BIT 36:1 (1996), 001–003.

ON CONJUGATE GRADIENT-LIKE METHODS FOR EIGEN-LIKE PROBLEMS*

A. EDELMAN† S. T. SMITH‡

[1]*Department of Mathematics*
Massachusetts Institute of Technology
Cambridge, MA 02139 USA
email: `edelman@math.mit.edu`

[2]*Lincoln Laboratory*
Massachusetts Institute of Technology
Lexington, MA 02173 USA
email: `stsmith@ll.mit.edu`

AMS subject classification: 65F15, 49M07, 49M15, 53B20.

Key words: Conjugate gradient, Lanczos, Newton's Method, Optimization, Signal Processing, Electronic Structures, Differential Geometry

Abstract.

Numerical analysts, physicists, and signal processing engineers have proposed algorithms that might be called conjugate gradient for problems associated with the computation of eigenvalues. There are many variations, mostly one eigenvalue at a time though sometimes block algorithms are proposed. Is there a correct "conjugate gradient" algorithm for the eigenvalue problem? How are the algorithms related amongst themselves and with other related algorithms such as Lanczos, the Newton method, and the Rayleigh quotient?

1 Introduction

If we minimize y^TAy on the unit sphere, perhaps with conjugate gradient optimization, we compute the smallest eigenvalue of A (assumed to be symmetric). Our objective function is quadratic, so we obtain the exact eigenvalue after n steps of conjugate gradient. The computation is mathematically equivalent to the Lanczos procedure. At the kth step, we obtain the optimal answer in a k dimensional Krylov space.

*Received November 1995. Revised March 1996.

†Supported by a fellowship from the Alfred P. Sloan Foundation and NSF Grant 9404326-CCR.

‡This work was sponsored by DARPA under Air Force contract F19628-95-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Air Force.

This paper will also appear in the Proceedings of the AMS/IMS/SIAM Joint Summer Research Conference on Linear and Nonlinear Conjugate Gradient-Related Methods held in Seattle, 9–13 July 1995.

The above paragraph may sound correct, but it is wrong! The objective function is quadratic, but only because of the constraint. The Rayleigh quotient $r(y) = y^TAy/y^Ty$ can be minimized without constraint, but it is not quadratic. The conjugate gradient link with the Lanczos process here is also dubious. How tempting it is to hear "eigenvalue" juxtaposed with "conjugate gradient" and instantly, almost by reflex, yell "Lanczos." To add to the confusion, conjugate gradient optimization of the Rayleigh quotient does compute an answer in a Krylov space, but not the optimal answer.

Our purpose in this paper is to 1) dispel a common myth, 2) create an intellectual framework tying together the plethora of conjugate gradient algorithms, 3) provide a scholarly review of the literature, and 4) show how the differential geometry point of view introduces new algorithms in a context that allows intellectual ties to older algorithms. We believe each of these purposes is justified, but the most important goal is to show that the differential geometry point of view gives idealized algorithms that unify the subject. Therefore in Section 1 we present the myth. In Sections 2 and 3 we dispel the myth and identify the common features and design choices for a large class of algorithms. Section 4 reviews the literature in the context of this framework. Section 5 discusses the differential geometry methods in this framework. Finally Section 6 presents an entirely new perspective on Newton's method and discusses connections to previous work by Chatelin, Demmel, and others.

2 Conjugate Gradient: linear, nonlinear, and idealized

Our first task is to eliminate any confusion among the various conjugate gradient algorithms with the Lanczos algorithm. This confusion has appeared in the literature. We hope the naming convention of Table 1 will help distinguish the algorithms in the future.

The three conjugate gradient algorithms are equivalent for the special case of quadratic objective functions with positive definite Hessians. To say this in a slightly different way, the LCG algorithm, already a special case of the more general NCG algorithm, may be derived with the "one line minimization per iteration" point of view, and nevertheless, the algorithm has the "global k-dimensional" property of an ICG anyway.

To contrast, if the function is not quadratic with positive definite Hessian, there is no expectation that the kth step of NCG will have the ICG property of being the global minimum in a k-dimensional space. The best that we hope for is that the ICG property holds approximately as we approach the minimum.

Similar to ICG is s-step steepest descent. This algorithm minimizes an objective function in the s-dimensional space spanned by the most recent s gradients. If the gradients fall in a certain natural Krylov space, then there is no distinction between ICG and s-step steepest descent.

