Some insights on restarting symmetric eigenvalue methods with Ritz and harmonic Ritz vectors

ANDREAS STATHOPOULOS

DEDICATION

Dedicated to Dr. David M. Young, Jr. on the occasion of his 75th birthday

Abstract

Eigenvalue iterative methods, such as Arnoldi and Jacobi-Davidson, are typically used with restarting. This has significant performance shortcomings, since important components of the invariant subspace may be discarded. One way of saving more information at restart is the idea of "thick" restarting which keeps more Ritz vectors than needed. Our previously proposed dynamic thick restarting chooses these vectors in a way that has proved efficient on a wide variety of matrices. It is also possible to keep this information more compactly by combining thick restarting with a technique based on a three term recurrence.

In this paper, we give strong experimental evidence that saving more information with thick restarting is not necessarily beneficial, and provide an explanation to the efficiency of the dynamic scheme. In addition, we show through a variety of experiments that restarting with harmonic Ritz instead of Ritz vectors does not improve the convergence of symmetric eigenvalue methods when an extreme part of the spectrum or some eigenpair within this part is needed. However, for computing highly interior eigenpairs, harmonic Ritz vectors may be the only viable alternative.

Keywords: Jacobi-Davidson, Arnoldi, Lanczos, thick, implicit restarting, deflation, eigenvalue, preconditioning, harmonic Ritz pairs

1 Introduction

Many problems in science and engineering require the solution of large, sparse, symmetric eigenvalue problems, $Au = \lambda u$, for a few of the lowest or highest (extreme) eigenvalues and eigenvectors (eigenpairs). As the difficulty and size of the problems grow, traditional methods such as the Lanczos method and its equivalent in the non symmetric case the Arnoldi method [16] become increasingly dependent on preconditioning to compensate for the loss of efficiency and robustness. The Davidson and its generalization the Jacobi-Davidson method [5, 11, 3, 18] are popular extensions to the Arnoldi method. Instead of extracting the eigenvectors from a generated Krylov space, these methods gradually build a different space by incorporating into the

Iterative Methods in Scientific Computation II
David R. Kincaid et. al. (eds.), pp. 1–15.

ISBN 0–123–45678–9

existing basis the approximate solution of a correction equation. Procedurally, the two methods are similar to the FGMRES method [17], and in this sense, we refer to the approximate solution of the correction equation as preconditioning.

In spite of using preconditioning, for many hard problems the (Jacobi-)Davidson method may still require a large number of steps. Because the vector iterates must be stored for computing the eigenvectors, the storage requirements are overwhelming. The problem is even more important in the symmetric case, where the better theoretical framework and software has led researchers to consider matrices of huge size that allow only a few vectors to be stored. Even in the Lanczos method where a three-term recurrence is known, orthogonality problems and spurious solutions prevent the application of the method for a large number of steps. For these reasons, many restarting variants of the Lanczos and (Jacobi-)Davidson methods are used in practice [4, 15, 20, 1, 6].

Contrary to methods for linear systems of equations where only one vector is needed, when restarting eigenvalue methods more than one vectors need to be stored. In the framework of Lanczos and Arnoldi methods this problem has been solved efficiently by the Implicitly Restarted Lanczos/Arnoldi methods (IRL/IRA) [20, 1, 7, 8]. IRL provides an implicit way of applying a polynomial filter during restarting and thus removing unwanted spectral information. Implicit restarting provides also an elegant formulation of several other proposed restarting schemes. For the (Jacobi-)Davidson method, restarting does not exhibit similar difficulties, because all the wanted Ritz vectors can be retained explicitly, and even additional information can be incorporated in the basis at restart. This latter characteristic facilitates many of the experiments presented in this paper.

During restarting, important components of the required invariant subspace may be discarded. This results in convergence deterioration and, frequently, stagnation of iterative methods. To reduce these effects we can save more information at every restart. In [22] we investigated the idea of "thick restarting", which implements this principle by keeping more Ritz vectors than needed. The question to be addressed is which, and how many Ritz vectors to retain at restart. For symmetric, non-preconditioned cases, a dynamic thick restarting scheme that keeps Ritz vectors on both sides of the spectrum has proved extremely efficient. With preconditioning, although it is still efficient, a less expensive scheme provides similar benefits. In [21], we combined thick restarting with a technique [12] that keeps, for each sought eigenvector, the two corresponding Ritz vectors from two successive iterations, the current and the previous one. The motivation stems from the proximity of the spaces built by the (Jacobi-)Davidson and the Preconditioned Conjugate Gradient (PCG) methods, and the fact that the two successive Ritz iterates span approximately the same space as the PCG iterates of the three term recurrence. A different approach has been followed in [1], where the dynamically obtained Leja points are given as shifts to the IRL, dampening the unwanted part of the spectrum.

The above restarting techniques have provided powerful ways of improving the convergence of iterative methods. However, there are still many issues that await either solution or explanation. The effect of keeping increasingly larger numbers of Ritz vectors in thick restarting has not been studied in the literature. A first, naive approach would be to expect that more retained information should improve convergence. We show that there is usually an optimal number of vectors to keep in thick restarting, beyond which there is convergence loss rather than gain. In addition, the success of the dynamic thick restarting is not well understood. We present an explanation for both of these phenomena by numerically studying the roots of the Arnoldi polynomials that these techniques generate, and point out some similarities between the dynamic scheme and the Leja points. We also show that the combination scheme can take advantage of any number of Ritz vectors retained.

Recently, there has been a lot of discussion about the advantages of using harmonic Ritz pairs to approximate eigenvalues and eigenvectors in the interior of the spectrum [13, 9]. Because, harmonic pairs are simply Ritz pairs of an inverted operator, they maintain the optimality of the Rayleigh-Ritz procedure but close to a specific shift. Harmonic Ritz vectors have been used successfully to deflate the eigenvector closest to zero in the GMRES procedure [2], and to obtain interior eigenpairs in the Jacobi-Davidson [6]. It is natural to ask whether harmonic vectors can be used instead of Ritz vectors to restart eigenvalue methods. We show numerically that although the eigenvector approximations may be more accurate during restarting for non extreme pairs, the harmonic vectors are not appropriate for restarting eigenvalue iterative methods when an extreme part of the spectrum is needed. We also provide a reasoning based on the Rayleigh-Ritz procedure. However, when only a few, highly interior eigenpairs are sought, harmonic pairs may be the only available choice.

