Nonsmooth analysis of eigenvalues: a summary

A.S. Lewis [1]
Department of Combinatorics and Optimization
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
aslewis@orion.uwaterloo.ca
http://orion.uwaterloo.ca/~aslewis

August 21, 1996

Abstract

I outline a unified approach to characterizing Fréchet, limiting Fréchet, and Clarke subgradients of an arbitrary function of the eigenvalues of a real symmetric matrix. In particular, I compute various subdifferentials of the k'th largest eigenvalue. This paper summarizes the results and techniques presented in detail in [4].

[1]The author wishes to thank the University of Milan, and in particular Dr Roberto Lucchetti, for their kind hospitality during the writing of this article. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada

1 Introduction

Variational problems involving eigenvalues arise in at least two major areas of applied mathematics. First, relaxing the integrality constraints in hard combinatorial optimization problems often leads to 'semidefinite programs' (analogues of linear programs with positive semidefinite matrix variables). Secondly, common engineering problems seek to vary the eigenvalues of some design matrix to enhance properties such as stability. The survey [5] gives many examples.

An example of the second type is Rayleigh's problem of finding which shape of vibrating drum has lowest natural frequency. Mathematically, if the shape has a Laplacian with eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots$, then we seek to minimize λ_1: the answer is the circular disk. In 1956, Payne, Pólya and Weinberger conjectured that the circle is also optimal for the problem of minimizing λ_1/λ_2, a conjecture proved recently in [1].

Varitational analysis of such problems is complicated by the nonsmooth dependence of the eigenvalues on the design variables. Eigenvalue coalescence is the reason: for example, the circular disk has a Laplacian satisfying $\lambda_2 = \lambda_3$. Nonsmooth analysis is therefore a natural approach.

2 Notation

For brevity, I discuss only the finite-dimensional, real, self-adjoint case. In other words, I study the eigenvalues of matrices X in the space of $n \times n$ symmetric matrices S_n. I denote these (real) eigenvalues $\lambda_1(X) \geq \lambda_2(X) \geq \ldots \geq \lambda_n(X)$ (written out by multiplicity), and hence define a function $\lambda : S_n \to \mathbf{R}^n$. I endow S_n with the trace inner product, defined, for matrices Y in S_n, by $\langle X, Y \rangle = \operatorname{tr} XY$.

Simply put, my aim is to study the calculus of the function λ. As I noted, since eigenvalues may coalesce, λ is not smooth. For example, real α and β clearly satisfy

$$\lambda_1 \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} = \max\{\alpha, \beta\}.$$

More precisely, I study *eigenvalue functions* $G : S_n \to [-\infty, +\infty]$. These are functions satisfying the following two equivalent properties. (The group of $n \times n$ orthogonal matrices I denote O_n, and the subgroup of permutation matrices I denote P_n.)

- G is *orthogonally invariant*: that is, any matrices U in O_n and X in S_n satisfy $G(U^T XU) = G(X)$.
- $G = f \circ \lambda$, where the function $f : \mathbf{R}^n \to [-\infty, +\infty]$ is *permutation-invariant*: that is, any matrix P in P_n and vector x in $\mathbf{R}^n$ satisfy $f(Px) = f(x)$.

Clearly, the equivalence of these properties does not depend on the permutation-invariance of the function f. But this condition turns out to be very natural. For example, just as the eigenvalue function G depends on f through the formula

$$G(X) = f(\lambda(X)) \quad (X \in S_n),$$

so f depends on G through the formula

$$f(x) = G(\mathrm{Diag}\, x) \quad (x \in \mathbf{R}^n),$$

providing f is permutation-invariant. The calculus I seek analyzes the eigenvalue function $G = f \circ \lambda$ in terms of the vector function f (which is essentially the restriction of G to the diagonals).