The word *conjugacy* merits discussion. For LCG, all search directions are all conjugate with respect to the Hessian matrix A, i.e., $p_i^TAp_j = 0$, $i \neq j$. For NCG, consecutive directions are conjugate with respect to the function's Hessian matrix; Fletcher-Reeves and Polak-Ribière approximate this conjugacy.

Table 1.

Conjugate Gradient Algorithms	
LCG	**Linear Conjugate Gradient:** Standard method for solving $Ax = b$, where A is a symmetric positive definite matrix. Found in all numerical linear algebra references. May be viewed as minimizing the objective function $f(x) = \frac{1}{2}x^TAx - x^Tb$.
NCG	**Nonlinear Conjugate Gradient:** A well-known algorithm often used for unconstrained minimization. Found in all numerical optimization texts. May be used to minimize very general objective functions. Each step in the algorithm requires the solution to a one dimensional line minimization problem.
ICG	**Idealized Conjugate Gradient:** A fictional algorithm introduced here for purposes of exposition. We define ICG as an algorithm whose kth iterate x_k satisfies $f(x_k) = \min_{x\in\mathcal{K}_k} f(x)$, where $\mathcal{K}_k$ denotes a k dimensional Krylov space. The Krylov spaces are nested.

Generally, ICG search directions are not conjugate; perhaps a different name is appropriate. However, ICG can be viewed as an extension of LCG.

Block versions of NCG and ICG may be defined. A block NCG replaces a 1-dimensional line search with a p-dimensional line search. A block ICG maximizes over block Krylov spaces of size p, $2p$, $3p$, etc.

Consider the case when the objective function is the Rayleigh quotient, $f(y) = (y^TAy)/(y^Ty)$.

- An NCG algorithm for Rayleigh quotient optimization computes an answer in a Krylov space. This answer is not the optimal choice from that subspace. Therefore the NCG algorithm is not an ICG algorithm. The exact eigenvalue will not be obtained in n steps. Such an algorithm is not equivalent to Lanczos.

- The Lanczos algorithm (including the computation of the tridiagonal's smallest eigenvalue) is an ICG algorithm for the Rayleigh quotient. Lanczos does compute the exact eigenvalue in n steps. This algorithm is not an NCG algorithm because it is not equivalent to an algorithm that performs line searches.

- There is also the LCG link with Lanczos through tridiagonalization. The residual vectors computed by conjugate gradient are (up to normalization) Lanczos vectors that tridiagonalize the matrix. The link is through the tridiagonalization and not the eigenvalue computation. There is no link to the NCG algorithm for eigenvalue computation.

Also we note that

- any algorithm that requires that the symmetric matrix A be positive definite is not a correct NCG algorithm for the Rayleigh quotient. Since all derivatives of the objective function are invariant under shifts, so should be the algorithm.

Understanding this last point clearly is the key sign that the reader understands the distinction between NCG for the Rayleigh quotient and LCG which is an optimization on quadratic functions. LCG requires positive definite quadratic optimization functions so that a minimum exists. The Rayleigh quotient always has a minimum for symmetric A; no positive definite condition is needed.

The first choice in our design space for algorithms is to consider whether to include the constraints or work with the Rayleigh quotient. The second choice is whether to have a one eigenvalue algorithm or a block algorithm. We indicate these choices below:

Rayleigh Quotients	
Unconstrained one eigenvalue Rayleigh quotient	$r(y) = (y^T Ay)/(y^T y)$
Constrained Rayleigh quotient	$r(y) = y^T Ay \; (y^T y = 1)$
Unconstrained block Rayleigh quotient	$R(Y) = \mathrm{tr}(Y^T AY)(Y^T Y)^{-1}$
Constrained block Rayleigh quotient	$R(Y) = \mathrm{tr} Y^T AY \quad (Y^T Y = I_p)$

In the above box Y denotes an $n \times p$ matrix. Of the four choices, only the constrained $r(y) = y^T Ay$ is (generically) non-degenerate in that the global minima are points in n dimensions, as opposed to lines or subspaces.

The third choice in our design space for an NCG eigenvalue algorithm is how to pick the new search direction. The various options are

NCG SEARCH DIRECTION CHOICES	
FR	The Fletcher-Reeves approach
PR	The Polak-Ribière approach
CA	Conjugacy through the matrix A
CSH	Conjugacy through the singular free space Hessian H
CCH	Conjugacy through a non-singular constrained Hessian H

All of these choices, except for CA, are reasonable. There is no sound mathematical justification for CA, though it has been proposed in the literature.