The paper is organized as follows. We first review thick and dynamic thick restarting in a generic Jacobi-Davidson framework. Following, we present results from numerical experiments on the Harwell-Boeing collection, and we explain the observed effects of restarting size to thick, dynamic thick, and the combination restarting schemes. Finally, we focus on the use of harmonic vectors, and discuss why they are not appropriate for restarting eigenvalue methods.

2 Thick and dynamic thick restarting

To provide a framework for studying various restarting schemes we first present the generic Davidson method. We assume that the matrix A is symmetric of order N, with eigenpairs (λ_i, u_i) of which the l lowest (or highest) are sought. The Davidson method first appeared as a diagonally preconditioned version of the Lanczos method for the symmetric eigenproblem. Extensions, to both general preconditioners and to the nonsymmetric case have been given since [10, 3, 6]. The following describes the algorithm for the symmetric case, where the maximum basis size is $m > l$, and at every restart $k \geq l$ Ritz vectors are retained. *Ortho* denotes any stable orthogonalization procedure.

ALGORITHM **2.1 Davidson**

0. *Choose initial unit vectors* $V_k = \{v_1, \ldots, v_k\}$
1. *For* $s = 0, 1, \ldots$
2. $\quad w_i = Av_i,\ i = 1, \ldots, k-1$
3. $\quad T_{k-1} = (AV_{k-1}, V_{k-1})$
4. $\quad$ *For* $j = k, \ldots, m$
5. $\quad\quad w_j = Av_j$
6. $\quad\quad t_{i,j} = (w_j, v_i),\ i = 1, \ldots, j,$ *the last column of* T_j
7. $\quad\quad$ *Compute some wanted eigenpair, say* (μ, c) *of* T_j
8. $\quad\quad x = V_j c$ *and* $r = \mu x - Ax,$ *the Ritz vector and its residual*
9. $\quad\quad$ *Test* $\|r\|$ *for convergence.*
If satisfied target a new vector and return to 7
10. $\quad\quad$ *Solve* $M_{(s,j)}\delta = r$, *for* δ
11. $\quad\quad v_{j+1} = Ortho(V_j, \delta)$
12. $\quad$ *Endfor*

13. $V_k = \{x_i$, *the* k *lowest Ritz vectors* $\}$, $l \leq k < m$. *Restart*
14. Endfor

The preconditioning is performed by solving the equation at step 10, with $M_{(s,j)}$ approximating $(A-\mu I)$ in some sense. Without step 10 the above algorithm is simply a more expensive implementation of the Arnoldi method. Originally, Morgan and Scott [11] proposed to solve approximately the Generalized Davidson (GD) equation: $(A-\sigma I)\ \delta = r$, where σ is an approximation to the sought eigenvalue. For stability, robustness, as well as efficiency, the Jacobi-Davidson (JD) method [18] solves an equation where the operator $M_{(s,j)}$ has a range orthogonal to x, i.e., $(I-xx^T)(A-\sigma I)(I-xx^T)\ \delta = (I-xx^T)(\mu I - A)x$. For preconditioners that approximate A directly, such as incomplete factorizations and approximate inverses, the above orthogonality condition is enforced through an equivalent formulation known as Olsen method.

Step 13 of the above algorithm implements the thick restarting technique by explicitly keeping $k > l$ Ritz pairs at every restart. The same effect can be achieved with the IRA/IRL methods, by implicitly removing the unwanted Ritz vectors through the use of the corresponding Ritz values as shifts. The explicit restarting of step 13, however, has the advantage of allowing any vectors (not necessarily Ritz or polynomial transformations of v_0) to be incorporated in the restarted basis [21].

Assuming either implicit or explicit restarting the natural question is which and how many vectors to keep at every restart. In [22] it was shown that if some Ritz vectors are retained, the Lanczos process can be approximated gradually by another Lanczos process on a matrix from which the eigenvectors corresponding to these Ritz vectors have been deflated. This provided the motivation for the thick restarting shown in step 13. Keeping the vectors with Ritz values closest to the required eigenvalue, would deflated them and thus increase the gap and the convergence rate to the wanted one. We denote this thick restarting scheme by TR(k). For symmetric cases however, convergence depends on the gap ratio of the eigenvalues and therefore both ends of the spectrum are important [14]. A generalization of thick restarting would keep L lowest and R highest Ritz vectors. Dynamic thick restarting chooses these numbers using a heuristic that captures the trade off between better error reduction through more, non-restarted Lanczos steps, and larger gap ratios from a thicker restart. Our heuristic minimizes the approximate error bound on Ritz values of the Lanczos process, which is described by a Chebyshev polynomial:

$$\frac{1}{C^2_{m-L-R}(1+2\gamma_i)} \approx 2e^{-2(m-L-R)\sqrt{\gamma_i}},$$

where $m-L-R$ Lanczos steps can be taken before a new restart is necessary, and $\gamma_i = \frac{\mu_i - \mu_{L+1}}{\mu_{L+1}-\mu_{m-R}}$ is the current approximation to the gap ratio of the $i-$th eigenvalue.

2.1 The proper thickness

Both thick and dynamic thick restarting have proved extremely efficient on a wide variety of matrices. We borrow a toy example from [22] that demonstrates the effectiveness of these schemes. In figure 1, we consider a matrix of dimension 100 with eigenvalues shown in the upper part of the figure. There are two eigenvalue clusters near the origin, each containing eight equidistant eigenvalues. The rest of the eigenvalues coincide with the integers 1, 2, and so on. The lower part of the figure shows the eigenvalue convergence for IRL with basis size of 20, and with TR(k) and the dynamic restarting schemes. Significant convergence improvements