Example 2.1 (Logarithmic Barrier) The classical logarithmic barrier function for the positive orthant is the permutation-invariant function

$$f(x) = \begin{cases} -\sum_{i=1}^n \log x_i, & \text{if } x > 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

It corresponds to the eigenvalue function

$$(f \circ \lambda)(X) = \begin{cases} -\log \det X, & \text{if } X \text{ positive definite,} \\ +\infty, & \text{otherwise,} \end{cases}$$

which is the classical logarithmic barrier for the positive definite cone.

Example 2.2 (Order Statistics) The k'th *order statistic* is the permutation-invariant function

$$\phi_k(x) = k\text{'th largest component of } x.$$

It corresponds to the k'th largest eigenvalue:

$$\phi_k \circ \lambda = \lambda_k.$$

3 Nonsmooth analysis

The k'th order statistic is a nonsmooth function, so variational analysis of it needs generalized ideas of gradient. In this section I therefore outline some central notions from nonsmooth analysis.

These notions originated in the convex case. For a convex function g on a Euclidean space E, we say an element v of E is a *subgradient* of g at the point u in E (written $v \in \partial g(u)$) if it satisfies

$$g(u+w) \geq g(u) + \langle v, w \rangle, \quad \text{for all } w \text{ in } E.$$

The set of subgradients $\partial g(u)$ we call the *subdifferential.* This idea has a number of merits.

Easy optimality conditions: a point maximizes g if and only if 0 is a subgradient there.

Good calculus: for example, useful sum and chain rules apply.

Stability: for lower semicontinuous g, the graph of ∂g is closed.

Nonconvex functions pose more difficulties. One natural approach studies good one-sided local approximations. Thus for any function $g : \mathbf{R}^n \to [-\infty, +\infty]$, we say v is a *Fréchet subgradient* at u (written $v \in \partial^F g(u)$) if it satisfies

$$g(u+w) \geq g(u) + \langle v, w \rangle + o(w), \quad \text{for all small } w \text{ in } E.$$

This notion is well-suited to necessary optimality conditions, and has good geometric properties. The following easy result is an example.

Proposition 3.1 (Orthogonality) *Suppose the point u is a local maximizer of the function g on a manifold M in E. Then any Fréchet subgradient of g at u is orthogonal to M at u.*

On the other hand, the Fréchet subdifferential is unstable. For example, the function $u \mapsto -|u|$ on $\mathbf{R}$ has a Fréchet subdifferential whose graph is not closed:

$$\partial^F(-|\cdot|)(u) = \begin{cases} 1, & \text{if } u < 0, \\ \emptyset, & \text{if } u = 0, \\ -1, & \text{if } u > 0. \end{cases}$$

We can correct this essentially by closing the graph. More precisely, we say v is a *limiting Fréchet subgradient* at u (written $v \in \partial^L g(u)$) if there are sequences of points u^r approaching u in E, with $g(u^r)$ approaching $g(u)$, and Fréchet subgradients v^r (at u^r) approaching v. In the example above, $\partial^L(-|\cdot|)(0)$ is the (nonconvex) set $\{-1, 1\}$.

The notion of the limiting Fréchet subdifferential is still effective for optimality conditions, is stable by construction, and, remarkably, satisfies good calculus rules, developed independently by Ioffe, Kruger and Mordukhovich in the 1980's. For more details, see [6].

The original idea of subgradients for nonconvex functions, due to Clarke, was rather different (see [2]). For a locally Lipschitz function g we define the Clarke subdifferential by the formula

$$\partial^C g(u) = \operatorname{conv}\{\lim g'(u^r) : u^r \to u\} :$$

this subdifferential in fact coincides with $\operatorname{conv} \partial^L g(u)$. Thus these three notions of subdifferential are ordered by inclusion:

$$\partial^F g(u) \subset \partial^L g(u) \subset \partial^C g(u).$$

In my example, $\partial^C(-|\cdot|)$ is the interval $[-1, 1]$.