3 Comparison between NCG and Lanczos

Since Lanczos is an ICG algorithm for the Rayleigh quotient, why would anyone consider the construction of an NCG algorithm which is guaranteed to give a worse answer for the same number of matrix vector multiplies? Ignoring the extra storage that may be needed in the Lanczos algorithm, the more practical point is that in many algorithms it is not the eigenvalues of a matrix that are desired at all.

In the applications of interest, a matrix may be changing in "space" or time. What we want from an algorithm is to be able to track the changes. In many such applications, the problem sizes are huge so computational efficiency is of utmost importance. We have two particular examples in mind. In the Local Density Approximation (LDA) to Schrödinger's equation, the function to be minimized is not the Rayleigh quotient at all, but rather the LDA energy. In many applications, the energy function has the same degeneracy as the Rayleigh quotient in that it depends only on the span of the columns of Y rather than all elements of Y. Lanczos would not directly apply unless one wants to pretend that the function really is the Rayleigh quotient for a few iterations. Such pretense has lead to instabilities in the past.

Another class of problems arise in signal processing where the matrix is changing in time. Once again, Lanczos does not directly apply, unless one wants to pretend that the matrix is constant for a few intervals.

4 A History of CG for eigen-like problems

The purpose of this section is to demonstrate the need for a unifying theory of conjugate gradient algorithms for the eigenproblem by giving an annotated chronology of the references. We note that algorithms have been proposed by researchers in a number of disciplines, so it seems likely that researchers in one field may well be unaware of those from another.

In order to keep the chronology focused, we only included papers that linked conjugate gradient to eigenvalue or eigenvalue-like problems. We found it convenient to divide the papers into five broad categories:

- **Single Eigenvalue (NCG) Algorithms** This refers to the one eigenvalue at a time algorithm that was originally proposed by Bradbury and Fletcher [4]. This algorithm minimizes $r(y)$. Other eigenvalues may be obtained by deflation. The historical development covers a number of points in the design space.

- **Block (NCG) Algorithms** These are attempts at computing multiple eigenvalues simultaneously using a conjugate gradient style algorithm. The objective function is $R(Y)$. It is our point of view that none of these algorithms are a true conjugate gradient algorithm, though many important features of a true algorithm may be recognized in either explicit or rudimentary form in the proposed algorithms. The block case exhibits difficulties not found in the single eigenvalue case, because not only is there the orthogonality constraint $Y^TY = I$, but also there is the "Grassmann equivalence" $R(Y) = R(YQ)$ for orthogonal matrices Q. This degeneracy may well be overcome without differential geometry, but one effective and beautiful theoretical approach towards working with these equivalence classes is the total space/base space viewpoint from differential geometry.

- **ICG: The Lanczos Link** Here our goal was not to mention papers that were connected to the Lanczos algorithm, but papers that looked at the Lanczos algorithm as an optimization algorithm with something of a conjugate gradient flavor. As mentioned earlier, the famous link between Lanczos and LCG is of no relevance here and is not included.

- **Differential Geometry Viewpoint** These papers make the link between numerical linear algebra, optimization, and differential geometry. They give a new theoretical viewpoint on the algorithms discussed in the other areas. This viewpoints shows that constrained algorithms may be implemented without explicit constraints. Ultimately, they answer the question of what it *means* to do conjugate gradient minimization for the eigenvalue problem.

- **Application I: LDA Schrödinger's Equation** Rightly or wrongly, conjugate gradient has become an extremely popular new method for scientists working with the local density approximation to Schrödinger's Equation. Here the problem is akin to minimizing the block Rayleigh quotient, but it is in fact more complicated. Many researchers are currently trying variations on the same theme. We only need to list the few that we were most aware of. One point of view is that they are computing eigenvalues of a matrix that is in some sense varying with space.

- **Application II: Adaptive Signal Processing** Here the algorithm is an eigenvalue algorithm, but the matrix may be thought of as changing with time. The goal is to do subspace tracking. All the papers in this category have already been listed in a previous category, but we thought it worthwhile to collect these papers under one heading as well.

Chronology of Algorithms

Key to search direction choices: FR=Fletcher-Reeves, PR=Polak-Ribière, CA=Conjugacy through A, CSH=Conjugacy through singular Hessian H, CCH= Conjugacy through constrained Hessian H.

