}$$

and the mean value of the resolvant of A will be reduced to its Padé approximant of order m:

$$R_x(\beta) = [n-1/n]_{R_x}(\beta) = \cdots = [m_0 - 1/m_0]_{R_x}(\beta)$$

On the other hand, by using equation (29) we see that if a point $\beta^* \neq 0$ exists such that $f_{m+1}(\beta^*) = f_m(\beta^*)$, then $D_m = 0$ and therefore:

$$f_{m+k}(\beta) = f_m(\beta) = f(\beta) \qquad \text{for all } k \text{ such that} \quad 0 \leq k \leq (n-1) - m \tag{32}$$

Apart from this particular case, we see that for $\beta > 0$ the $f_{m+1}(\beta)$ and $f_m(\beta)$ functions have no point of intersection. Consequently, in order to fix their relative position, we consider their relative values in the vicinity of zero. In fact, we deduce from (29) that:

$$f_{m+1}(\beta) - f_m(\beta) = (D_m)^2 \beta^{2m+1} + O(\beta^{2m+2}) \tag{33}$$

Because the Taylor's development, in the vicinity of zero of the function $f_m(\beta) = \arctan(Z_m)$ with $Z_m = \beta[m-1/m]_{R_x}(\beta)$ provides that:

$$f_m(\beta) = \arctan(Z_m) = Z_m + O(\beta^{2m+1}) = [m/m]_{\beta R_x}(\beta) + O(\beta^{2m+1}) \tag{34}$$

we have:

$$\begin{cases} f_{m+1}(\beta) > f_m(\beta) & \text{for } \beta > 0 \\ f_{m+1}(\beta) < f_m(\beta) & \text{for } \beta < 0 \end{cases}$$

□

As long as the Krylov's vectors $x, Ax, ..., A^{m-1}x$ are linearly independent, the poles of $[m-1/m]_{R_x}(\beta)$, which are the eigenvalues of A_m, are all *real* and *distinct.* Suppose they are ordered in the following scheme:

$$\overline{\beta}_{m^-}^{(m)} < \cdots < \overline{\beta}_2^{(m)} < \overline{\beta}_1^{(m)} < 0 < \tilde{\beta}_1^{(m)} < \tilde{\beta}_2^{(m)} < \cdots < \tilde{\beta}_{m^+}^{(m)} \tag{35}$$

where $m^+ + m^- = m$. According to the definition of f_m (for $k = 1, \cdots, m^-$ and $j = 1, \cdots, m^+$) we have:

$$f_m(\overline{\beta}_k^{(m)}) = -(2k-1)\frac{\pi}{2} \quad \textit{and} \quad f_m(\tilde{\beta}_j^{(m)}) = (2j-1)\frac{\pi}{2} \tag{36}$$

On the other hand, the orthogonality properties of the denominators of the Padé approximations $[m/m]_{\beta R_x}(\beta)$ imply that between two successive poles of $[m/m]_{\beta R_x}(\beta)$, we have *exactly* one pole of $[m+1/m+1]_{\beta R_x}(\beta)$, except between the two poles nearest to the origin of the $[m/m]_{\beta R_x}(\beta)$. In fact, between those two poles (i.e.: $\tilde{\beta}_1^{(m)}$ and $\overline{\beta}_1^{(m)}$), there are two poles of $[m+1/m+1]_{\beta R_x}(\beta)$, one positive and one negative. In other words:

$$\overline{\beta}_1^{(m)} < \overline{\beta}_1^{(m+1)} < 0 < \tilde{\beta}_1^{(m+1)} < \tilde{\beta}_1^{(m)} \tag{37}$$

Because in the contrary case, we have:

$$\overline{\beta}_1^{(m+1)} < \overline{\beta}_1^{(m)} < 0 \quad \textit{or} \quad 0 < \tilde{\beta}_1^{(m)} < \tilde{\beta}_1^{(m+1)} \tag{38}$$

