

A NEW ALGORITHM FOR THE GENERALIZED EIGENVALUE PROBLEM

Knut Hüper and Uwe Helmke

Department of Mathematics, University of Würzburg

D-97074 Würzburg, Germany

hueper@mathematik.uni-wuerzburg.de

ABSTRACT

The problem of finding the generalized eigenvalues and eigenvectors of a pair of real symmetric matrices A and B , with $B > 0$, can be viewed as a smooth optimization problem on a smooth manifold. We present a cost function approach to the generalized eigenvalue problem which is posed on the product of the n -sphere and Euclidian space $\mathbb{R}$. The critical point set of this cost function is studied. An algorithm is presented based on constrained optimization. A proof of local quadratic convergence is given.

1. INTRODUCTION

In this paper a cost function approach for the generalized eigenvalue problem is presented. It is related to methods minimizing the Rayleigh quotient, cf. [1] and [2], but is different in a number of important points.

The generalized Rayleigh quotient when restricted to the n -sphere S^{n-1} has a minimum, a maximum, as well as saddle points. Our cost function is defined on a noncompact manifold, the product $S^{n-1} \times \mathbb{R}$. Moreover, it is bounded from below by 0 and we will show that there exist only global minima but no further critical points.

The optimization of our cost function in terms of Jacobi-type rotations coupled with a

shift-type strategy, known from inverse iteration and QR-type methods, results in a very efficient algorithm. Because we use orthogonal Jacobi-type transformations a VLSI-realization is probably easy to implement due to its inherent regularity and parallelism. Local quadratic convergence is shown using elementary tools of analysis.

2. COST FUNCTION APPROACH TO THE GENERALIZED EIGENVALUE PROBLEM

We aim at finding real generalized eigenvalues for a definitizable pair of real symmetric matrices. That is, for

$$\det(A - \lambda B) \neq 0,$$

solve

$$\det(A - \lambda B) = 0$$

with

$$\lambda \in \mathbb{R},$$

$$A = A' \in \mathbb{R}^{n \times n},$$

and

$$B = B' \in \mathbb{R}^{n \times n}, \quad B > 0.$$

Equivalently, we want to find all real solutions $\lambda \in \mathbb{R}, x \in \mathbb{R}^n$ with $\|x\| = 1$, of

$$(A - \lambda B)x = 0.$$

Then $x \in S^{n-1}$ is called a real generalized eigenvector of (A, B) and $\lambda \in \mathbb{R}$ a real generalized eigenvalue; characterized as

$$\det(A - \lambda B) = 0.$$

Real generalized eigenvalues/eigenvectors are thus the real eigenvalues/eigenvectors of $B^{-1}A$.

The next result characterizes the structure of the set of critical points of our cost function.

Theorem 2.1 *Consider the smooth function*

$$f : S^{n-1} \times \mathbb{R} \rightarrow \mathbb{R},$$

defined by

$$\begin{aligned} f(x, \lambda) &= \|(A - \lambda B)x\|^2 \\ &= x'(A - \lambda B)^2 x. \end{aligned}$$

(a) $(x, \lambda) \in S^{n-1} \times \mathbb{R}$ is a critical point of $f : S^{n-1} \times \mathbb{R} \rightarrow \mathbb{R}$ if and only if

$$B^{-1}Ax = \lambda x$$

holds.

(b) *All critical points are global minima with critical value equal to 0.*

Proof. The derivative of $f : S^{n-1} \times \mathbb{R} \rightarrow \mathbb{R}$ at a point $(x, \lambda) \in S^{n-1} \times \mathbb{R}$ vanishes if and only if

$$(A - \lambda B)^2 x = \mu^2 x, \quad \mu \in \mathbb{R},$$

as well as

$$x'(A - \lambda B)Bx = 0$$

hold. Because B is nonsingular and $A - \lambda B$ is symmetric it follows that $\mu = 0$ and thus

$$x \in \ker(A - \lambda B)$$

must hold. Equivalently,

$$Ax = \lambda Bx$$

holds and by the invertibility of B we get

$$B^{-1}Ax = \lambda x.$$

It is easily seen that the critical values are all equal to 0 which is the global minimum value of the function $f : S^{n-1} \times \mathbb{R} \rightarrow \mathbb{R}$. $\square$

Recall the generalized Rayleigh quotient

$$r_{A,B} : S^{n-1} \rightarrow \mathbb{R},$$

defined by

$$r_{A,B}(x) = \frac{x'Ax}{x'Bx}.$$

Here x is a critical point of $r_{A,B}$ with critical value λ if and only if

$$B^{-1}Ax = \lambda x$$

holds. But not all critical points of $r_{A,B}$ are minima. There are a maximum and saddle points as well. For a discussion of the ordinary Rayleigh quotient, cf. [3].