1 Single Eigenvalue Algorithms

1966	Bradbury and Fletcher [4]	Notes degeneracy of Rayleigh quotient. Proposes projection to ∞- and 2-norm unit spheres. FR.
1969	Fox and Kapoor [15]	Application in Structural Dynamics. Downplays importance of the constraint. FR.
1969	Fried [16]	Application in Structural Dynamics. FR.
1971	Andersson [2]	Compares two norms mentioned above [4].
1971	Geradin [20]	Unconstrained CSH.
1972	Fried [17]	Unconstrained FR.
1974	Ruhe [30]	Systematically compares above CG algorithms with other non-CG algorithms. For CG, prefers unconstrained approach in [15] and [20]. Observes spectral influence on convergence.
1984	Fuhrmann and Liu [19]	Precursor to intrinsic geometry: Searches along geodesics. FR.
1986	Perdon and Gambolati [29]	CA with a proposed preconditioner.
1986	Chen, Sarkar, et al. [7]	Rederives known algorithms. CA.
1987	Haimi-Cohen and Cohen [23]	Rederives results such as those in [4].
1989	Yang, Sarkar, and Arvas [39]	Considers effect of normalization. Concludes that their 1986 proposal is not competitive with CSH or other choices. FR or PR style. [38].

2 Block Algorithms:

1971	Alsén [1]	Assumes A positive definite. Exact line maximization (not minimization) replaced by one step of orthogonal iteration. Corresponding minimization algorithm would require orthogonal inverse iteration. FR.
1982	Sameh and Wisniewski [31]	Not exactly a CG method but very similar. Constrained objective function with unconstrained block line searches. Section 2.3 suggests block line minimization on unconstrained function as a least squares problem to be solved with LCG. Makes the Lanczos style link through simultaneous iteration.
1995	Fu and Dowling [18]	Optimization of unconstrained $R(Y)$. Block search directions arranged column by column. Projects back to constraint surface with Gram Schmidt. CA.

3 ICG: The Lanczos Link:

1951	Karush [24]	An ICG algorithm (with restarts) for the Rayleigh quotient. Not explicitly identified with Lanczos or CG.
1974	Cullum and Donath [9, 10]	Identifies block Lanczos as a block ICG.
1978	Cullum [8]	Shows that block NCG on the Rayleigh quotient computes a (generally non-optimal) answer in the block Krylov space over which the ICG would finds the minimum.
1985	Cullum and Willoughby [11]	Chapter 7 summarizes optimization interpretation. CG used at times to mean NCG and other times ICG.

4 Differential Geometry Approach:

1993	Smith [32, 33]	Introduces Differential Geometry viewpoint. Hessian is replaced with second covariant derivative. Line searches replaced with geodesic searches.
1995	Edelman, Arias, Smith [14]	Works out details of the Grassmann manifold and Stiefel manifold approaches, including the total space-base space formulation, parallel transport, differing metrics, and the linear algebra links.

5 LDA Schrödinger's Equation

1989	Gillan [21]	Projects onto constraint space, search in tangent space, suggests NCG preconditioning, FR.
1989	Štich, Car, et al. [35]	Minimizes Rayleigh quotient only (not LDA energy) for simplicity.
1989	Teter, Payne, and Allan [37]	
1992	Arias [3]	Forms unconstrained analogue of Rayleigh quotient for LDA.
1992	Payne, Teter, Allan et al. [28]	Survey of minimization in LDA.
1993	Kresse and Hafner [25]	Molecular dynamics of liquid metals.
1994	Sung, Kawai, and Weare [36]	Computes structure of Lithium.

6 Adaptive Signal Processing

1986	Chen, Sarkar, et al. [7]	See above.
1989	Yang, Sarkar, and Arvas [39]	See above.
1995	Fu and Dowling [18]	See above.

There are many closely related ideas such as steepest descents or coordinate relaxation that are not discussed here. It is instructive to consider the subtle variations between similar algorithms, as well as the key differences between algorithms that sound the same, but are quite different.

Almost all the algorithmic ideas carry over to the generalized eigenvalue problem, as was recognized by most of the authors. For purposes of exposition, we will only discuss the ordinary eigenvalue problem in this paper. We have chosen to confine our discussion to conjugate gradient methods for the general eigenproblem. The larger history of conjugate gradient methods may be found in the survey by Golub and O'Leary [22]. The search direction choice(s) appear in the annotations.

