An Implicit Riemannian Trust-Region Method for the Symmetric Generalized Eigenproblem ⋆

C. G. Baker[1,2], P.-A. Absil[3,4], and K. A. Gallivan[1]

[1] School of Computational Science, Florida State University, Tallahassee, FL, USA,
[2] Computational Mathematics & Algorithms, Sandia National Laboratories, Albuquerque, NM, USA,
[3] Département d'ingénierie mathématique, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium,
[4] Peterhouse, University of Cambridge, UK

Abstract. The recently proposed Riemannian Trust-Region method can be applied to the problem of computing extreme eigenpairs of a matrix pencil, with strong global convergence and local convergence properties. This paper addresses inherent inefficiencies of an explicit trust-region mechanism. We propose a new algorithm, the Implicit Riemannian Trust-Region method for extreme eigenpair computation, which seeks to overcome these inefficiencies while still retaining the favorable convergence properties.

1 Introduction

Consider $n \times n$ symmetric matrices A and B, with B positive definite. The generalized eigenvalue problem

$$Ax = \lambda Bx$$

is known to admit n real eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$, along with associated B-orthonormal eigenvectors $v_1, \ldots, v_n$ (see [1]). We seek here to compute the p leftmost eigenvectors of the pencil (A, B). It is known that the leftmost eigenspace $\mathcal{U} = \operatorname{colsp}(v_1, \ldots, v_p)$ of (A, B) is the column space of any minimizer of the generalized Rayleigh quotient

$$f : \mathbb{R}_*^{n \times p} \to \mathbb{R} : Y \mapsto \operatorname{trace}\left((Y^T BY)^{-1}(Y^T AY)\right), \tag{1}$$

where $\mathbb{R}_*^{n \times p}$ denotes the set of full-rank $n \times p$ matrices.

⋆ This work was supported by NSF Grant ACI0324944. The first author was in part supported by the CSRI, Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy; contract/grant number: DE-AC04-94AL85000. The second author was partially supported by Microsoft Research through a Research Fellowship at Peterhouse, Cambridge.

This result underpins a number of methods based on finding the extreme points of the generalized Rayleigh quotient (see [2–7] and references therein). Here, we consider the recently proposed [8, 9] Riemannian Trust-Region (RTR) method. This method formulates the eigenvalue problem as an optimization problem on a Riemannian manifold, utilizing a trust-region mechanism to find a solution. Similar to Euclidean trust-region methods [10, 11], the RTR method ensures strong global convergence properties while allowing superlinear convergence near the solution. However, the classical trust-region mechanism has some inherent inefficiencies. When the trust-region radius is too large, valuable time may be spent computing an update that may be rejected. When the trust-region radius is too small, we may reject good updates lying outside the trust-region. A second problem with the RTR method is typical of methods where the outer stopping criterion is evaluated only after exiting the inner iteration: in almost all cases, the last call to the inner iteration will perform more work than necessary to satisfy the outer stopping criterion.

In the current paper, we explore solutions to both of the problems described above. We present an analysis providing us knowledge of the model fidelity at every step of the inner iteration, allowing our trust-region to be based directly on the trustworthiness of the model. We propose a new algorithm, the Implicit Riemannian Trust-Region (IRTR) method, exploiting this analysis.

2 Riemannian Trust-Region Method with Newton Model

The RTR method can be used to minimize the generalized Rayleigh quotient (1). The right-hand side of this function depends only on $\mathrm{colsp}(Y)$, so that f induces a real-valued function on the set of p-dimensional subspaces of $\mathbb{R}^n$. (This set is known as the Grassmann manifold, which can be endowed with a Riemannian structure [4, 12].) The RTR method iteratively computes the minimizer of f by (approximately) minimizing successive models of f. The minimization of the models is done via an iterative process, which is referred to as the *inner iteration*, to distinguish it with the principal *outer iteration*. We present here the process in a way that does not require a background in differential geometry; we refer to [13] for the mathematical foundations of the technique.