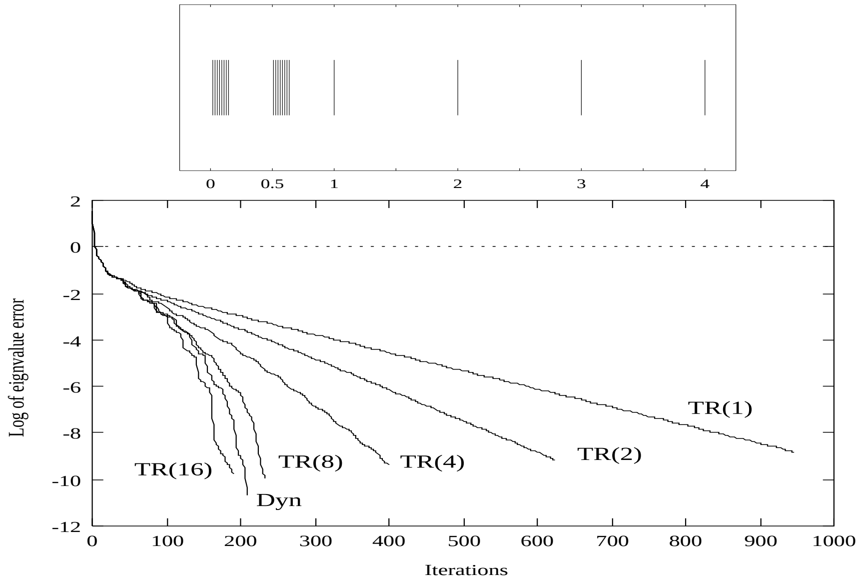

Figure 1: The effect of the TR(k) and Dynamic schemes on a toy problem. The lower part of the matrix spectrum is shown above.

are achieved by both schemes, but note the superiority of the dynamic scheme without any knowledge of the underlying spectrum.

Our experiments on Harwell-Boeing matrices have shown similar behavior for both restarting schemes [22]. We should note that these improvements are mainly apparent in difficult eigenproblems where several hundreds of steps are required for convergence. For eigenproblems that require a small number of steps, restarting occurs only a couple of times and convergence is not significantly affected. Because restarting techniques attempt to recover the convergence of the non restarted algorithm the goal is different from preconditioning. Moreover, preconditioning reduces the number of steps of iterative methods thus diminishing the effects of restarting.

Figure 1 seems to imply that convergence improves when keeping more information in the TR(k) scheme. This is a reasonable assumption, because, in TR(k), the few discarded Ritz vectors should have the least overlap with the required eigenvector. However, this is only true up to a certain value of k, beyond which the number of iterations increases. For example, in figure 1 TR(k) with $k > 16$ offers no convergence improvements.

We have conducted extensive tests on the symmetric matrices from the Harwell-Boeing collection to see if there is a common optimal thickness for restarting. Using the Davidson algorithm with basis size of 20, we seek the five smallest eigenpairs of the matrices. We run the algorithm both without preconditioning and with diagonal preconditioning. Despite different characteristics of the matrices, the convergence behavior as a function of k has been very similar throughout the matrices with noticeably few exceptions. We have accumulated our results without preconditioning in figure 2, and the results with diagonal preconditioning in figure 3. Both figures show two graphs. The left graph shows the harmonic averages of the ratios of the number of matrix vector multiplications required by TR(k) over the number required by the dynamic scheme. We choose to display the harmonic average, because in some cases the dynamic scheme is far better than TR(k) and simple averaging would skew the results

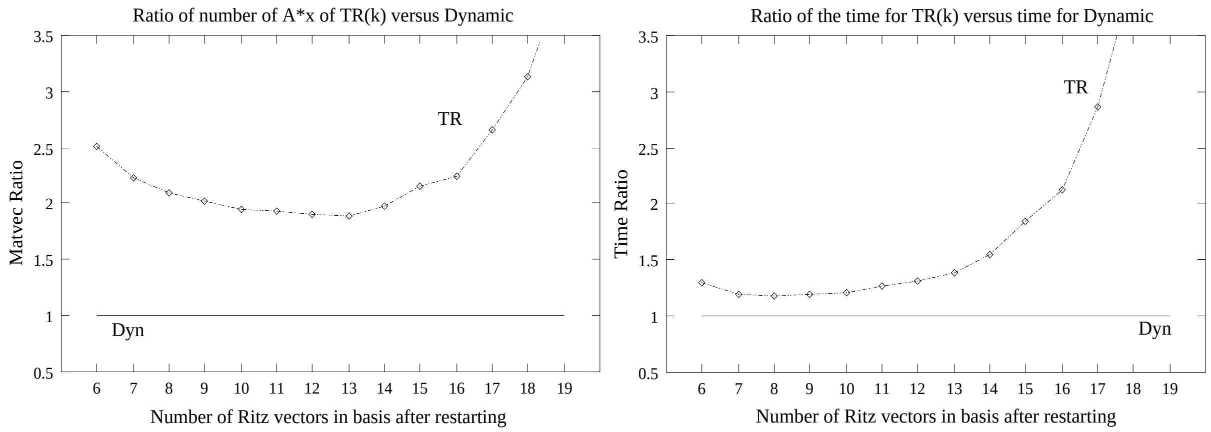

Figure 2: TR(k) versus Dynamic as a function of k. No preconditioner.

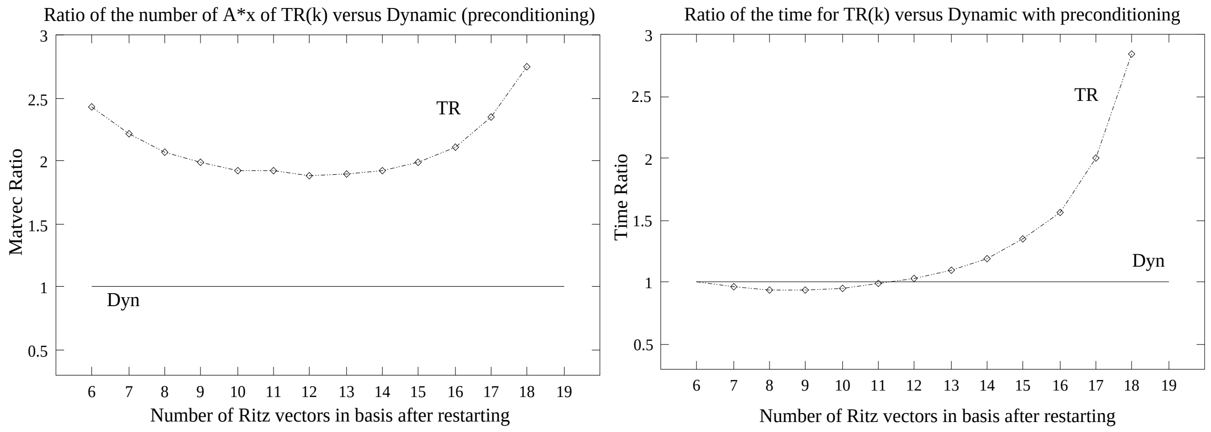

Figure 3: TR(k) versus Dynamic as a function of k. Diagonal preconditioner.

significantly. For the same reason, we also limit the number of matrix vector multiplications to 5000. The right graph shows the harmonic averages of the ratios of the corresponding times of the techniques.