5 The Differential Geometry Viewpoint for NCG

In a recent paper [14], we proposed an algorithm to perform minimization on the block Rayleigh quotient that has appealing theoretical properties. Our algorithm takes into account both the constraints $Y^TY = I_p$ and the degeneracy of the function ($R(Y) = R(YQ)$ for orthogonal $p \times p$ matrices Q) in a clean manner. Following the lead proposed by Smith [32, 33], we derived an algorithm based on intrinsic geometry that is practical when expressed in extrinsic coordinates.

Like French-English false cognates (*faux amis*), there are a few terms used both in optimization and differential geometry with somewhat different meanings. This may cause confusion to readers of our paper [14] who are specialists in optimization.

Optimization vs. Differential Geometry

Metric

Optimization:	Sometimes the inner product defined by the Hessian or its inverse.
Diff. Geom.:	Any positive definite inner product defined on a tangent space from which all geometrical information about a space may be derived.

Curvature

Optimization:	Hessian. Usage refers to non-linearity of the graph of a function with nonzero Hessian.
Diff. Geom.:	A rank four tensor that refers roughly to the non-flatness of a space.

The Grassmann algorithm [14] does not suffer from any of the deficiencies of the other block algorithms for the eigenvalue problem. There is no requirement that the matrix be positive definite, there is no theoretical difficulty related to the singular Hessian, and the algorithm converges quadratically.

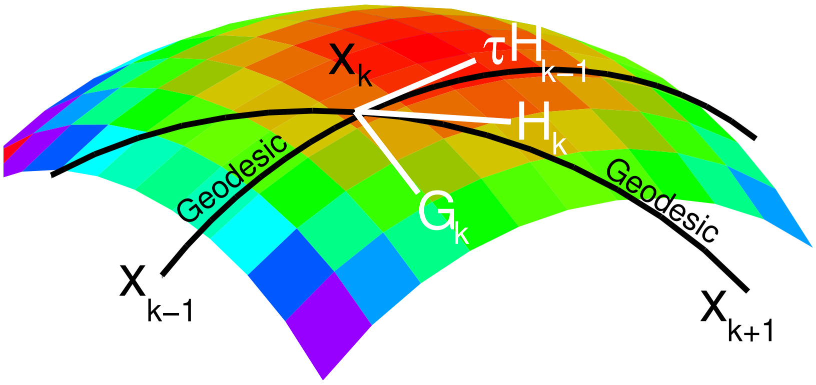

Figure 5.1: Conjugate gradient on Riemannian manifolds.

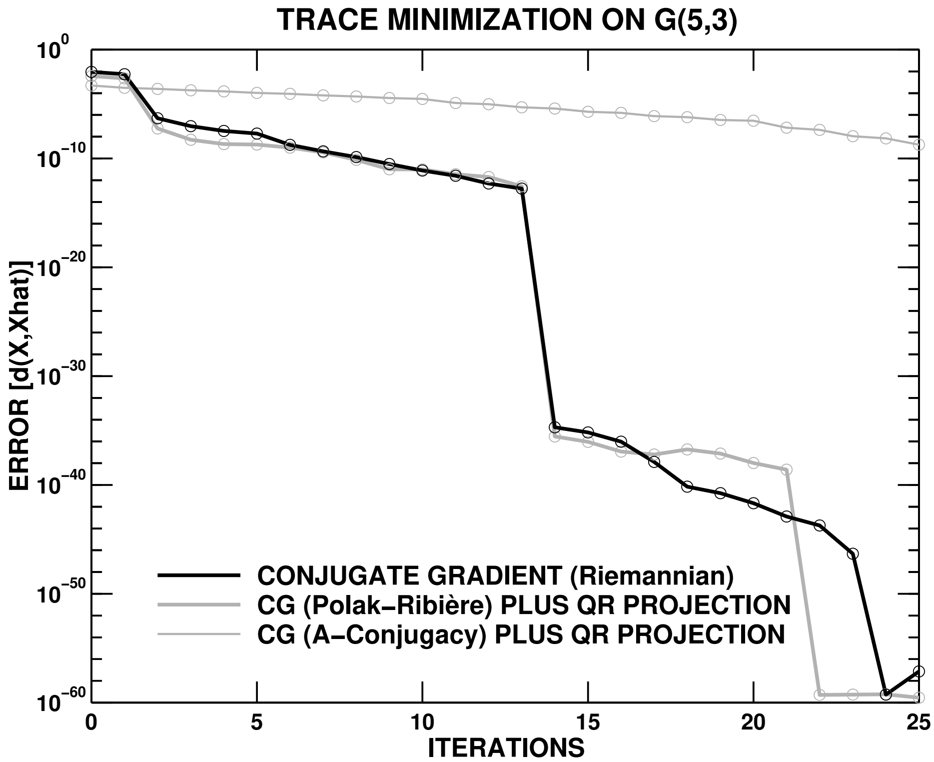

Figure 5.2: Riemannian conjugate gradient, unconstrained block Polak-Ribière, and the A-conjugacy algorithms compared.