Let Y be a full-rank, $n \times p$ matrix. We desire a correction S of Y such that $f(Y + S) < f(Y)$. A difficulty is that corrections of Y that do not modify its column space do not affect the value of the cost function. This situation leads to unpleasant degeneracy if it is not addressed. Therefore, we require S to satisfy some complementarity condition with respect to the space $\mathcal{V}_Y := \{YM : M \in \mathbb{R}^{p\times p}\}$. Here, in order to simplify later developments, we impose complementarity via B-orthogonality, namely $S \in \mathcal{H}_Y$ where

$$\mathcal{H}_Y = \{Z \in \mathbb{R}^{n\times p} : Y^T B Z = 0\}.$$

Consequently, the task is to minimize the function

$$\hat{f}_Y(S) := \mathrm{trace}\left(((Y+S)^T B (Y+S))^{-1}((Y+S)^T A (Y+S))\right), \quad S \in \mathcal{H}_Y.$$

The RTR method constructs a model m_Y of $\hat{f}_Y$ and computes an update S which approximately minimizes m_Y, so that the inner iteration attempts to solve the following problem:

$$\min m_Y(S), \quad S \in \mathcal{H}_Y, \quad \|S\|_2 \leq \Delta,$$

where Δ (the *trust-region radius*) denotes the region in which we trust m_Y to approximate $\hat{f}_Y$. The next iterate and trust-region radius are determined by the performance of m_Y with respect to $\hat{f}_Y$. This performance ratio is measured by the quotient:

$$\rho_Y(S) = \frac{\hat{f}_Y(0) - \hat{f}_Y(S)}{m_Y(0) - m_Y(S)}.$$

Low values of $\rho_Y(S)$ (close to zero) indicate that the model m_Y at S is not a good approximation to $\hat{f}_Y$. In this scenario, the trust-region radius is reduced and the update $Y + S$ is rejected. Higher values of $\rho_Y(S)$ allow the acceptance of $Y + S$ as the next iterate, and a value of $\rho_Y(S)$ close to one suggests good approximation of $\hat{f}_Y$ by m_Y, allowing the trust-region radius to be enlarged.

Usually, the model m_Y is chosen as a quadratic function approximating $\hat{f}_Y$. In the sequel, in contrast to [9] where the quadratic term of the model was unspecified, we assume that m_Y is the *Newton model*, i.e., the quadratic expansion of $\hat{f}_Y$ at $S = 0$. Then, assuming from here on that $Y^T BY = I_p$, we have

$$\begin{aligned} m_Y(S) &= \operatorname{trace}\left(Y^T AY\right) + 2\operatorname{trace}\left(S^T AY\right) + \operatorname{trace}\left(S^T\left(AS - BS(Y^T AY)\right)\right) \\ &= \hat{f}_Y(0) + \operatorname{trace}\left(S^T \nabla \hat{f}_Y\right) + \frac{1}{2}\operatorname{trace}\left(S^T H_Y[S]\right), \end{aligned}$$

where the gradient and the effect of the Hessian of $\hat{f}_Y$ are identified as

$$\nabla \hat{f}_Y = 2P_{BY} AY \qquad H_Y[S] = 2P_{BY}\left(AS - BS(Y^T AY)\right),$$

and where $P_{BY} = I - BY(Y^T BBY)^{-1} Y^T B$ is the orthogonal projector on the space perpendicular to the column space of BY.

Simple manipulation shows the following:

$$\begin{aligned} \hat{f}_Y(0) - \hat{f}_Y(S) &= \operatorname{trace}\left(Y^T AY - (I + S^T BS)^{-1}(Y + S)^T A(Y + S)\right) \\ &= \operatorname{trace}\left((I + S^T BS)^{-1}(S^T BS(Y^T AY) - 2S^T AY - S^T AS)\right). \end{aligned}$$

Consider the case where $p = 1$. The above equation simplifies to

$$\begin{aligned} \hat{f}_y(0) - \hat{f}_y(s) &= (1 + s^T Bs)^{-1}\left(s^T Bsy^T Ay - 2s^T Ay - s^T As\right) \\ &= (1 + s^T Bs)^{-1}\left(m_y(0) - m_y(s)\right), \end{aligned}$$

so that

$$\rho_y(s) = \frac{\hat{f}_y(0) - \hat{f}_y(s)}{m_y(0) - m_y(s)} = \frac{1}{1 + s^T Bs}. \tag{2}$$

This allows the model performance ratio ρ_y to be constantly evaluated as the model minimization progresses, simply by tracking the B-norm of the current update vector.

3 Implicit Riemannian Trust-Region Method

In this section, we explore the possibility of selecting the trust-region as a sublevel set of the performance ratio ρ_Y. We dub this approach the Implicit Riemannian Trust-Region method.

