On The Need For Special Purpose Algorithms For Minimax Eigenvalue Problems

by

E. Panier

ON THE NEED FOR SPECIAL PURPOSE ALGORITHMS
FOR MINIMAX EIGENVALUE PROBLEMS [1] [2]

Eliane R. Panier[3]

Abstract. It has been recently reported that minimax eigenvalue problems can be formulated as nonlinear optimization problems involving smooth objective and constraint functions. This result seems very appealing since minimax eigenvalue problems are known to be typically nondifferentiable. In this paper, we show however that general purpose nonlinear optimization algorithms usually fail to find a solution to these smooth problems even in the simple case of minimization of the maximum eigenvalue of an affine family of symmetric matrices, a convex problem for which efficient algorithms are available.
Key Words. Minimax problems, eigenvalue problems, constrained optimization.

1. Introduction

In many optimization problems arising from engineering considerations, such as robustness or sensitivity analysis (see, e.g., [1–4]), one has to deal with objective or constraint functions involving the spectral radius, the maximum singular value, or the maximum real part of the eigenvalues of a matrix depending on a parameter vector. These functions are generally not differentiable but merely directionally differentiable in the presence of eigenvalues of multiplicity larger than 1 [5]. Special purpose algorithms have been proposed [6–10] for solving such problems. Recently Goh and Teo [11] showed that minimax eigenvalue problems (minimization of the maximum real part of the eigenvalues of a matrix) are equivalent to constrained optimization problems involving only differentiable functions. In this paper, we study some of the features of the problem obtained in [11] for the simplified case of minimization of the maximum eigenvalue of a symmetric matrix. We show that, although this problem involves only smooth functions, it generally cannot be solved by general purpose nonlinear optimization algorithms.

1

In Section 2, we recall briefly the nonlinear formulation in [11] for the case of minimization of the largest eigenvalue of a symmetric matrix depending on a parameter vector. The constrained optimization problem thus obtained is studied in Section 3. Numerical results support the analysis. Finally, concluding remarks are presented in Section 4.

2. Background

Throughout this paper, we will consider a family of symmetric matrices $Q(x) \in \mathbb{R}^{n \times n}$ depending on a parameter vector $x \in \mathbb{R}^m$. The eigenvalues of any matrix $Q(x)$ in the family will be denoted by $\lambda_1(x) \geq \lambda_2(x) \geq \ldots \geq \lambda_n(x)$. We will study the nondifferentiable optimization problem

$$(P) \quad \min_{x \in \mathbb{R}^m} \lambda_1(x).$$

In this particular context, Goh and Teo's [11] basic idea consists in rewriting Problem (P) as

$$\min_{(x,\gamma,\rho) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n} \gamma$$

$$\text{s.t.} \quad \rho_i \leq \gamma, \quad i = 1, \ldots, n$$

$$\rho_i = \lambda_i(x), \quad i = 1, \ldots, n.$$

A necessary and sufficient condition for the quantities ρ_i to be equal to the eigenvalues $\lambda_i(x)$ is that the two polynomials in λ, $\det(\lambda I - Q(x))$ and $\prod_{i=1}^n (\lambda - \rho_i)$ be equal. Since these two polynomials are monic and of degree n, they will be equal if, and only if, their values at any n distinct points $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ coincide. Based on this observation, Goh and Teo conclude that the solution to (P) can be obtained by solving instead

$$(\tilde{P}) \begin{cases} \min\limits_{(x,\gamma,\rho) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n} \gamma \\ \\ \qquad \text{s.t.} \quad \rho_i \leq \gamma, \quad i = 1, \ldots, n \\ \\ \qquad f_k(x, \rho) = 0, \quad k = 1, \ldots, n \end{cases}$$

where

$$f_k(x, \rho) \equiv \det(\tilde{\lambda}_k I - Q(x)) - \prod_{i=1}^{n} (\tilde{\lambda}_k - \rho_i), \quad k = 1, \ldots, n$$

2

and $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ are some arbitrary distinct values. Although this is not explicit in [11], it can be easily seen that Problems (P) and $(\tilde{P})$ are equivalent in the following slightly stronger sense.

Fact.

x^* is a local solution for Problem (P) if, and only if, (x^*, γ^*, ρ^*) is a local solution for Problem $(\tilde{P})$ for some γ^*, ρ^*.