The key components of the Grassmann algorithm are the total space/base space point of view, the following of geodesics, and the parallel transportation of tangent vectors. Below we plot a schematic figure that describes conjugate gradient on a Riemannian manifold:

In Figure 2, we show convergence curves in exact arithmetic for three block algorithms. The A-conjugacy algorithm is clearly inferior and not worth considering. The Riemannian algorithm and our own unconstrained version of the block Polak-Ribière algorithm, which approximates the Riemannian CG algorithm up to second order and is less computationally intensive than the full-blown Riemannian algorithm, are quadratically convergent.

6 Beyond Conjugate Gradient

Potential users of a conjugate gradient algorithm may well consider the viewpoint expressed by Jorge Nocedal [26]:

> The recent development of limited memory and discrete Newton methods have narrowed the class of problems for which conjugate gradient methods are recommended. Nevertheless, in my view, conjugate gradient methods are still the best choice for solving very large problems with relatively inexpensive objective functions. They can also be more suitable than limited memory methods on several types of multiprocessor computers.

Though we believe that we have derived the correct NCG algorithm for functions such as the block Rayleigh quotient, it is very possible that the best algorithms for applications in the LDA community and the signal processing community may well be a Newton iteration rather than an algorithm in the conjugate gradient family. Rightly or wrongly, the CG algorithms may yet remain popular for the largest problems because of the simplicity of programming and limited memory requirements.

It is important to understand the relationship between preconditioned conjugate gradient and Newton's method. We sometimes consider the ideal preconditioner to be one that is easily computed, and yet closely resembles the inverse of the Hessian. Multiplying the gradient by the exact inverse of the Hessian is exactly the Newton method. Therefore the Newton method is equivalent to preconditioned conjugate gradient without any use of a previous search direction. One expects that taking advantage of Hessian information, if convenient, would lead to superior convergence.

We now explore Newton's method for invariant subspace computation more closely. Suppose (for simplicity only) that A is symmetric. We would then wish to find matrices Y and B such that

$$AY - YB = 0. \tag{6.1}$$

The above system is degenerate because there are more unknown variables than equations; constraints of some form or another need to be imposed. One approach is the general

$$\text{affine constraint } Z^T Y = I \tag{6.2}$$

introduced by Chatelin [5, 6] in which case $B = Z^T AY$. We observe that the algorithms proposed by Dongarra, Moler, and Wilkinson [13] and Stewart [34]. represent two special choices for the affine constraint matrix Z. In the Dongarra et al. case, Z may be obtained by inverting and transposing an arbitrary $p \times p$ minor of the $n \times p$ matrix Y. In Moler's Matlab notation, `Z=zeros(n,p); Z(r,:)=inv(Y(r,:))'`, where `r` denotes a p-vector of row indices. For Stewart, $Z = Y(Y^T Y)^{-1}$. Demmel [12, Section3] originally observed that the three algorithms were related. Being linear, the affine constraint allows a direct application of Newton's method without any difficulty.

An alternative choice for the constraint is the

$$\text{orthogonality constraint } Y^T Y = I. \tag{6.3}$$

Newton's method with this constraint is slightly more complicated because of the nonlinearity. The value of B is now $Y^T AY$. However the equation $AY - Y(Y^T AY) = 0$, even with the orthogonality constraints, is degenerate. This makes this problem even more difficult to handle.

The geometric approach to the Grassmann manifold gives an approach to resolving this problem. We do not derive the details here, but the Grassmann point of view on the Newton method starts with the second covariant derivative of the block Rayleigh quotient (see [14]):

$$\mathrm{Hess}(\Delta_1, \Delta_2) = \mathrm{tr}\Big(\Delta_1^T A \Delta_2 - (\Delta_1^T \Delta_2) Y^T AY\Big). \tag{6.4}$$

From there to pick the Newton search direction, we must solve for Δ such that $Y^T \Delta = 0$ in the Sylvester equation

$$\Pi(A\Delta - \Delta(Y^T AY)) = -G, \tag{6.5}$$

where $\Pi = (I - YY^T)$ denotes the projection onto the tangent space of the Grassmann manifold, and $G = \Pi AY$ is the gradient. We then may use this search direction to follow a geodesic.