Then, the function $\beta \mapsto f_m(\beta)$ is strictly increasing and the sequence $(f_m(\beta))_m$ is strictly monotonic. Combining theses two properties of f_m with equation (38), we have:

$$-\frac{\pi}{2} = f_{m+1}(\overline{\beta}_1^{(m+1)}) < f_m(\overline{\beta}_1^{(m)}) = -\frac{\pi}{2} \tag{39}$$

$$\frac{\pi}{2} = f_m(\tilde{\beta}_1^{(m)}) < f_{m+1}(\tilde{\beta}_1^{(m+1)}) = \frac{\pi}{2} \tag{40}$$

which is a contradiction in the two cases. Consequently, the sequence of the poles $(\tilde{\beta}_k^{(m)})_m$ (resp. $(\overline{\beta}_k^{(m)})_m$) is strictly decreasing (resp. strictly increasing)

Hence, if the eigenvalues $\tilde{\lambda}_i = (\tilde{\beta}_i)^{-1}$ and $\overline{\lambda}_i = (\overline{\beta}_i)^{-1}$ of the matrix A and ones $\tilde{\lambda}_i^{(m)} = (\tilde{\beta}_i^{(m)})^{-1}$ and $\overline{\lambda}_i^{(m)} = (\overline{\beta}_i^{(m)})^{-1}$ of the matrix A_m are ordered as follows:

$$\overline{\lambda}_1 < \overline{\lambda}_2 < \cdots < \overline{\lambda}_{s^-} < 0 < \tilde{\lambda}_{s^+} < \cdots < \tilde{\lambda}_2 < \tilde{\lambda}_1 \tag{41}$$

$$\overline{\lambda}_1^{(m)} < \overline{\lambda}_2^{(m)} < \cdots < \overline{\lambda}_{m^-}^{(m)} < 0 < \tilde{\lambda}_{m^+}^{(m)} < \cdots < \tilde{\lambda}_2^{(m)} < \tilde{\lambda}_1^{(m)} \tag{42}$$

they fulfill:

$$\overline{\lambda}_k^{(m_{end})} < \cdots < \overline{\lambda}_k^{(m+1)} < \overline{\lambda}_k^{(m)} < 0 \tag{43}$$

$$0 < \tilde{\lambda}_k^{(m)} < \tilde{\lambda}_k^{(m+1)} < \cdots < \tilde{\lambda}_k^{(m_{end})} \tag{44}$$

In other words, for $m = 1, \cdots, m_{end}$, the sequence of *positive* approximated eigenvalues $(\tilde{\lambda}_k^{(m)})_m$ is *strictly increasing* and the sequence of *negatives* approximated eigenvalues $(\overline{\lambda}_k^{(m)})_m$ is *strictly decreasing.*

But, what is the m_{end}? It is clear that the *strict monotonicity* of the approximated eigenvalues is true as long as the eigenvalues of A_m are all real and distinct. We have seen before that this is true as long as the sequence of the Krylov's vectors S_x^m is *linearly independent.* Therefore, m_{end} is the value of the parameter m beyond which we have linear dependence of S_x^m. Consequently, we are interested in locating the moment from where the Krylov's vectors sequence S_x^m is no longer linearly independent. We will address this question in the next section.

3 A Necessary and Sufficient Condition for the Linear Dependence of a Krylov's Vectors Sequence. Let $M_1, M_2, ..., M_p$ be the eigenspaces corresponding to p distinct eigenvalues of A. We denote the orthogonality of the vector x with the subspace M_i by $(x, M_i) = 0$. We can show the following:

Theorem 3 (Emad) *Let A be a normal (in particular hermitian) matrix of order n and let x be any non null vector of $\mathcal{C}^n$. If $(x, M_i) \neq 0$ for i=1,..,t and $(x, M_i) = 0$ for i=t+1,...,p, then S_x^t is linearly independent and S_x^m is linearly dependent for all $m > t$, where $S_x^m = (x, Ax, ..., A^{m-1}x)$.*