$\square$

While the objective function in Problem (P) is usually nondifferentiable in the presence of a multiple maximum eigenvalue, Problem $(\tilde{P})$ involves only smooth functions.

3. Analysis

Using standard techniques for solving Problem $(\tilde{P})$ instead of tackling directly Problem (P) seems very appealing. However, we show in this section that the most common nonlinear optimization algorithms will usually fail to find a local (a *fortiori* a global) solution for Problem $(\tilde{P})$.

We first show that at any feasible point for Problem $(\tilde{P})$ at which two eigenvalues are equal, the gradients (vectors of partial derivatives with respect to x and ρ) associated with the equality constraints are linearly dependent.

Observation 3.1.

Let $x^* \in \mathbb{R}$ be such that $\lambda_r(x^*) = \lambda_{r+1}(x^*)$ for some $r \in \{1, \ldots, n-1\}$ and let $\rho^* = (\lambda_1(x^*), \ldots, \lambda_n(x^*))^T$. Then, the vectors $\nabla f_k(x^*, \rho^*)$, $k = 1, \ldots, n$, defined by,

$$\nabla f_k(x^*, \rho^*) = (\frac{\partial f_k}{\partial x}(x^*, \rho^*), \frac{\partial f_k}{\partial \rho}(x^*, \rho^*))^T$$

are linearly dependent.

Proof.

3

We proceed to show that there exist some scalars β_k, $\quad k = 1, \ldots, n$, not all zero, such that, for any $\begin{pmatrix} d \\ \alpha \end{pmatrix}$ in $\mathbb{R}^{m+n}$,

$$\langle \sum_{k=1}^{n} \beta_k \nabla f_k(x^*, \rho^*), \begin{pmatrix} d \\ \alpha \end{pmatrix} \rangle = 0$$

where $\langle \cdot, \cdot \rangle$ represents the Euclidian scalar product in $\mathbb{R}^{m+n}$. It will therefore hold,

$$\sum_{k=1}^{n} \beta_k \nabla f_k(x^*, \rho^*) = 0.$$

The functions $\lambda_i(x)$, $\quad i = 1, \ldots, n$, admit a directional derivative at x^* in the direction d (see [5]) that we will denote below by $(\lambda_i)'(x^*; d)$, $\quad i = 1, \ldots, n$. Now, by definition,

$$f_k(x, \rho) \equiv \prod_{i=1}^{n}(\tilde{\lambda}_k - \lambda_i(x)) - \prod_{i=1}^{n}(\tilde{\lambda}_k - \rho_i), \quad k = 1, \ldots, n.$$

Differentiating and evaluating the derivatives at (x^*, ρ^*) in the direction $\begin{pmatrix} d \\ \alpha \end{pmatrix}$ gives

$$\langle \nabla f_k(x^*, \rho^*), \begin{pmatrix} d \\ \alpha \end{pmatrix} \rangle = -\sum_{j=1}^{n}(\lambda_j)'(x^*; d) \prod_{i \neq j}(\tilde{\lambda}_k - \lambda_i(x^*)) + \sum_{j=1}^{n} \alpha_j \prod_{i \neq j}(\tilde{\lambda}_k - \lambda_i(x^*)), \quad k = 1, \ldots, n.$$

Letting $\xi_{j,k} \equiv -\prod_{i \neq j}(\tilde{\lambda}_k - \lambda_i(x^*))$ $\quad j, \; k = 1, \ldots, n$, we obtain

$$\langle \nabla f_k(x^*, \rho^*), \begin{pmatrix} d \\ \alpha \end{pmatrix} \rangle = \sum_{j=1}^{n}((\lambda_j)'(x^*; d) - \alpha_j)\xi_{j,k}, \quad k = 1, \ldots, n. \tag{1}$$

Now, since $\lambda_r(x^*) = \lambda_{r+1}(x^*)$, the vectors $(\xi_{1,k}, \ldots, \xi_{n,k})^T$, $\quad k = 1, \ldots n$, have identical rth and $r + 1$th components so that they are linearly dependent. Therefore, there exist some values β_k, $\quad k = 1, \ldots, n$, not all zero, such that, $\sum_{k=1}^{n} \beta_k \xi_{j,k} = 0$, $\quad \forall \; j = 1, \ldots, n$. But this yields, in view of (1),

4

$$\langle \sum_{k=1}^{n} \beta_k \nabla f_k(x^*, \rho^*), \begin{pmatrix} d \\ \alpha \end{pmatrix} \rangle = 0.$$

□ .