Clearly, increasing the thickness of the restarting beyond a point adversely affects convergence. For the basis size of 20, this optimal k seems to be between 12 and 14. The time-graph shows a similar behavior, only the optimal range for k is between 8 and 10. There are three reasons for this. First, with large values of k, the average number of vectors present in the basis is also large and orthogonalization, which depends on the square of k, becomes an important factor. Second, we have observed that the dynamic scheme consistently keeps 16-17 vectors at restart. Although extremely efficient iteration wise, its average computational expenses per step are rather high. Third, the matrices of the Harwell-Boeing collection are very sparse, and the time to perform a matrix vector multiplication is less than the time to perform the rest computations of a Davidson step. With preconditioning, the figures confirm that the differences between the methods become smaller, because the number of iterations, and thus the information loss at restarts, is reduced.

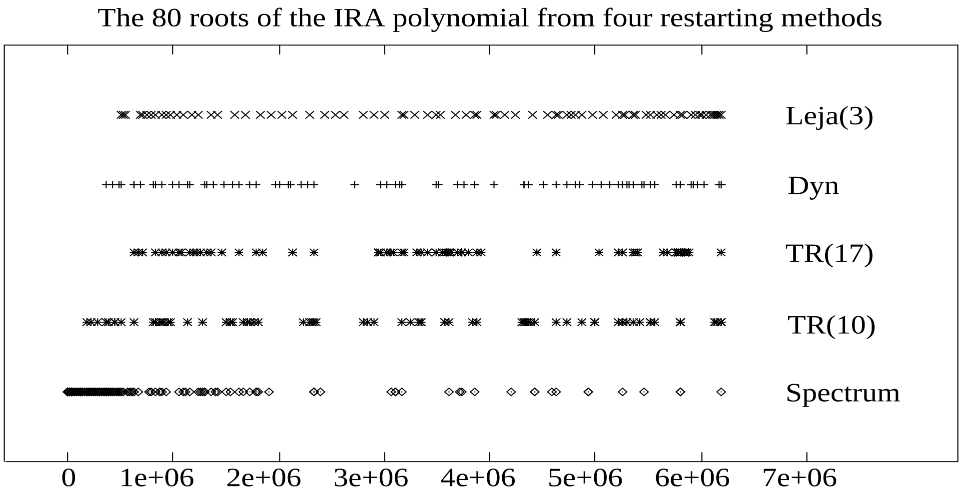

Figure 4: The 80 roots of the implicitly restarted Arnoldi polynomial as generated after 100 iterations from four different restarting techniques. The matrix is BCSSTK05 from Harwell-Boeing and its spectrum is also shown.

2.2 An experimental justification

The convergence deterioration of TR(k) with larger values of k seems counter-intuitive, if considered the outcome of keeping more information at restart. To understand this behavior we have to consider restarting from the view of annihilating shifts in implicit restarting. The equivalence between explicit thick restarting and implicit restarting is well known [22, 10], when the shifts used in the latter are the Ritz values of the pairs discarded in the former. Therefore, the restarted Lanczos/Davidson can be studied through their polynomials which have all the discarded Ritz values as roots.

In figure 4, we display the discarded Ritz values after 80 shifts have been applied in the IRL method, for four restarting techniques: TR(10), TR(17), Dynamic thick restarting, and Leja(3). The BCSSTK05 matrix from Harwell-Boeing is considered as a representative example, and its spectrum is also plotted for reference. The restarting technique Leja(3) computes three Leja points at every restart, and provides them as shifts to the implicit restarting procedure. The advantage of the Leja points is that they are computed dynamically over a changing domain, yet their distribution converges to the zeros of the Chebyshev polynomial in the same region. Leja shifts have proved a competitive restarting strategy for the IRL [1]. If run to convergence, the above methods find the five smallest eigenpairs of BCSSTK05 in the following number of iterations: TR(10) in 1110, TR(17) in 1733, Dynamic in 591, and Leja(3) in 596.

From figure 4, a factor affecting convergence is the region covered by the annihilating shifts. The residual polynomial of TR(17) has a significantly smaller region than other schemes, and the worst convergence. However, the distribution of the shifts is more important. Note that the shifts from TR(17) cluster around 3-4 areas. The residual polynomial dampens these small areas very well, but its absolute value increases sharply in the space between them. The TR(10) shifts also tend to cluster around certain areas, but these are closer together providing a more uniform range. On the other hand, the dynamic and Leja(3) schemes do not exhibit such clustering. To explain this, note that with the TR(17) the three Ritz pairs discarded each time are always the largest ones. Since they have not had time to fully converge before restarting, they keep coming back, and slowly reproduce that part of the spectrum. Similar but less pronounced effects are observed with TR(10). The dynamic scheme picks Ritz values from various areas of the spectrum producing a more uniform residual polynomial [22].

In our experiments, the dynamic scheme has outperformed Leja(3) marginally. One of the reasons is that Leja shifts try to dampen a whole region without taking into account the distribution of the eigenvalues in it. The dynamic scheme uses Ritz values as shifts and therefore it might avoid regions without eigenvalues, where the residual polynomial does not have to be equally dampened. Another important observation is that low degree annihilating polynomials at each restart, i.e., few discarded pairs, are much more effective. If we force the dynamic scheme to discard more than 5 Ritz pairs at restart, the convergence deteriorates dramatically. Similarly, Leja(10) converges in 670 iterations and Leja(15) in 800 iterations. Evidently, by annihilating only a few pairs at a time, we get the most recently updated boundary or spectrum information from IRL.