3. CONSTRAINED OPTIMIZATION PROCEDURE

Consider the following optimization procedure of the function $f : S^{n-1} \times \mathbb{R} \rightarrow \mathbb{R}$ where x_{k+1} and λ_{k+1} are recursively defined by

$$\begin{aligned} x_{k+1} &= \arg \min_{x \in S^{n-1}} x'(A - \lambda_k B)^2 x, \\ \lambda_{k+1} &= \frac{x'_{k+1} A x_{k+1}}{x'_{k+1} B x_{k+1}}. \end{aligned}$$

Thus we minimize $f(x, \lambda) = x'(A - \lambda B)^2 x$ on S^{n-1} leaving λ fixed and then minimize $(x'(A - \lambda B)x)^2$ with respect to λ leaving invariant x . By this choice of λ_k it follows that x_k becomes isotropic with respect to the quadratic form corresponding to the matrix $A - \lambda_k B$. The vector $x_{k+1} \in S^{n-1}$ then is chosen as an eigenvector of $(A - \lambda_k B)^2$ corresponding to the smallest eigenvalue. Equivalently, x_{k+1} is a right singular vector of $A - \lambda_k B$ corresponding to the smallest singular value.

4. CONVERGENCE PROPERTIES

In this section we discuss the local convergence properties of our algorithm. At the moment we have no proof for global convergence. Standard approaches, e.g., a Lyapunov-type argument combined with continuity properties of the functions

$$f : S^{n-1} \times \mathbb{R} \rightarrow \mathbb{R},$$

$$f(x, \lambda) = x'(A - \lambda B)^2 x$$

and

$$h : S^{n-1} \rightarrow \mathbb{R},$$

$$h(x) = \min_{\lambda \in \mathbb{R}} \{x'(A - \lambda B)^2 x\}$$

will fail to work. On the other hand there is numerical evidence that the algorithm is indeed globally convergent.

To prove the next result we use elementary tools from global analysis.

Theorem 4.1 *Let two symmetric matrices $A, B \in \mathbb{R}^{n \times n}$ be given. Let B be positive definite. Given an arbitrary but fixed $\lambda \in \mathbb{R}$, let*

$$s_\lambda : S^{n-1} \rightarrow S^{n-1} \quad (1)$$

with

$$x_{k+1} = s_\lambda(x_k), \quad k \in \mathbb{N}_0,$$

be a quadratical fast algorithm or, alternatively, one step of a quadratical fast algorithm to compute

$$\min_{x \in S^{n-1}} x'(A - \lambda B)^2 x.$$

Consider the sequence $((x_k, \lambda_k))$, with $k = 0, 1, 2, \dots$, generated by the recursions

$$x_{k+1} = s_\lambda(x_k),$$

$$\lambda_{k+1} = \frac{x'_{k+1} A x_{k+1}}{x'_{k+1} B x_{k+1}},$$

with $x_0 \in S^{n-1}$ arbitrary and $\lambda_0 := \frac{x'_0 A x_0}{x'_0 B x_0}$.

If the sequence $((x_k, \lambda_k))$ converges to (x, λ) , where $B^{-1}Ax = \lambda x$ holds, then it converges locally quadratically.