Although some methods may be less sensitive then others (see, e.g., REQP (Recursive Equality Quadratic Programming) [12] *vs.* SQP (Sequential Quadratic Programming) [13]), in view of Observation 3.1, numerical difficulties may arise when the iterate is close to a feasible point for Problem $(\tilde{P})$ corresponding to multiple eigenvalues. Also, standard nonlinear optimization algorithms may converge to such a point, at which the gradients of the equality constraints are dependent, instead of a stationary point for Problem $(\tilde{P})$.

An even more serious problem can be seen from the next observation.

Observation 3.2.

Let $x^* \in \mathbb{R}^m$ and $\rho^* \in \mathbb{R}^n$ with $\rho_l^* = \rho_r^*$ for some l, $r \in \{1, \dots, n\}$. Then,

$$\frac{\partial f_k}{\partial \rho_l}(x^*, \rho^*) = \frac{\partial f_k}{\partial \rho_r}(x^*, \rho^*), \quad k = 1, \dots, n.$$

□

In view of the above, it can be easily checked that if $q + 1$ values ρ_i^* are equal, with $q > m$, then the vectors $\nabla f_k(x^*, \rho^*)$, $k = 1, \dots, n$, are linearly dependent even when the values ρ_i^* do not correspond to eigenvalues at x^*. In this case, an SQP algorithm may fail to find a search direction. Such a situation is likely to occur when we have matrices of big dimension depending on a few parameters only.

A more important consequence of Observation 3.2 is that Problem $(\tilde{P})$ is symmetrical in all the variables ρ_i so that, if two values ρ_l and ρ_r are identical at some time within a usual optimization process (see, e.g., SQP or penalty methods), they will remain identical thereafter. In that case, the algorithm will never be able to detect a solution at which the eigenvalues are simple. We give below an illustration of that phenomenon. In that

5

example, the matrices $Q(x)$ are affine in x. It is well known (see, e.g., [8] Theorem $A.1$) that, in that case the objective function of Problem (P) is convex (although not necessarily differentiable). All local solutions for (P) are therefore global solutions and efficient algorithms for such problems do exist (see, e.g., [7,9]).

Let

$$Q(x) = \begin{bmatrix} 1 & x & 0 \\ x & -1 & 0 \\ 0 & 0 & 2x \end{bmatrix}.$$

The characteristic polynomial associated with any element $Q(x)$ in the family is $\det(\lambda) \equiv (\lambda^2 - (1+x^2))(\lambda - 2x)$ so that the eigenvalues are $\sqrt{1+x^2}$, $2x$ and $-\sqrt{1+x^2}$. The largest eigenvalue is minimized at $x = 0$. Notice that, at that point, the eigenvalues are distinct so that, locally, the objective function in Problem (P) is differentiable. At $x = 1/\sqrt{3}$ (resp. $x = -1/\sqrt{3}$), the eigenvalues are $2/\sqrt{3}$, $2/\sqrt{3}$, $-2/\sqrt{3}$ (resp. $2/\sqrt{3}$, $-2/\sqrt{3}$, $-2/\sqrt{3}$). Therefore, in view of Observation 3.1, at the points $(x,\rho)_1 = (1/\sqrt{3}, 2/\sqrt{3}, 2/\sqrt{3}, -2/\sqrt{3})$ and $(x,\rho)_2 = (-1/\sqrt{3}, 2/\sqrt{3}, -2/\sqrt{3}, -2/\sqrt{3})$, the gradients of the active constraints are dependent. These points are also local solutions for the problems

$$\min_{(x,\gamma,\rho)\in\mathbb{R}^m\times\mathbb{R}\times\mathbb{R}^n} \gamma$$

$$\text{s.t.} \quad \rho_i \leq \gamma, \quad i = 1,\ldots,n$$

$$f_k(x,\rho) = 0, \quad k = 1,\ldots,n$$

$$\rho_1 = \rho_2 \quad (\text{resp. } \rho_2 = \rho_3).$$

(but not for Problem $(\tilde{P})$) since they are isolated feasible points. Problem $(\tilde{P})$ (as well as all the examples presented in this Section) has been tested using the routine $VF02AD$ of the Harwell library [14]. We report below the final iterates obtained starting from different initial points. These results show that the iterates can be trapped at the isolated points $(x,\rho)_1$ and $(x,\rho)_2$. For these tests, we chose the values $\tilde{\lambda}_1 = 0$, $\tilde{\lambda}_2 = 1$, $\tilde{\lambda}_3 = 3$. The final accuracy in $VF02AD$ was taken as $1\cdot 10^{-6}$ (i.e., the calculation ended when the objective function plus suitably weighted multiples of the constraint functions were predicted to

6

differ from their optimal values by at most $1 \cdot 10^{-6}$).