Proof 3 *Since A is a normal matrix, there is an orthogonal basis $u_1, u_2, \cdots, u_n$ of its eigenvectors. Let $u_{1_i}, u_{2_i}, \cdots, u_{d_i}$ be the eigenvectors associated to the eigenvalue λ_i, and let $M_i = span\{u_{1_i}, u_{2_i}, \cdots, u_{d_i}\}$ be the eigenspaces corresponding to λ_i. Then:*

$$n = \sum_{i=1}^{p} d_i \qquad and \qquad \mathcal{C}^n = \oplus_{i=1}^{p} M_i \tag{45}$$

The vector $x \in \mathcal{C}^n \Rightarrow x = \sum_{i=1}^{p} x_i$ where each vector $x_i \in M_i$. Our hypothesis for x implies $x_i \neq 0$ for $i = 1, \cdots, t$ and $x_i = 0$ for $i = t+1, \cdots, p$. Hence:

$$x = \sum_{i=1}^{t} x_i \tag{46}$$

Consider now:

$$\sum_{j=0}^{m-1} \alpha_j A^j x \qquad for \qquad y = (\alpha_0, \alpha_1, \cdots, \alpha_{m-1})^t \in \mathcal{C}^m \tag{47}$$

Equation (46) implies:

$$A^j x = \sum_{i=1}^{t} A^j x_i \qquad for\ all \qquad j \geq 0 \tag{48}$$

On the other hand, for all $i \in \{1, \cdots, t\}$, there exists a vector $(\beta_{1_i}, \beta_{2_i}, \cdots, \beta_{d_i})^t \neq 0$ such that:

$$x_i = \sum_{k=1_i}^{d_i} \beta_k u_k \qquad because \qquad x_i \in M_i \tag{49}$$

Hence, for all $j \geq 0$ and $i = 1, \cdots, t$:

$$A^j x_i = A^j \sum_{k=1_i}^{d_i} \beta_k u_k = \lambda_i^j x_i \tag{50}$$

$$\sum_{j=0}^{m-1} \alpha_j A^j x = \sum_{i=1}^{t} (\sum_{j=0}^{m-1} \alpha_j \lambda_i^j) x_i \tag{51}$$

Linear independence of the vectors $(x_1, \cdots, x_t)$ implies:

$$\sum_{j=0}^{m-1} \alpha_j A^j x = 0 \Leftrightarrow \sum_{j=0}^{m-1} \alpha_j \lambda_i^j = 0 \qquad \forall i \in \{1, \cdots, t\} \tag{52}$$

This is equivalent to the following system:

$$\begin{cases} \alpha_0 + \alpha_1 \lambda_1 + \alpha_2 \lambda_1^2 + \cdots + \alpha_{m-1} \lambda_1^{m-1} &= 0 \\ \alpha_0 + \alpha_1 \lambda_2 + \alpha_2 \lambda_2^2 + \cdots + \alpha_{m-1} \lambda_2^{m-1} &= 0 \\ \qquad\qquad \cdots\cdots\cdots & \\ \alpha_0 + \alpha_1 \lambda_t + \alpha_2 \lambda_t^2 + \cdots + \alpha_{m-1} \lambda_t^{m-1} &= 0 \end{cases} \tag{53}$$

which is equivalent to:

$$\begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots\cdots & \lambda_1^{m-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots\cdots & \lambda_2^{m-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_t & \lambda_t^2 & \cdots\cdots & \lambda_t^{m-1} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{m-1} \end{pmatrix} = V_m y = 0 \tag{54}$$