Proof. The derivative of s_λ at $x \in S^{n-1}$ is the linear mapping that assigns to a tangent vector $\psi \in T_x S^{n-1}$ the value $Ds_\lambda(x) \cdot \psi$. Because $s_\lambda(x)$ minimizes $x'(A - \lambda B)^2 x$

$$Ds_\lambda(x_{\min}) \cdot \xi = 0 \quad (2)$$

must hold, with $\xi \in T_{x_{\min}} S^{n-1}$. Now consider the mapping

$$\tau : S^{n-1} \rightarrow S^{n-1},$$

defined by

$$\tau(x) = s_{\lambda(x)}(x).$$

Note that $\tau : S^{n-1} \rightarrow S^{n-1}$ is smooth in an open neighborhood of $x_{\min}$ if (1) is the so-called Sort-Jacobi algorithm, cf. [4]. This can be shown by applying the implicit function theorem to a suitable function, cf. [4], [5], and [6] for a general discussion of Jacobi-type methods and related arguments.

Then

$$\begin{aligned} D\tau(x_{\min}) \cdot \xi &= \frac{\partial s}{\partial \lambda} D\lambda(x_{\min}) \cdot \xi + Ds(x_{\min}) \cdot \xi \\ &= \frac{\partial s}{\partial \lambda} D\lambda(x_{\min}) \cdot \xi \end{aligned}$$

holds because of (2). Furthermore, $D\lambda(x_{\min}) \cdot \xi = 0$ holds if λ is chosen as

$$\lambda_{k+1} = \frac{x'_{k+1} A x_{k+1}}{x'_{k+1} B x_{k+1}}.$$

Thus

$$D\tau(x_{\min}) \cdot \xi = 0$$

vanishes. Now quadratic convergence follows by the smoothness of the mapping $s_\lambda : S^{n-1} \rightarrow S^{n-1}$ around the point $x_{\min}$ together with a standard Taylor expansion argument

$$\|s_{\lambda(x_k)}(x_k) - x_{\min}\| \leq \sup \|D^2 s_{\lambda(x)}(x)\| \|x_k - x_{\min}\|^2.$$

□

5. DISCUSSION

As stated above one step

$$x_{k+1} = s_\lambda(x_k)$$

of a quadratical fast algorithm is sufficient to ensure the local quadratic convergence of the whole scheme. For that purpose one sweep of the so-called Sort-Jacobi algorithm for singular value computations, introduced in [4], seems to be well suited. On the one hand because the Sort-Jacobi algorithm converges globally and locally quadratically even in the case of multiple singular values, on the other hand because this method is able to sort the singular values in any prescribed order on the diagonal. Both features are neither shared by standard Jacobi-type methods nor by Kogbetliantz algorithms, cf. [1] and related literature. The basic difference to standard textbook Jacobi algorithms is that the Sort-Jacobi algorithm minimizes Brockett's linear trace function, cf. [7], in each step, rather than minimizing the quadratic off-norm function.

It seems to be new to interpret shifts known from QR-type methods and inverse iterations as additional optimization parameters for function optimization on differentiable manifolds. On the other hand as a shortcoming of this strategy the function $f : S^{n-1} \times \mathbb{R} \rightarrow \mathbb{R}$ is no longer strictly decreasing on the sequence $((x_k, \lambda_k))$.

Numerical experiments and a more detailed discussion on the convergence properties of our algorithm will be published elsewhere.

6. REFERENCES

- [1] G. Golub and C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 2nd edition, 1989.
- [2] B.N. Parlett. *The symmetric eigenvalue problem*. Prentice Hall, Englewood Cliffs, 1980.
- [3] S. Batterson and J. Smillie. The dynamics of Rayleigh quotient iteration. *SIAM J. Num. Anal.*, 26(3):624–636, 1989.
- [4] K. Hüper. *Structure and convergence of Jacobi-type methods for matrix computations*. PhD thesis, Technical University of Munich, June 1996.
- [5] K. Hüper and U. Helmke. Structure and convergence of Jacobi-type methods. *Num. Math.*, 1996. to appear.
- [6] K. Hüper, U. Helmke, and J.B. Moore. Structure and convergence of conventional Jacobi-type methods minimizing the off-norm function. In *Proc. IEEE of the 35th Conference on Decision and Control*, Kobe, Japan, 1996.
- [7] R.W. Brockett. Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems. In *Proc. IEEE of the 27th Conference on Decision and Control*, pages 799–803, Austin, TX, 12 1988. See also *Lin. Algebra & Applic.*, 146:79-91, 1991.