3 Thickness and the TR-CG scheme

As mentioned above, dynamic thick restarting outperforms TR(k) but it keeps more vectors at restart, thus increasing the average computational expense per iteration. For symmetric matrices, the TR-CG scheme we have proposed recently [21] manages to store similar information but with fewer vectors. We have shown that if the Ritz value at the i-th step of Lanczos is known a priori, there is a 3-term recurrence (obtained by CG) that yields the corresponding Ritz vector. Both the Lanczos and the 3-term recurrence start with the same initial vector. In this case restarting is obviated, because all history information is contained in the last three vector iterates. Although the eigenvalue is usually not known, often we can assume that this future Ritz value is approximately known, either because eigenvalues converge faster than eigenvectors in the symmetric case, or because of slow convergence. In that case, we have shown that the recurrence yields a vector with a relative distance from the Ritz vector bounded by the approximation error in the Ritz value times a constant.

This property can be used effectively in restarting. Assuming that, starting from $x^{(0)}$, a JD/Lanczos process generates a sequence of Ritz vectors $x^{(i)}$ as shown below, there is a CG sequence $y^{(i)}$, starting also from $x^{(0)}$, that yields $x^{(i+1)}$ after $i+1$ steps. Based on the theory outlined above, if the Ritz value is almost constant between steps $i-1, i, i+1$, then $x^{(i-1)}$ and $x^{(i)}$ approximate $y^{(i-1)}$ and $y^{(i)}$ respectively, and thus $x^{(i+1)}$ approximately lies in their span. The idea is to restart with the Ritz vector from step $i-1$ along with the one from step i.

$$
\begin{array}{lcccccc c}
 & & & & & & \text{Restart} & \\
 & & & & & & \downarrow & \\
\text{JD}: & x^{(0)} & x^{(1)} & x^{(2)} & \dots & x^{(i-1)} & x^{(i)} & \\
 & & & & & & & x^{(i+1)} \\
\text{CG}: & x^{(0)} & y^{(1)} & y^{(2)} & \dots & y^{(i-1)} & y^{(i)} &
\end{array}
$$

The above 3-term recurrence is obtained by applying the CG iteration on the correction equation for a specific eigenpair. In this sense, this scheme resembles preconditioning, improving convergence towards that specific eigenpair, but delaying convergence to the rest. For more eigenpairs, Murray et al. [12] proposed to restart with Ritz vectors from the previous step for all required eigenpairs. However, this does not work well if all lower eigenpairs have not

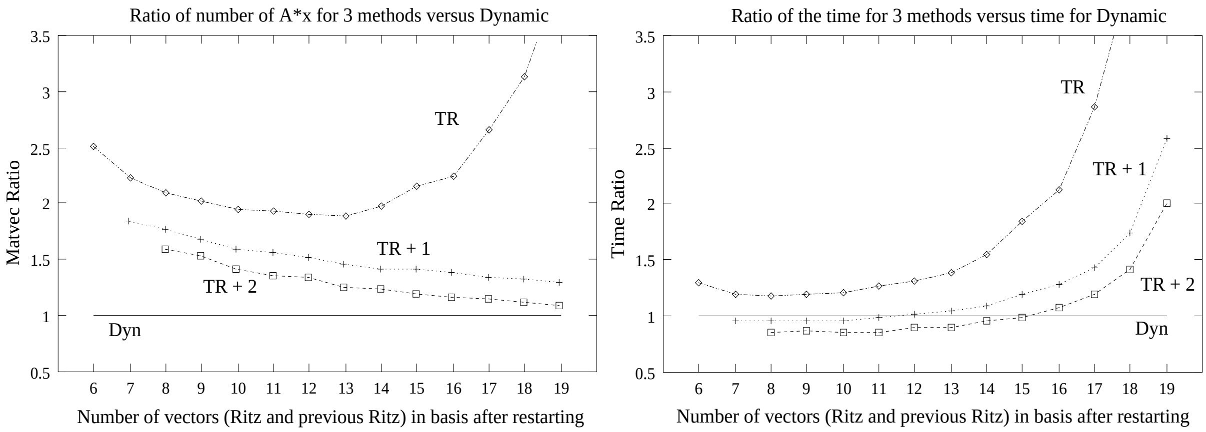

Figure 5: TR(k)+1 and TR(k)+2 versus Dynamic as a function of k. No preconditioner.

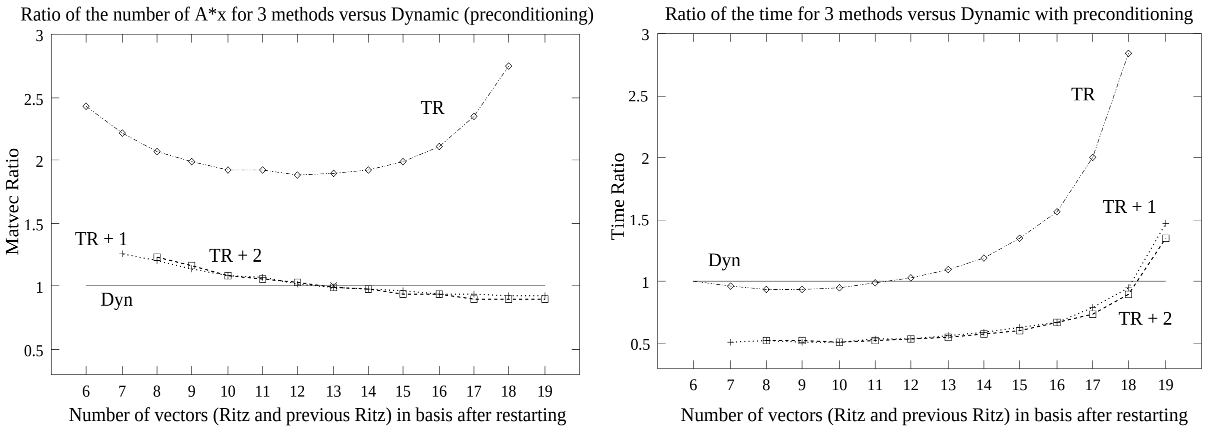

Figure 6: TR(k)+1 and TR(k)+2 versus Dynamic as a function of k. Diagonal preconditioner.

converged adequately. Instead, we proposed to combine this scheme with thick restarting (TR-CG). At every restart, we keep k lowest Ritz vectors and p Ritz vectors from the previous step corresponding to the lowest non converged pairs. We denote this scheme by TR(k)+p.