No.	$(x,\rho)_{init}$	$(x,\rho)_{final}$
1	$(2,1,2,3)$	$(-3.362962 \cdot 10^{-5}, -0.9999997, -6.757097 \cdot 10^{-5}, 0.9999994)$
2	$(-2,-2.5,-2,1)$	$(-1.501202 \cdot 10^{-6}, -1.000000, -2.997826 \cdot 10^{-6}, 1.000000)$
3	$(-4,1,-2,-3)$	$(0.5773497, 1.154700, -1.154700, 1.154700)$
4	$(-0.5,-2.5,-2,-4)$	$(0.5773498, 1.154700, -1.154700, 1.154700)$
5	$(0.55,-0.8,-0.8,1)$	$(-0.2603311 \cdot 10^{-3}, -1.000000, 1.000000, -0.131790 \cdot 10^{-5})$
6	$(-0.5,-1.6,-1.7,1.9)$	$(-0.5400000, -1.13628, -1.08026, 1.13646)$.

In tests number 1 and 2, convergence to a solution occurred. In tests number 3 and 4, the iterate first hit the subspace $\rho_1 = \rho_3$. From then on, it remained in that subspace progressing toward the point $(x,\rho)_1$. Surprisingly, in test number 5, while we start with an iterate in the subspace $\rho_1 = \rho_2$, there is convergence to the solution of Problem (P), corresponding to simple eigenvalues. Although such a result was not expected in view of Observation 3.2, the components ρ_1 and ρ_2 became distinct due to roundoff error. The perfect symmetry between ρ_1 and ρ_2 was probably destroyed due to the order in which operations are performed in $VF02AD$. Also, it took 1934 iterations of that program before obtaining convergence. Finally, in test number 6 the algorithm exited due to inconsistency in the search direction computation due to near dependence of the gradients associated with the equality constraints. This could not have been predicted by either Observation 3.1 or 3.2.

Another example is

$$Q(x) = \begin{bmatrix} x^2 + 1 & 0 \\ 0 & x^2 + x \end{bmatrix}.$$

It is still a convex problem so that every local solution for Problem (P) is global. Again, convergence may occur to nonstationary points. Indeed, using the values $\tilde{\lambda}_1 = 0$, $\tilde{\lambda}_2 = 1$ and starting with the iterate $(x,\rho)_{init} = (0.5, 1.9, 2.05)$ we got the final point $(x,\rho)_{final} =$

7

$(1.000000, 2.000000, 2.000000)$. As may be expected, due to the dependency of the gradients associated with the equality constraints, convergence was very slow.

The next example shows that, even if the eigenvalues are distinct for all x, the successive iterates may be trapped again by the subspace $\tilde{\rho}_1 = \rho_2$. Let

$$Q(x) = \begin{bmatrix} x^2 + 1 & 0 \\ 0 & x^2 \end{bmatrix}.$$

With the parameter values $\tilde{\lambda}_1 = 0$, $\tilde{\lambda}_2 = 1$ and the initial iterate $(x, \rho)_{init} = (0.2, 1.05, 1.03)$ the optimization process stopped at the iterate $(x, \rho)_{final} = (0.1997500, 0.5399000, 0.5399000)$ at which the gradients of the equalities are dependent.