As the approximated eigenvalues $\lambda_1, \cdots, \lambda_t$ are distinct, the Vander Monde matrix V_m is invertible for $m = t$. Hence:

$$V_t y = 0 \Leftrightarrow y = 0 \tag{55}$$

Consequently, the sequence S_x^m is linearly independent. For $m = t + 1$ the above system of equations has at least one (nonzero) solution. This is equivalent to the linear dependence of S_x^m for $m > t$. These results allow us to say:

$$S_x^m \textit{ is linearly dependent} \Leftrightarrow m > t \tag{56}$$

□

An immediate consequence of this theorem is that for $m \geq t$ the projection subspace F_m is an invariant subspace of the *normal* matrix A. A_m is no longer a projection of A onto F_m but a restriction of A to F_m. Consequently, the eigenvalues of A_m are the *exact* eigenvalues of A. We propose then, a suggestion for possible choices of the initial vector x and parameter m: the number t is unknown in practice, but for any x and large n, it is, in general, large. This means that if we choose m *large*, we can hope to have $m \geq t$. But this choice is in disagreement with the general principle of projection methods for computing a few eigenvalues of a very large matrix (i.e.: $m \ll n$). Now, with m small, according to the above result, we must choose an initial vector x in such a way that it belongs to an invariant subspace $M^t = M_1 \oplus M_2 \oplus ... \oplus M_t$ and does not belong to any other invariant subspace (i.e. $x \notin M^{p-t} = M_{t+1} \oplus M_{t+2} \oplus ... \oplus M_p$).

Suppose we want to compute r eigenvalues of matrix A ($r \leq p \leq m \ll n$). As a consequence of the above theorem, we must attempt to find an initial vector x whose components are *nonzero* (resp. *zero*) in the r eigenspaces corresponding to the *wanted* (resp. *unwanted*) eigenvalues. The research of such initial vectors has been the aim of much investigation, particularly in [4, 5, 17, 16].

In the specific case of the PRR method, we can consider that the vector x defined in this method fulfills the condition of the above theorem for $t = s$, the poles number of $R_x(\beta)$. In other words, the vector x belongs to the invariant subspace $M^s = M_1 \oplus ... \oplus M_s$ and does not belong to the rest of invariant subspaces: $M^{p-s} = M_{s+1} \oplus ... \oplus M_p$. Consequently, for $m \geq s$ the approximated eigenvalues obtained by PRR are the *exact* eigenvalues of A and the m_{end} parameter, defined in the last section, is equal to s. The inequalities (43, 44) can now be rewriten:

$$\overline{\lambda}_k = \overline{\lambda}_k^{(s)} = \overline{\lambda}_k^{(m_{end})} < \cdots < \overline{\lambda}_k^{(m+1)} < \overline{\lambda}_k^{(m)} < 0 \tag{57}$$

$$\tilde{\lambda}_k = \tilde{\lambda}_k^{(s)} = \tilde{\lambda}_k^{(m_{end})} > \cdots > \tilde{\lambda}_k^{(m+1)} > \tilde{\lambda}_k^{(m)} > 0 \tag{58}$$

This shows that the ideal choice of parameter m is $m = s = m_{end}$. As we have just seen, this is accomplished by a *good* choice of the initial vector x.

Combining the strict monotonicity of the approximated eigenvalues sequence $(\lambda^{(m)})_m$ obtained by the PRR method and the above theorem, we can show that if x is not orthogonal to any of the eigenspaces, the approximated eigenvalues in the extremity of the spectrum converge to the corresponding exact eigenvalues. Furthermore, the error bounds given by [16] show that these approximations are more and more precise as they go up to the extremity of the spectrum of A.

4 The PRR Method Consider an even number of moments $C_0, C_1, ..., C_{2m-1}$. We consider the denominator of the Padé approximation $[m/m]_{\beta R_x}(\beta)$ built up on the resolvant of A:

$$Q_m(\lambda) = det \begin{pmatrix} C_0 & C_1 & ..... & C_m \\ C_1 & C_2 & ..... & C_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ C_{m-1} & C_m & ..... & C_{2m-1} \\ 1 & \lambda & ..... & \lambda^m \end{pmatrix} \tag{59}$$

The PRR method consists of approximating m distinct eigenvalues of A by the m real and distinct roots $(\lambda_i^{(m)})_{i=1,m}$ of the polynomial $Q_m(\lambda)$. The $\lambda_i^{(m)}$ are the eigenvalues of A_m, the part of A onto $F_m = span(x, Ax, A^2x, ..., A^{m-1}x)$ subspace with dimension$(F_m) = m$.