Our previous results have shown that the TR(k)+p scheme frequently matches the number of iterations of the dynamic scheme, and often it takes less time to converge. The benefits are especially apparent in the case of preconditioning. Figures 5 and 6 depict the effect of restarting thickness to the TR-CG method without and with diagonal preconditioning respectively. These figures are superpositions of the new results for TR(k)+1 and TR(k)+2 on figures 2 and 3. As before, we show harmonic averages of the ratios of matrix vector multiplications (left graphs) and ratios of the corresponding times (right graphs), using the dynamic scheme as reference.

In contrast to TR(k), both of the TR(k)+p schemes can effectively use additional vectors in thick restarting to reduce the number of iterations. The right graphs of both figures show that time increases with thickness. This is expected because for larger values of k, the quadratic factor of the orthogonalization process dominates the moderate convergence improvements. To explain this behavior, we note that for most Harwell-Boeing matrices the Ritz values do not change drastically between restarts, because of the many steps required for convergence. Therefore, the TR-CG scheme targets a specific eigenpair retaining almost all necessary information needed to obtain the next Ritz vector. Keeping more Ritz vectors will not harm convergence, but it can improve nearby eigenpairs both for future use and for helping isolate the required one. Finally, because of its relation to CG, the TR-CG works well when the preconditioner does not vary between steps. In case of variable preconditioners (such as solving the correction equation with an iterative method) the efficiency of the TR-CG is expected to decrease.

4 Restarting with harmonic Ritz vectors

The approximate solution, with a preconditioner or an iterative method, of the correction equation 10 in the Davidson algorithm yields a correction to a specific eigenvector. The more accurate this solution, the less it contributes towards other eigenvectors. Similarly, the ability of the TR-CG scheme to target an eigenpair better impairs the convergence to the rest. When one eigenpair is needed, preconditioning is superior to gradient based methods such as non preconditioned Lanczos, and TR-CG matches and could even outperform the dynamic thick restarting. However, when several extreme eigenpairs are needed, their effectiveness diminishes with the number of eigenpairs (see experiments and comments in [21] and [6]). To face this problem it is necessary to combine these methods with some sort of thick restarting.

4.1 Why harmonic pairs?

In TR(k) the lowest k Ritz vectors are retained after restarting, as the best approximations to the k lowest eigenvectors. However, the Rayleigh-Ritz procedure guarantees optimality only for the lowest eigenpair. This optimality involves the minimization of the Rayleigh quotient over all the vectors in the subspace. The rest of the Ritz vectors can be arbitrary linear combinations of several (not necessarily nearby) eigenvectors. Therefore, the question arises whether we can restart our method with a set of different vectors that approximate the k lowest eigenvectors in a better way.

Recently, there has been a lot of discussion on the use of harmonic Ritz vectors as a way to extract more meaningful approximations from a subspace [13, 9, 6, 19], but their effect on restarting has not been examined. To extend the Rayleigh-Ritz optimality to interior eigenpairs, we have to consider a shifted and inverted operator. In the notation of the Davidson algorithm, consider the matrix $\tilde{A} = (A - \sigma I)^{-1}$, and take its Rayleigh-Ritz projection onto a subspace spanned by the non orthogonal basis: $\tilde{V} = (A - \sigma I)V = W - \sigma V$. The interior eigenvalues of A closest to σ are extreme eigenvalues of the matrix $\tilde{A}$, and therefore the Ritz pairs extracted from $\tilde{V}$ would have the Rayleigh-Ritz optimality. These Ritz pairs are called harmonic Ritz pairs. Because of the choice of subspace, the need to invert a matrix is avoided, and the small eigenvalue problem $\tilde{V}^T \tilde{A} \tilde{V} c = \lambda \tilde{V}^T \tilde{V} c$ becomes: $(H - \sigma I)c = \lambda \tilde{V}^T \tilde{V} c$. Since the vectors $W = AV$ are already stored in the JD, the only additional computation needed is $W^T W$. Efficient ways of obtaining harmonic Ritz pairs are described in [6, 9].

Because of their properties, harmonic Ritz pairs have been used for computing interior eigenpairs of a matrix [9, 6] as well as for computing the eigenvector near the origin for deflating GMRES [2]. Thus, one would expect them to be good candidates for restarting eigenvalue iterative methods.

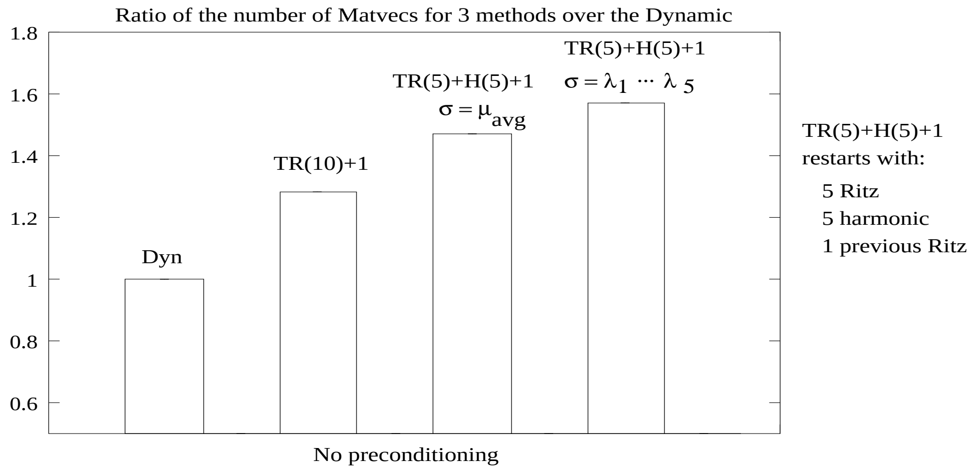

Figure 7: Comparison of combinations of TR-CG and harmonic restarting schemes versus the Dynamic. No preconditioner.

4.2 Experiments and justification

To test the use of harmonic vectors in restarting, we set up a new restarting scheme that we denote as TR(k)+H(q)+p. This scheme restarts with a combination of k lowest Ritz vectors, q lowest harmonic Ritz vectors, starting from the *second non converged* pair, and p lowest Ritz vectors from the previous step, starting from the first non converged pair. The reason for this laborious bookkeeping is that we would like to retain Ritz vectors for all converged pairs and for the extreme one. Because of deflation, the extreme pair is simply the first non converged one. For the same reason, we do not want to restart with TR-CG or harmonic vectors for the converged eigenpairs, or with a harmonic Ritz vector for the extreme one.