Finally, we tested the first example given in [11] with values $\tilde{\lambda}_1 = 0$, $\tilde{\lambda}_2 = 1$ and $\tilde{\lambda}_3 = 2$. For that example,

$$Q(x) = \begin{bmatrix} 2 + x_3^2 + (x_1 - x_2)^2 & (x_1 - x_2)(1 - x_3) & x_2 x_3 \\ (x_1 - x_2)(1 - x_3) & x_1^2 + x_2^2 + 3x_3^2 & x_1 x_3 \\ x_2 x_3 & x_1 x_3 & 2 + (x_1 x_3)^2 \end{bmatrix},$$

the minimum of the maximum eigenvalues is equal to 2, and any parameter vector x such that $x_1 = x_2 \leq 1$ and $x_3 = 0$ is optimal. Starting with initial values $(x, \rho)_{init} = (0.5, 0.5, 0.5, 1.5, 1.5, 1.5)$, we get the final iterate $(x, \rho)_{final} = (0.9998771, 0.9995843, 0.2011660 \cdot 10^{-5}, 1.998848, 2.000075, 2.000000)$, consistent with [11]. The algorithm is again successful if we start with $(x, \rho)_{init} = (0.5, 0.5, 0.5, 0, 2, 3)$, in which case convergence occurs to $(x, \rho)_{final} = (0.3588680, 0.3588784, 0.3604587 \cdot 10^{-3}, 0.2575800, 2.000000 \ 2.000000)$. But if we now take $(x, \rho)_{init} = (1, 2, 3, 0, 0.5, 1.5)$ the optimization stops at the infeasible point for Problem $(\tilde{P})$, $(x, \rho)_{final} = (0.3455119, 0.6681274, -0.7033917, 6.193530, 8.168414, 0.1767469 \cdot 10^{+2})$ at which the gradients of the equality constraints are linearly dependent. Different tests have been performed on that same example with different starting points. In many instances, convergence to a solution failed to occur.

4. Conclusion

While it is true that minimax eigenvalue problems can be formulated as nonlinear optimization problems involving only smooth functions, nonlinear optimization algorithms cannot be applied blindly for solving them. Indeed, they have a special structure that is not taken into account by general purpose algorithms.

References

[1] A.J. Laub, "An Inequality and Some Computations Related to the Robust Stability of Linear Dynamic Systems," *IEEE Transactions on Automatic Control* AC-24 (1979), 318–320.

[2] J.C. Doyle, "Robustness of Multiloop Linear Feedback Systems," in *Proc. 17th IEEE Conf. on Decision and Control*, 1979.

[3] N. R. Sandell Jr, "Robust Stability of Systems with Applications to Singular Perturbations," *Automatica* (1979).

[4] J.S. Freudenberg, D.P. Looze & J.B. Cruz, "Robustness Analysis Using Singular Value Sensitivities," *Internat. J. Control* 35 (1982), 95–116.

[5] T. Kato, in *Perturbation Theory for Linear Operators, 2nd edition*, Springer, ed., Berlin, 1976.

[6] J. Cullum, W.E. Donath & P. Wolfe, "The Minimization of Certain Nondifferentiable Sums of Eigenvalues of Symmetric Matrices ," in *Nondifferentiable optimization*, M. Balinski & P. Wolfe, eds., Mathematical programming study #3, North-Holland, 1975.

[7] E. Polak & Y. Wardi, "Nondifferentiable Optimization Algorithm for the Design of Control Systems Subject to Singular Value Inequalities," *Automatica*, 18 (1982), 267–283.

[8] R. Fletcher, "Semi-Definite Matrix Constraints in Optimization," *Siam J. Control and Optimization* 23 (1985), 493–513.

[9] M.L. Overton, "On Minimizing the Maximum Eigenvalue of a Symmetric Matrix," Centre for Mathematical Analysis, Australian National University , Report CMA-R03-87, Canberra, Australia, January, 1987, to appear in SIAM J. Matrix Anal. Appl..

[10] M.L. Overton & R.S. Womersley, "On Minimizing the Spectral Radius of a Nonsymmetric Matrix Function - Optimality Conditions and Duality Theory," 1987, to appear in SIAM J. Matrix Anal. Appl..

[11] C.J. Goh & K.L. Teo, "On Minimax Eigenvalue Problems via Constrained Optimization," *J. Optim. Theory Appl.* 57 (1988), 59–68.

[12] M.C. Bartholomew-Biggs, "A Globally Convergent Version of REQP for Constrained Minimization," *IMA Journal of Numerical Analysis* (1988).

[13] M.J.D. Powell, "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," in *Numerical Analysis, Dundee, 1977, Lecture Notes in Mathematics 630*, G.A. Watson, ed., Springer-Verlag, 1977, 144–157.

[14] "Harwell Subroutine Library," Library Reference Manual, Harwell, England, 1982.