Computing roots of $Q_m(\lambda)$ requires the polynomial coefficients $b_0, b_1, ..., b_{m-1}$ of

$$Q_m(\lambda) = D_{m-1}(b_0 + b_1\lambda + ... + b_{m-1}\lambda^{m-1} + \lambda^m) \tag{60}$$

This can be done using the following relations:

$$b_j = -D_{m-1}^{-1}C_{m+j} \qquad \forall j \in \{0, \cdots, m-1\} \tag{61}$$

or by solving the following linear system:

$$\underbrace{\begin{pmatrix} C_0 & C_1 & ..... & C_{m-1} \\ C_1 & C_2 & ..... & C_m \\ \vdots & \vdots & \vdots & \vdots \\ C_{m-1} & C_m & ..... & C_{2m-2} \end{pmatrix}}_{\Delta_{m-1}} \underbrace{\begin{pmatrix} b_{m-1} \\ b_{m-2} \\ \vdots \\ b_0 \end{pmatrix}}_{b} = - \underbrace{\begin{pmatrix} C_m \\ C_{m+1} \\ \vdots \\ C_{2m-1} \end{pmatrix}}_{c} \tag{62}$$

As soon as the vector b is known, we can use either:

1. A method to compute the roots of an explicit polynomial (for example: Bairstow).
2. A method to compute the eigenvalues of a non-symmetric and sparse matrix (for example: QR). Since, if we let $p_i = -b_i$ (for $i = 0, 1, ..., m-1$), then:

$$Q_m(\lambda) = \lambda^m - p_{m-1}\lambda^{m-1} - p_{m-2}\lambda_{m-2} - \cdots - p_1\lambda - p_0 \tag{63}$$

where the companion matrix is:

$$H_m = \begin{pmatrix} p_{m-1} & p_{m-2} & ..... & p_1 & p_0 \\ 1 & 0 & ..... & 0 & 0 \\ 0 & 1 & ..... & 0 & 0 \\ \vdots & \vdots & ..... & \vdots & \vdots \\ 0 & 0 & ..... & 1 & 0 \end{pmatrix} \tag{64}$$

It is not difficult to show that the approximated eigenvector u_i of A corresponding to $\lambda_i^{(m)}$ can be obtained by the following relation:

$$u_i^{(m)} = \sum_{j=0}^{j=m-1} d_j Q_j(\lambda_i^{(m)}) Q_j(A)x \tag{65}$$

where $Q_j(t)$ is the denominator of the Padé approximant [j-1/j](t) and $d_j = (D_j D_{j-1})^{-1}$ with $D_j = det(\Delta_j)$ and $D_{-1} = 1$. The special form of Δ_j matrices and their inter-dependence allows an important simplification for computation of $u_i^{(m)}$ and/or for solving the $\Delta_{m-1} b = c$ linear system. In fact, we can show:

$$D_j = C_{2j} D_{j-1} - \| E_j \|^2 \qquad \text{for all} \qquad j \overset{\leq m-1}{\geq 0} \tag{66}$$

$$Q_j(\lambda) = \lambda^j D_{j-1} - (E_j, \Lambda_j) \qquad \text{for all} \qquad j \overset{\leq m-1}{\geq 0} \tag{67}$$

$$Q_j(A)x = A^j x D_j - (E_j, \Lambda_j^x) \qquad \text{for all} \qquad j \overset{\leq m-1}{\geq 0} \tag{68}$$

where $E_j = (C_j, C_{j+1}, \cdots, C_{2j-1})^t$, $\Lambda_j = (1, \lambda, \cdots, \lambda^{j-1})^t$ and $\Lambda_j^x = (x, Ax, \cdots, A^{j-1}x)^t$.