The choice of the shift σ plays an important role in the quality of the harmonic Ritz pairs. To ascertain that an extreme harmonic Ritz pair corresponds to a given eigenpair, the value of σ must be closer to the given eigenvalue than to any other one. This shift placement, however, is not possible because the eigenvalues are not known. Further, solving a different harmonic Ritz problem for each non converged eigenpair involves additional computations. In our experiments, where we look for 5 lowest eigenpairs, we choose $\sigma = \gamma\mu_{target} + (1-\gamma)\mu_5$, where μ's are the Ritz values from the current step, and γ a real number between 0 and 1. In other words, we choose our shift between the first non converged and the innermost eigenvalue sought. Because Ritz values are overestimates of the eiganvalues, the parameter γ is usually chosen closer to 1. We also test a best case scenario for harmonic Ritz pairs by providing perturbations of the exact eigenvalues as shifts and solving a different harmonic problem for each non converged eigenpair.

As with our previous figures, we test the symmetric Harwell-Boeing collection and we report harmonic averages of ratios of matrix vector multiplications for various restarting techniques versus the dynamic scheme. Because the new scheme can only be implemented with explicit restarting, our results are obtained with the Davidson method. We provide no timing comparisons because we want to study the effectiveness of the harmonic restarting and not the efficiency of implementation of the harmonic procedure.

In figures 7 and 8 we show the results without and with diagonal preconditioning respectively. The results are contrary to our initial expectations. First, restarting with harmonic

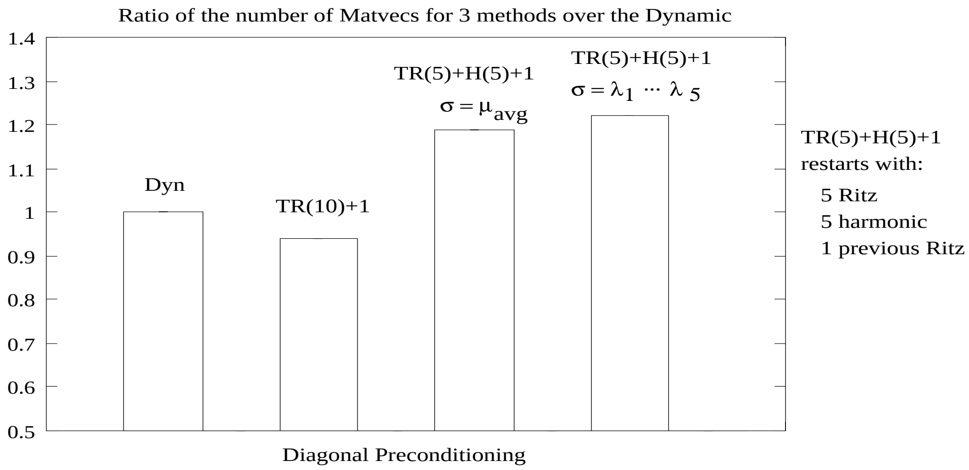

Figure 8: Comparison of combinations of TR-CG and harmonic restarting schemes versus the Dynamic. Diagonal preconditioner.

Ritz vectors is far worse than both the dynamic and TR(10)+1 schemes. Second, using the exact eigenvalues as shifts should produce better eigenvector approximations, but using these for restarting seems to be even worse than using the harmonic vectors from some approximate shift. Finally, we observed (not shown in the figures) that if we did not retain Ritz vectors for all required eigenpairs the method converged extremely slowly. The observed behavior suggests that five lowest eigenpairs is still an extreme part of the spectrum where Rayleigh-Ritz may provide better restarting information. This is especially true for most Harwell-Boeing matrices, which have spectra that are packed towards the lower end. This eigenvalue clustering may also cause eigenpair misselection in the harmonic Ritz procedure.

The next experiment examines the case where only the 5th lowest eigenpair of the matrices is needed. Our Davidson program achieves this by keeping all five lowest Ritz pairs but targeting only the fifth at each step. We use diagonal preconditioning because without preconditioning the residuals of Ritz pairs in Arnoldi are all co-directional. Figure 9 shows the results from several methods. The information is given similarly to previous figures, except for the ratios which are reported versus the TR(10)+1 scheme. Again, we include two variations of harmonic restarting. One uses the current Ritz value as the shift, and the other the perturbed exact eigenvalue. Although the differences between the methods are smaller, the results are qualitatively the same as with figures 7 and 8. Even in this interior case the use of harmonic vectors does not lead to a better restarting strategy. As an explanation we note that our Davidson program maintains all five lowest pairs, attempting to build the fifth one faster than the rest. However, the diagonal preconditioning is not powerful enough to give selective convergence in a few steps, and therefore a good approximation of the lower eigenspace must be obtained first. This lower eigenspace is approximated better by restarting with Ritz vectors than with harmonic ones.

The above results do not preclude the use of harmonic Ritz pairs. Both of the above experiments required the computation of a relatively extreme part of the spectrum (either all of this part or a slightly interior eigenpair), where Ritz vectors are known to perform well. When a highly interior eigenpair is required, the lower eigenspace is usually too large to store or even to capture approximately within a few Ritz vectors. For these problems, harmonic Ritz pairs

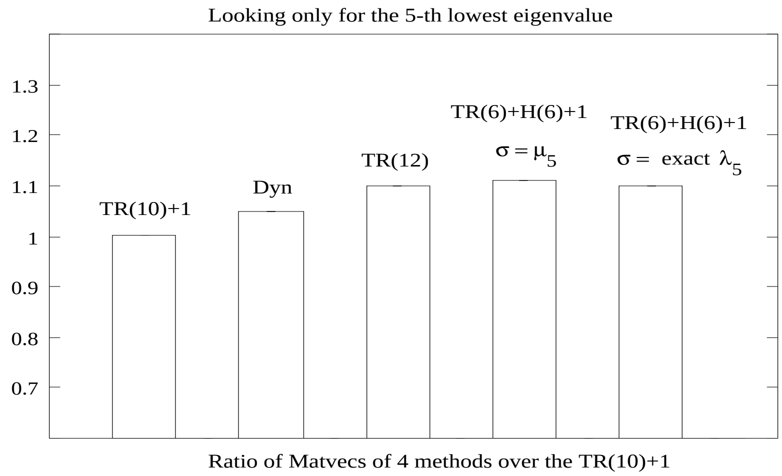

Figure 9: TR(6)+H(6)+1 versus the TR(10)+1 technique when looking only for 5th lowest pair. Diagonal preconditioner.

may be the only alternative. In [9] several examples are given where Lanczos with harmonic Ritz pairs outperforms the classical Ritz Lanczos. Similar examples with the JD can be found in [6, 19]. With preconditioning, however, the space depends on the choice of residuals, and therefore harmonic Ritz residuals should be preferred for highly interior eigenpairs.