Equation (6.5) is really a column by column Rayleigh quotient iteration in disguise. To see this, write $Y^T AY =: Q\Theta Q^T$, Θ diagonal, and project Equation (6.5) onto Image(Π). The projected equation is

$$\bar{A}\bar{\Delta} - \bar{\Delta}\Theta = -\bar{G}, \tag{6.6}$$

where the barred quantities are $\bar{A} = \Pi A \Pi$, $\bar{\Delta} = \Pi \Delta Q$ is a column matrix of Ritz vectors [27], and $\bar{G} = GQ$. If $\bar{\Delta}$ is any solution to the above equation, then $\Delta = \bar{\Delta} Q^T$ is a solution to Equation (6.5). One way to interpret Equation (6.6) is that we turned the orthogonality constraint $Y^T Y = I$ into an affine constraint $\Delta^T Y = 0$ by differentiating.

Solving for Δ and exponentiating amounts to performing the Rayleigh quotient iteration on each of the Ritz vectors associated with the subspace Y. Therefore, Newton's method applied to the function $\mathrm{tr} Y^T A Y$ on the Grassmann manifold converges cubically. This is a generalization of the identification between RQI and Newton's method applied to Rayleigh's quotient on the sphere [32, 33]. This method also has very much in common with Chatelin's method, yet it is the natural algorithm from the differential geometry point of view given the orthogonality constraints $Y^T Y = I$.

REFERENCES

1. B. M. Alsén, *Multiple step gradient iterative methods for computing eigenvalues of large symmetric matrices*, Tech. Rep. UMINF-15, University of Umeå, 1971.
2. I. Andersson, *Experiments with the conjugate gradient algorithm for the determination of eigenvalues of symmetric matrices*, Tech. Rep. UMINF-4.71, University of Umeå, 1971.
3. T. A. Arias, M. C. Payne, and J. D. Joannopoulous, Ab initio *molecular dynamics: analytically continued energy functionals and insights into iterative solutions*, Physical Review Letters, 71 (1992), pp. 1077–1080.
4. W. W. Bradbury and R. Fletcher, *New iterative methods for solutions of the eigenproblem*, Numerische Mathematik, 9 (1966), pp. 259–267.
5. F. Chatelin, *Simultaneous Newton's iteration for the eigenproblem*, Computing, Suppl., 5 (1984), pp. 67–74.
6. ——, *Eigenvalues of Matrices*, John Wiley & Sons, New York, 1993.
7. H. Chen, T. K. Sarkar, S. A. Dianat, and J. D. Brulé, *Adaptive spectral estimation by the conjugate gradient method*, IEEE Trans. Acoustics, Sppech, and Signal Processing, ASSP-34 (1986), pp. 272–284.
8. J. K. Cullum, *The simultaneous computation of a few of the algebraically largest and smallest eigenvalues of a large, sparse, symmetric matrix*, BIT, 18 (1978), pp. 265–275.
9. J. K. Cullum and W. E. Donath, *A block generalization of the symmetric s-step Lanczos algorithm*, Tech. Rep. RC 4845, IBM Research, Yorktown Heights, New York, 1974.
10. ——, *A block Lanczos algorithm for computing the q algebraically largest eigenvalues and a corresponding eigenspace of large, sparse, real symmetric matrices*, in Proceedings of the 1974 Conference on Decision and Control, Phoeniz, Arizona, November 1974, pp. 505–509.
11. J. K. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations*, vol. 1 Theory, Birkhäuser, Stuttgart, 1985.
12. J. W. Demmel, *Three methods for refining estimates of invariant subspaces*, Computing, 38 (1987), pp. 43–57.
13. J. J. Dongarra, C. B. Moler, and J. H. Wilkinson, *Improving the accuracy of computed eigenvalues and eigenvectors*, SIAM J. Num. Analysis, 20 (1983), pp. 46–58.