Suppose, we want to compute some number r of the eigenelements of a matrix A of order n ($r \leq m \ll n$). By using the PRR method, the realized accuracy of approximated eigenvalues can be unsatisfactory.

We propose to use this method iteratively. That is, to restart a PRR step with a new initial vector, until obtaining the desired accuracy. Concerning the choice of an initial vector, according to the theorem in the previous section, we need to find a vector with *nonzero* (resp. *zero*) components in the eigenspaces corresponding to the *wanted* (resp. *unwanted*) eigenvalues.

4.1 Iterative PRR Algorithm

1. Choice of m.
2. Choice of initial vector x.
3. Normalization of x : $y_0 = x / \| x \|$, $C_0 = \| y_0 \|^2 = 1$
4. Computation of $C_1, C_2, ..., C_{2m-1}$.
   - $y_1 = A y_0$
   - For $k = 1, m-1$, do
     $C_{2k-1} = (y_k, y_{k-1})$
     $C_{2k} = (y_k, y_k)$
     $y_{k+1} = A y_k$
   - End for k
   - $C_{2m-1} = (y_m, y_{m-1})$

5. Linear system solving : $\Delta_{m-1} b = c$.
6. Computation of the eigenvalues of H_m matrix.
7. Computation of the approximated eigenvectors $u_i^{(m)}$ for $i = 1, ..., m$ by (65).
8. If $(\min_{1 \leq i \leq r} \parallel (A - \lambda_i^{(m)} I) u_i^{(m)} \parallel > p$: *requested precision*) Then
   with a new initial vector go to 3.
   End if
9. End

5 Comparison with the Lanczos iterative method. Let us suppose $dim(F_m) = m$ with $F_m = \{x, Ax, \cdots, A^{m-1}x\}$.

The Padé approximants are deeply linked with the theory of orthogonal polynomials. As a consequence, one can establish a relationship between the PRR and Lanczos methods. In fact, if we orthogonalize the Krylov's vectors sequence S_x^m by Gramm Schimdt orthogonalization, we can show [7, 1] that the results of the Lanczos method will be reproduced. For this, let $y^* = (y_0, y_1, \cdots, y_{m-1})$ be this orthonormal vectors basis of F_m. It is not difficult to show that the following relations are satisfied (for $j \overset{\geq 0}{\leq m-2}$):

$$\begin{cases} y_0 & \text{such that } \parallel y_0 \parallel = 1 \\ \alpha_j = (Ay_j, y_j) & \\ y'_{j+1} = Ay_j - \alpha_j y_j - \beta_{j-1} y_{j-1} & \text{with } \beta_{-1} = 0 \\ \beta_j = \parallel y'_{j+1} \parallel & \\ y_{j+1} = y'_{j+1} / \beta_j & \end{cases}$$

Now, in the y^* basis, $A_m = P_m A P_m$ is a tridiagonal and symmetrical matrix. In fact, if Y represents the matrix of $y_0, y_1, \cdots, y_{m-1}$ vectors, we have:

$$T_m = Y^t A_m Y = Y^t A Y \tag{69}$$

with $(T_m)_{i,i} = \alpha_i = (Ay_i, y_i)$, $(T_m)_{i,i+1} = (T_m)_{i+1,i} = \beta_i = (Ay_i, y_{i+1})$ and $(T_m)_{i,j} = (T_m)_{j,i} = 0$ for $j \geq i + 2$. If $w_i^{(m)}$ is the eigenvector of T_m corresponding to the eigenvalue $\lambda_i^{(m)}$, then the approximated eigenvector $v_i^{(m)}$ of A corresponding to $\lambda_i^{(m)}$ can be obtained by:

$$v_i^{(m)} = Y w_i^{(m)} \tag{70}$$

Equation (69) shows clearly that the A_m and T_m matrices have the same eigenvalues. This process of tridiagonalization of a matrix by orthogonalization of the corresponding Krylov's sequence is the Lanczos method. Because the PRR and Lanczos methods produce the same results it is natural to compare them. Recall the iterative Lanczos method:

5.1 Iterative Lanczos Algorithm

1. Choice of m.
2. Choice of initial vector x.
3. Normalization of x : $y_0 = x / \parallel x \parallel$ and $\beta_{-1} = 0$

4. Computation of T_m matrix elements.
   - For $j = 0, m-2$, do
     $$\alpha_j = (Ay_j, y_j)$$
     $$y'_{j+1} = Ay_j - \alpha_j y_j - \beta_{j-1} y_{j-1}$$
     $$\beta_j = \| y'_{j+1} \|$$
     $$y_{j+1} = y'_{j+1}/\beta_j$$
   - End for j
   - $\alpha_{m-1} = (Ay_{m-1}, y_{m-1})$
5. Computation of the eigenvalues of T_m matrix.
6. Computation of the eigenvectors $w_i^{(m)}$ of T_m matrix.
7. Computation of the approximated eigenvectors $v_i^{(m)}$ for $i = 1, ..., m$ by (70).
8. If ($\min_{1 \le i \le r} \| (A - \lambda_i^{(m)} I) v_i^{(m)} \| > p$: *requested precision*) Then
   with a new initial vector go to 3.
   End if
9. End

A drawback of the Lanczos method, calling into question its stability, is that the vectors obtained by the algorithm loose their orthogonality very rapidly. Several strategies of re-orthogonalization [5, 4, 18] can be used, but this can become very expensive.

The PRR method does not have this problem. Instead, the projection matrix whose eigenvalues approximate those of A is a non-symmetric and sparse upper Hessenberg matrix: our symmetric eigenproblem is reduced to a non-symmetric problem! In order to circumvent this difficulty, an alternative seems to compute the roots of the characteristic polynomial $Q_m(\lambda)$ of which A_m is the companion matrix.

In the PRR method, we must solve a symmetric linear system which is well conditioned if there is not *considerable* variation in the size of matrix elements of A. Furthermore, its special form permits use the Bordering [3, 6, 15] method, which is particularly well suited to these kind of matrices. Computing approximated eigenvectors $u_i^{(m)}$ requires the evaluation of some expressions (see (65)- (68)). The special form of these expressions and their inter-dependence allows a very important simplification for their computation.

The mainly expensive parts of the Lanczos and PRR algorithms are their projection phases. The complexity of which is $m(\alpha + 2\beta + 5\gamma) - (\beta + 7\gamma)$ for Lanczos and $m(\alpha + 2\beta) - \beta$ for PRR, where α, β and γ are, respectively, the complexities of matrix-vector multiplication, the inner product of two vectors, and an elementary operation between a scalar and a vector of order n.

Suppose that we have $O(n^2)$ processors. Then, if we do not consider the communications time and the mapping problems, we have:

$$\alpha = \beta \qquad \text{and} \qquad \gamma \approx 1 \tag{71}$$

Nevertheless, we must recall that this is target machine dependent. Consequently, with $O(n^2)$

processors and the above hypothesis, the complexity of the projection phases of the above version of PRR and Lanczos algorithms are respectively:

$$\alpha(3m-1) \qquad \text{and} \qquad \alpha(3m-1)+5m-7 \tag{72}$$

with $\alpha = 1+\log_2 n$ in the case of dense matrices and $\alpha = 1+\log_2 c$ in the case of sparse matrices, where c is the maximum number of the non zero elements in a column of A.

Consequently, for this portion of the algorithm, the PRR method can be more efficient than the Lanczos method. Furthermore, the inclusion of re-orthogonalization makes the projection phase of the Lanczos method almost *two times* more expensive. For the Arnoldi projective method on the massively parallel architecture of the Connection Machine 2, it has been observed [14] that with $O(n^2)$ processors in the general case and with $O(nc)$ processors in the sparse case, for $\frac{n}{m}$ large, we have:

$$d(\text{projection method}) \rightarrow d(\text{projection phase})$$

where $d(x)$ is the throughput of x. We can conclude that for $\frac{n}{m}$ large, the PRR method can be two times less expensive than Lanczos (with re-orthogonalization) on such architectures.