5 Conclusions

Efficiency and robustness of iterative methods depend significantly on the restarting techniques used. For eigenvalue methods, thick, dynamic thick, TR-CG, and Leja restarting techniques have proved effective on a wide variety of matrices. The numerical experiments in this paper show that thick restarting does not benefit from keeping more than a certain number of vectors at restart. On the contrary, dynamic thick, Leja, and the TR-CG schemes can efficiently use additional vector information. We have provided a justification of this behavior by interpreting the discarded Ritz values during restart as roots of the residual polynomial. We also observe the similarities between Leja and dynamic thick restarting in dampening more uniformly a wider part of the spectrum. The effectiveness of the TR-CG scheme is attributed to its relation with solving the correction equation.

Harmonic Ritz vectors are often used as better approximations from a subspace to interior eigenvectors of a matrix. We have used these vectors in thick restarting to improve the quality of the approximations to non extreme eigenvectors. Our experiments show that this use of harmonic Ritz vectors deteriorates the convergence of iterative methods. Further, choosing the harmonic vectors with shifts close to the required eigenvalues is usually worse. Finally, this scheme does not work even when we look for an interior eigenpair within an extreme part of the spectrum. The justification of this behavior is based on the extremity as well as the clustering of the required spectrum. Ritz pairs are the best way to obtain eigenpairs within such a region.

However, harmonic Ritz pairs have been used successfully for obtaining highly interior eigenpairs. In these cases, their efficient use goes beyond restarting into building an iterative method such as the JD around the harmonic procedure.

References

[1] D. Calvetti, L. Reichel, and D. Sorensen. An implicitly restarted Lanczos method for large symmetric eigenvalue problems. *Electronic Trans. Numer. Anal.*, 2:1–21, 1994.

[2] A. Chapman and Y. Saad. Deflated and augmented Krylov subspace techniques. *Numerical Linear Algebra with Applications*, 4:43–66, 1997.

[3] M. Crouzeix, B. Philippe, and M. Sadkane. The Davidson method. *SIAM J. Sci. Comput.*, 15:62–76, 1994.

[4] J. Cullum and R. A. Willoughby. *Lanczos algorithms for large symmetric eigenvalue computations*, volume 2: Programs of *Progress in Scientific Computing; v. 4*. Birkhauser, Boston, 1985.

[5] Ernest R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comput. Phys.*, 17:87–94, 1975.

[6] D. R. Fokkema, G. L. G. Sleijpen, and H. A. van der Vorst. Jacobi-Davidson style QR and QZ algorithms for the partial reduction of matrix pencils. Technical Report 941, Department of Mathematics, University of Utrecht, 1996. To appear in SIAM J. Sci. Comput.

[7] R. B. Lehoucq. *Analysis and implementation of an implicitly restarted Arnoldi iteration.* PhD thesis, Department of Computational and Applied Mathematics, Rice University, 1995. TR95-13.

[8] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK USERS GUIDE: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods.* SIAM, Phildelphia, PA, 1998.

[9] R. B. Morgan. Computing interior eigenvalues of large matrices. *Lin. Alg. Appl.*, 154–156:289–309, 1991.

[10] R. B. Morgan. On restarting the Arnoldi method for large nonsymmetric eigenvalue problems. *Math. Comput.*, 65:1213–1230, 1996.

[11] R. B. Morgan and D. S. Scott. Generalizations of Davidson's method for computing eigenvalues of sparse symmetric matrices. *SIAM J. Sci. Comput.*, 7:817–825, 1986.

[12] C. W. Murray, S. C. Racine, and E. R. Davidson. Improved algorithms for the lowest eigenvalues and associated eigenvectors of large matrices. *J. Comput. Phys.*, 103(2):382–389, 1992.

[13] C. C. Paige, B. N. Parlett, and Van der Vorst. Approximate solutions and eigenvalue bounds from Krylov spaces. *Num. Lin. Alg. Appl.*, 2:115–133, 1995.

[14] Beresford N. Parlett. *The Symmetric Eigenvalue Problem.* SIAM, Philadelphia, PA, 1998.

[15] Yousef Saad. Chebyshev acceleration techniques for solving nonsymmetric eigenvalue problems. *Math. Comp.*, 42:567–588, 1984.

[16] Yousef Saad. *Numerical methods for large eigenvalue problems.* Manchester University Press, 1992.

[17] Yousef Saad. A flexible inner-outer preconditioned GMRES algorithm. *SIAM J. Sci. Comput.*, 14(2):461–469, March 1993.

[18] G. L. G. Sleijpen and H. A. van der Vorst. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 17(2):401–425, 1996.

[19] G. L. G. Sleijpen, H. A. van der Vorst, and E. Meijerink. Efficient expansion of subspaces in the Jacobi-Davidson method for standard and generalized eigenproblems. Technical Report 1046, Department of Mathematics, University of Utrecht, 1998.

[20] D. C. Sorensen. Implicit application of polynomial filters in a K-step Arnoldi method. *SIAM J. Matrix Anal. Appl.*, 13(1):357–385, 1992.

[21] A. Stathopoulos and Y. Saad. Restarting techniques for (jacobi-)davidson symmetric eigenvalue methods. *Electr. Trans. Numer. Alg.*, 7:163–181, 1998.

[22] A. Stathopoulos, Y. Saad, and K. Wu. Dynamic thick restarting of the Davidson, and the implicitly restarted Arnoldi methods. *SIAM J. Sci. Comput.*, 19(1):227–245, 1998.

Andreas Stathopoulos
Department of Computer Science
College of William and Mary
P.O. Box 8705
Willamsburg, VA 23187-8795
andreas@cs.wm.edu