3.1 Case $p = 1$

The analysis of ρ in the previous section shows that for the generalized Rayleigh quotient with $p = 1$, the performance of the model decreases as the iterate moves away from zero. However, in the case of the $p = 1$ generalized Rayleigh quotient, $\rho_y(s)$ has a simple relationship with $\|s\|_B$. Therefore, by monitoring the B-norm of the inner iterate, we can easily determine the value of ρ for a given inner iterate. Furthermore, the relationship between ρ and the B-norm of a vector, allows us to move along a search direction to a specific value of ρ. These two things, combined, enable us to redefine the trust-region based instead on the value of ρ.

The truncated conjugate gradient proposed in [9] for use in the simple RTR algorithm seeks to minimize the model m_Y within a trust-region defined explicitly as $\{s : \|s\|_2 \leq \Delta\}$. Here, we change the definition of the trust-region to $\{s : \rho_y(s) \geq \rho'\}$, for some $\rho' \in (0, 1)$. The necessary modifications to this algorithm are very simple. The definition of the trust-region occurs in three places: when detecting whether the trust-region has been breached; when constraining the update vector in the case that the trust-region was breached; and when constraining the update vector in the case that we have detected a direction of negative curvature. The new inner iteration is listed in Algorithm 1, with the differences highlighted.

Having stated the definition of the implicit trust-region, based on ρ, we need a mechanism for following a search direction to the edge of the trust-region. That is, at some outer step k and given s_j and a search direction d_j, we wish to compute $s = s_j + \tau d_j$ such that $\rho_{y_k}(s) = \rho'$. Given ρ' and denoting

$$\Delta_{\rho'} = \sqrt{\frac{1}{\rho'} - 1}, \tag{3}$$

the desired value of τ is given by

$$\tau = \frac{-d_j^T B s_j + \sqrt{(d_j^T B s_j)^2 + d_j^T B d_j (\Delta_{\rho'}^2 - s_j^T B s_j)}}{d_j^T B d_j}. \tag{4}$$

A careful implementation precludes the need for any more matrix multiplications against B than are necessary to perform the iterations.

Another enhancement in Algorithm 1 is that the outer stopping criterion is tested during the inner iteration. This technique is not novel in the context of eigensolvers with inner iterations, having been proposed by Notay [14]. Our motivation for introducing this test is that, when it is absent, the final outer

Algorithm 1 (Preconditioned Truncated CG (IRTR))

Data: A,B symmetric, B positive definite, $\rho' \in (0,1)$*, preconditioner M*

Input: Iterate y, $y^T B y = 1$

Set $s_0 = 0$, $r_0 = \nabla \hat{f}_y$, $z_0 = M^{-1} r_0$, $d_0 = -z_0$

for $j = 0, 1, 2, \ldots$

    Check κ/θ ***stopping criterion***

    if $\|r_j\|_2 \leq \|r_0\|_2 \min\left\{\kappa, \|r_0\|_2^\theta\right\}$

        return s_j

    Check curvature of current search direction

    if $d_j^T H_y[d_j] \leq 0$

        Compute τ *such that* $s = s_j + \tau d_j$ *satisfies* $\boxed{\rho_y(s) = \rho'}$

        return s

    Set $\alpha_j = (z_j^T r_j)/(d_j^T H_y[d_j])$

    Generate next inner iterate

    Set $s_{j+1} = s_j + \alpha_j d_j$

    Check implicit trust-region

    if $\boxed{\rho_y(s_{j+1}) < \rho'}$

        Compute $\tau \geq 0$ *such that* $s = s_j + \tau d_j$ *satisfies* $\boxed{\rho_y(s) = \rho'}$

        return s

    Use CG recurrences to update residual and search direction

    Set $r_{j+1} = r_j + \alpha_j H_y[d_j]$

    Set $z_{j+1} = M^{-1} r_{j+1}$

    Set $\beta_{j+1} = (z_{j+1}^T r_{j+1})/(z_j^T r_j)$

    Set $d_{j+1} = -z_{j+1} + \beta_{j+1} d_j$

    Check outer stopping criterion

    $\boxed{\text{Compute } \|\nabla \hat{f}_{y+s_{j+1}}\|_2 \text{ and test}}$

end for.

step may reach a much higher accuracy than specified by the outer stopping criterion, resulting in a waste of computational effort. Also, while Notay proposed a formula for the inexpensive evaluation of the outer norm based on the inner iteration, we must rely on a slightly more expensive, but less frequent, explicit evaluation of the outer stopping criterion.