On the other hand, large degree of parallelism is possible in the PRR algorithm due to the form of the linear system in equation (62) and the expression of $u_i^{(m)}$.

6 Conclusion We have shown that the method developed in this paper and the Lanczos method produce the same results. A theoretical comparison of the stability, complexity and natural parallelism of these methods was provided. We conclude that, under assumptions introduced in the last section, the PRR method is more efficient in terms of complexity and parallelism.

Nevertheless, we think that the most important criterion for choosing a method must be its numerical stability. We have seen that in this respect these methods are comparable. Consequently, a reliable assessment of the PRR method requires using it and comparing numerical results with those of the nearest method (i.e.: Lanczos).

Some theoretical results developed in this paper are sufficiently general that they can be applied to other projection methods (i.e.: Lanczos).

Acknowledgement The author would like to thank Françoise Chatelin and Martin Schultz for helpful discussions.

References

[1] D. Bessis and M. Villani, *Perturbative-variational approximations to the spectral properties of semibounded hilbert space operators, based on the moment problem with finite or diverging moments. Applications to quantum mechanical systems*, Math. Phys., 16 (1975), pp. 462+.

[2] C. BREZINSKI, *Padé-type approximation and general orthogonal polynomials.*, Birknäuse Verlag, Basel. Boston. Stuttgart., 1980.

[3] ——, *Bordering methods and progressive forms for sequence transformation.*, USTL Technical Report, Publication ANO-192, 1987.

[4] F. CHATELIN, *Spectral Approximation of Linear Operators*, Academic Press, New-York, 1983.

[5] ——, *Valeurs Propres de Matrices*, Masson, 1988.

[6] J.-M. DELOSME AND I. C. IPSEN, *Parallel solution of symmetric positive definite systems with hyperbolic rotations.*, Linear Algebra and Its Applications 77:75-111, 1986.

[7] N. EMAD, *Contribution à la Résolution de Grands Problèmes de Valeur Propre*, PhD thesis, Université P. and M. Curie, Paris, January 1989.

[8] G. GOLUB AND C. VAN LOAN., *Matrix Computation.*, North Oxford Academic Press, Oxford, 1984.

[9] P. GRAVES-MORRIS., *Padé Approximants and their Applications.*, Proceeding of the conference held at the University of Kent 17-21 July 1972, Academic Press, New-York, 1972.

[10] A. HOUSEHOLDER, *Localization of the Characteristic Roots of Matrices.*, Recent Advances in Matrix Theory, Madison and Milwaukee, the University of Wisc. Press, 1964.

[11] P. LASCAUX AND R. THÉODOR, *Analyse Numérique Matricielle Appliquée à l'Art de l'Ingénieur.*, Tom II, Masson, Paris, 1986.

[12] H. PADÉ, *Sur la Représentation Approchée d'une Fonction par des Fractions Rationnelles*, Annale de l'Ecole Normale supérieure, Paris, 1892.

[13] B. PARLETT., *The Symmetric Eigenvalue Problem*, Prentice Hall, Engelwood Chiffs, 1980.

[14] S. PETITON., *Parallel subspace method for non-hermitian eigenproblems on the connection machine (cm2)*, YALE Technical Report, YALEU/DCS/RR-859, 1991.

[15] J. L. PHILLIPS, *The triangular decomposition of hankel matrices.*, Mathematics of Computation, Volume 25, Number 115, 1971.

[16] Y. SAAD., *Computation of Eigenvalue of Large Hermitian Matrices by Partitioning Techniques*, Technical Report, INPG, University of Grenoble, 1974.

[17] ——, *Projection Methods for Solving Large Sparse Eigenvalue Problems*, Lecture Notes in Math., Vol 973, Springer-Verlag, Berlin, 1982.

[18] J. WILKINSON, *The Algebric Eigenvalue Problem*, Oxford University Press, 1965.