14. A. Edelman, T. A. Arias, and S. T. Smith, *Conjugate gradient on the Stiefel and Grassmann manifolds*, submitted to SIAM J. Matrix Anal. Appl., (1995).
15. R. L. Fox and M. P. Kapoor, *A miminimization method for the solution of the eigenproblem arising in structural dynamics*, in Proceedings of the Second Conference on Matrix Methods in Structural Mechanics, L. Berke, R. M.Bader, W. J. Mykytow, J. S. Przemieniecki, and M. H. Shirk, eds., Wright-Patterson Air Force Base, Ohio, 1969, pp. 271–306.
16. I. Fried, *Gradient methods for finite element eigenproblems*, AIAA Journal, 7 (1969), pp. 739–741.
17. ———, *Optimal gradient minimization scheme for finite element eigenproblems*, J. Sound Vib, 20 (1972), pp. 333–342.
18. Z. Fu and E. M. Dowling, *Conjugate gradient eigenstructure tracking for adaptive spectral estimation*, IEEE Trans. Signal Processing, 43 (1995), pp. 1151–1160.
19. D. R. Fuhrmann and B. Liu, *An iterative algorithm for locating the minimal eigenvector of a symmetric matrix*, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, 1984, pp. 45.8.1–4.
20. M. Geradin, *The computational efficiency of a new minimization algorithm for eigenvalue analysis*, J. Sound Vibration, 19 (1971), pp. 319–331.
21. M. J. Gillan, *Calculation of the vacancy formation energy in aluminium*, Journal of Physics, Condensed Matter, 1 (1989), pp. 689–711.
22. G. Golub and D. O'Leary, *Some history of the conjugate gradient and Lanczos methods*, SIAM Review, 31 (1989), pp. 50–102.
23. R. Haimi-Cohen and A. Cohen, *Gradient-type algorithms for partial singular value decomposition*, IEEE Trans. Pattern. Anal. Machine Intell., PAMI-9 (1987), pp. 137–142.
24. W. Karush, *An iterative method for finding characteristic vectors of a symmetric matrix*, Pacific J. Math., 1 (1951), pp. 233–248.
25. G. Kresse and J. Hafner, *Ab initio molecular dynamics for liquid metals*, Physical Review B, (1993), pp. 558–561.
26. J. Nocedal, *Theory of algorithms for unconstrained optimization*, Acta Numerica, (1992), pp. 199–242.
27. B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1980.
28. M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients*, Rev. Mod. Phys, 64 (1992), pp. 1045–1097.
29. A. Perdon and G. Gambolati, *Extreme eigenvalues of large sparse matrices by Rayleigh quotient and modified conjugate gradients*, Comp. Methods Appl. Mech. Engin., 56 (1986), pp. 251–264.

30. A. Ruhe, *Iterative eigenvalue algorithms for large symmetric matrices*, in Numerische Behandlung von Eigenwertaaufgaben Oberwolfach 1972, Intl, Series Numerical Math. Volume 24, 1974, pp. 97–115.

31. A. H. Sameh and J. A. Wisniewski, *A trace minimization algorithm for the generalized eigenvalue problem*, SIAM Journal of Numerical Analysis, 19 (1982), pp. 1243–1259.

32. S. T. Smith, *Geometric Optimization Methods for Adaptive Filtering*, PhD thesis, Harvard University, Cambridge, Massachusetts, 1993.

33. ———, *Optimization techniques on Riemannian manifolds*, in Hamiltonian and Gradient Flows, Algorithms and Control, A. Bloch, ed., vol. 3 of Fields Institute Communications, Providence, RI, 1994, American Mathematical Society, pp. 113–146.

34. G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Review, 15 (1973), pp. 752–764.

35. I. Štich, R. Car, M. Parrinello, and S. Baroni, *Conjugate gradient minimization of the energy functional: A new method for electronic structure calculation*, Phys. Rev. B., 39 (1989), pp. 4997–5004.

36. M. W. Sung, R. Kawai, and J. Weare, *Packing transitions in nanosized* Li *clusters*, Physical Review Letter, 73 (1994), pp. 3552–3555.

37. M. P. Teter, M. C. Payne, and D. C. Allan, *Solution of Schrödinger's equation for large systems*, Phys. Review B, 40 (1989), pp. 12255–12263.

38. M. A. Townsend and G. E. Johnson, *In favor of conjugate directions: A generalized acceptable-point algorithm for function minimization*, J. Franklin Inst., 306 (1978).

39. X. Yang, T. P. Sarkar, and E. Arvas, *A survey of conjugate gradient algorithms for solution of extreme eigen-problems of a symmetric matrix*, IEEE Trans. Acoust., Speech, Signal Processing, 37 (1989), pp. 1550–1556.