The product of this iteration is an update vector s_j which is guaranteed to lie inside of the ρ-based trust-region. The result is that the ρ value of the new iterate need not be explicitly computed, the new iterate can be automatically accepted, with an update vector constrained by model fidelity instead of a discretely chosen trust-region radius based on the performance of the last iterate. An updated outer iteration is presented in Algorithm 2, which also features an optional subspace acceleration enhancement à la Davidson [15].

Algorithm 2 (Implicit Riemannian Trust-Region Algorithm)
Data: A,B symmetric, B positive definite, $\rho' \in (0,1)$
Input: Initial subspace $\mathcal{W}_0$
for $k = 0,1,2,\ldots$

Model-based Minimization
Generate y_k *using a Rayleigh-Ritz procedure on* $\mathcal{W}_k$
Compute $\nabla \hat{f}_{y_k}$ *and check* $\|\nabla \hat{f}_{y_k}\|_2$
Compute s_k *to approximately minimize* m_{y_k} *such that* $\rho(s_k) \geq \rho'$ *(Algorithm 1)*

Generate next subspace
if *performing subspace acceleration*
Compute new acceleration subspace $\mathcal{W}_{k+1}$ *from* $\mathcal{W}_k$ *and* s_k
else
Set $\mathcal{W}_{k+1} = \operatorname{colsp}(y_k + s_k)$
end

end for.

3.2 A Block Algorithm

The analysis of Section 2 seems to preclude a simple formula for ρ in the case that $p > 1$. We wish, however, to have a block algorithm. The solution is to decouple the block Rayleigh quotient into the sum of p separate rank-1 Rayleigh quotients, which can then be addressed individually using the IRTR strategy. This is done as follows.

Assume that our iterates satisfy $Y^T AY = \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_p)$, in addition to $Y^T BY = I_p$. In fact, this is a natural consequence of the Rayleigh-Ritz process. Then given $Y = \begin{bmatrix} y_1 \ldots y_p \end{bmatrix}$, the model m_Y can be rewritten:

$$\begin{aligned} m_Y(S) &= \operatorname{trace}\left(\Sigma + 2S^T AY + S^T(AS - BS\Sigma)\right) \\ &= \sum_{i=1}^{p} \left(\sigma_i + 2s_i^T Ay_i + s_i^T(A - \sigma_i B)s_i\right) = \sum_{i=1}^{p} m_{y_i}(s_i). \end{aligned}$$

It should be noted that the update vectors for the decoupled minimizations must have the original orthogonality constraints in place. That is, instead of requiring only that $y_i^T Bs_i = 0$, we require that $Y^T Bs_i = 0$ for each s_i. This is necessary to guarantee that the next iterate, $Y + S$, has full rank, so that the Rayleigh quotient is defined.

As for the truncated conjugate gradient, the p individual IRTR subproblems should be solved simultaneously, with the inner iteration stopped as soon as any of the iterations satisfy one of the inner stopping criteria (exceeded trust-region or detected negative curvature). If only a subset of iterations are allowed to continue, then the κ/θ inner stopping criterion may not be feasible.

The described method attempts to improve on the RTR, while retaining the strong global and local convergence properties of the RTR. The model fidelity guaranteed by the implicit trust-region mechanism allows for a straightforward

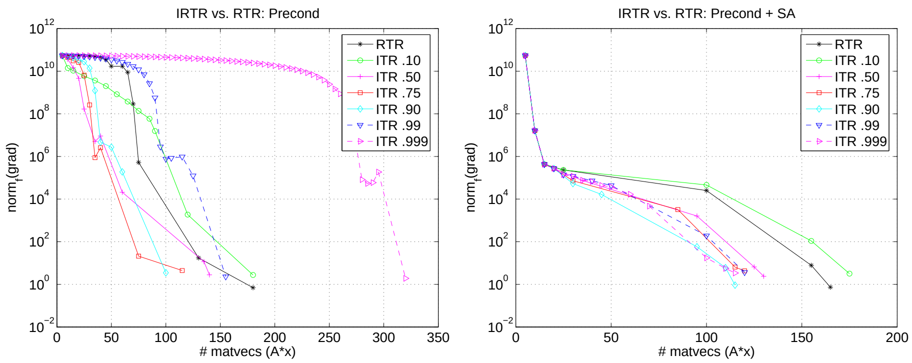

Fig. 1. Figures illustrating the efficiency of RTR vs. IRTR for different values of ρ', in the presence of a preconditioned inner iteration, for the BCSST24 data.

proof of global convergence. Related work [16] presents the proofs of global convergence, along with a discussion regarding the consequences of early termination of the inner iteration due to testing the outer stopping criterion and an exploration of the RTR method in light of the ρ analysis presented here.

4 Numerical Results

The IRTR method seeks to overcome the inefficiencies of the RTR method, such as the rejection of computed updates and the limitations due to the discrete nature of the trust-region radius. We compare the performance of the IRTR with that of the classical RTR. The following experiments were performed in MATLAB (R14) under Mac OSX. Figure 1 considers a generalized eigenvalue problem with a preconditioned inner iteration. The matrices A and B are from the Harwell-Boeing collection BCSST24. The problem is of size $n = 3562$ and we are seeking the leftmost $p = 5$ eigenvalues. The inner iteration is preconditioned using an exact factorization of A. Two experiments are run: with and without subspace acceleration. When in effect, the subspace acceleration strategy occurs over the 10-dimensional subspace $\mathrm{colsp}([Y_k, S_k])$. The RTR is tested with a value of $\rho' = 0.1$, while the IRTR is run for multiple values of ρ'. These experiments demonstrate that the IRTR method is able to achieve a greater efficiency than the RTR method.

5 Conclusion

This paper presents an optimization-based analysis of the symmetric, generalized eigenvalue problem which explores the relationship between the inner and outer iterations. The paper proposes the Implicit Riemannian Trust-Region method, which seeks to alleviate inefficiencies resulting from the inner/outer divide, while still preserving the strong convergence properties of the RTR method. This algorithm was shown in numerical experiments to be capable of greater efficiency than the RTR method.

Acknowledgments Useful discussions with Andreas Stathopoulos, Rich Lehoucq and Ulrich Hetmaniuk are gratefully acknowledged.

References

1. Stewart, G.W.: Matrix algorithms, Vol II: Eigensystems. Society for Industrial and Applied Mathematics, Philadelphia (2001)
2. Sameh, A.H., Wisniewski, J.A.: A trace minimization algorithm for the generalized eigenvalue problem. SIAM J. Numer. Anal. **19**(6) (1982) 1243–1259
3. Smith, S.T.: Optimization techniques on Riemannian manifolds. In: Hamiltonian and gradient flows, algorithms and control. Volume 3 of Fields Inst. Commun. Amer. Math. Soc., Providence, RI (1994) 113–136
4. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**(2) (1998) 303–353
5. Mongeau, M., Torki, M.: Computing eigenelements of real symmetric matrices via optimization. Comput. Optim. Appl. **29**(3) (2004) 263–287
6. Sameh, A., Tong, Z.: The trace minimization method for the symmetric generalized eigenvalue problem. J. Comput. Appl. Math. **123** (2000) 155–175
7. Knyazev, A.V.: Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method. SIAM J. Sci. Comput. **23**(2) (2001) 517–541
8. Absil, P.-A., Baker, C.G., Gallivan, K.A.: Trust-region methods on Riemannian manifolds with applications in numerical linear algebra. In: Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS2004), Leuven, Belgium, 5–9 July 2004. (2004)
9. Absil, P.A., Baker, C.G., Gallivan, K.A.: A truncated-CG style method for symmetric generalized eigenvalue problems. J. Comput. Appl. Math. **189**(1–2) (2006) 274–285
10. Moré, J.J., Sorensen, D.C.: Newton's method. In: Studies in numerical analysis. Volume 24 of MAA Stud. Math. Math. Assoc. America, Washington, DC (1984) 29–82
11. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operations Research. Springer-Verlag, New York (1999)
12. Absil, P.-A., Mahony, R., Sepulchre, R.: Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. Acta Appl. Math. **80**(2) (2004) 199–220
13. Absil, P.-A., Baker, C.G., Gallivan, K.A.: Trust-region methods on Riemannian manifolds. submitted (2005)
14. Notay, Y.: Combination of Jacobi-Davidson and conjugate gradients for the partial symmetric eigenproblem. Numer. Linear Algebra Appl. **9**(1) (2002) 21–44
15. Sleijpen, G.L.G., Van der Vorst, H.A.: A Jacobi-Davidson iteration method for linear eigenvalue problems. SIAM J. Matrix Anal. Appl. **17**(2) (1996) 401–425
16. Baker, C.G., Absil, P.-A., Gallivan, K.A.: An implicit Riemannian trust-region method for the symmetric generalized eigenproblem. Technical Report FSU-SCS-2006-152, School of Computational Science, Florida State University (2006) http://scseprints.scs.fsu.edu.