Example 3.2 (Order Statistics) The k'th order statistic $\phi_k : \mathbf{R}^n \to \mathbf{R}$ is piecewise linear and Lipschitz, and a little thought shows that, assuming $k > 1$, it is convex near a point x in $\mathbf{R}^n$ exactly when x satisfies $\phi_{k-1}(x) > \phi_k(x)$. Denoting the standard unit vectors in $\mathbf{R}^n$ by $e^1, e^2, \ldots, e^n$, and their convex hull (the *simplex*) by Δ, a little calculation then shows

$$\partial^F \phi_k(x) = \begin{cases} \operatorname{conv}\{e^i : x_i = \phi_k(x)\}, & \text{if } \phi_{k-1}(x) > \phi_k(x), \\ \emptyset, & \text{otherwise.} \end{cases}$$

I denote the number of nonzero components of a vector y in $\mathbf{R}^n$ by $|\operatorname{supp} y|$. Taking limits and convex hulls using the formula above, I obtain the following subdifferentials at the origin:

$$\begin{aligned} \partial^F \phi_k(0) &= \emptyset \quad (\text{if } k > 1). \\ \partial^L \phi_k(0) &= \{y \in \Delta : |\operatorname{supp} y| \le n - k + 1\}. \\ \partial^C \phi_k(0) &= \Delta. \end{aligned}$$

Notice that the Clarke subdifferential is independent of k, and is much larger than the limiting Fréchet subdifferential.

4 Isospectral manifolds and diagonal reduction

The eigenvalue map λ has distinctive and beautiful properties. In this section I use the differential geometry of the *isospectral manifold*

$$\lambda^{-1}(\lambda(X)) = \{Z \in S_n : \lambda(Z) = \lambda(X)\} = \{U^T X U : U \in O_n\}$$

(for a fixed matrix X in S_n). The geometry reduces the question of characterizing subgradients $Y \in S_n$ of an eigenvalue function at X to the same question when X and Y are diagonal.

The following result is the key tool.

Theorem 4.1 (Arnold, 1971) *For any matrix X in S_n, the set of matrices $\lambda^{-1}(\lambda(X))$ is a manifold in S_n, and a matrix Y in S_n is orthogonal to this manifold at X if and only if Y commutes with X.*

Given an eigenvalue function G, suppose $Y \in \partial^F G(X)$. Since G is clearly constant on the isospectral manifold $\lambda^{-1}(\lambda(X))$, the Orthogonality Proposition (3.1) and Arnold's Theorem (above) show Y commutes with X. Taking limits demonstrates the same property for $\partial^L G(X)$, and taking convex hulls proves it for $\partial^C G(X)$. In summary, if the symbol # is either F, L, or (in the locally Lipschitz case) C, then I have proved

$$Y \in \partial^\# G(X) \Rightarrow XY = YX.$$

Commuting matrices in S_n are simultaneously diagonalizable. Hence $Y \in \partial^\# G(X)$ implies the existence of a matrix U in S_n and vectors x and y in $\mathbf{R}^n$ satisfying

$$U^T X U = \mathrm{Diag}\, x \quad \text{and} \quad U^T Y U = \mathrm{Diag}\, y.$$

Since the function G is orthogonally invariant, I deduce easily $\mathrm{Diag}\, y \in \partial^\# G(\mathrm{Diag}\, x)$.

5 The diagonal case and the main result

Given a permutation-invariant function f, I have reduced the problem of characterizing subgradients of the eigenvalue function $f \circ \lambda$ to the diagonal case: which vectors x and y in $\mathbf{R}^n$ satisfy $\mathrm{Diag}\, y \in \partial^\# (f \circ \lambda)(\mathrm{Diag}\, x)$?

Checking the definition of Fréchet subgradients shows immediately

$$\mathrm{Diag}\, y \in \partial^{\#}(f \circ \lambda)(\mathrm{Diag}\, x) \quad \Rightarrow \quad y \in \partial^{\#} f(x).$$

In fact the converse is also true, but seems to depend on rather more sophisticated properties of the eigenvalue map λ: [4] contains the details. One approach uses the following result (see for example [3]), an 'envelope' representation of any nonincreasing linear combination of the eigenvalues as the maximum of a collection of linear functions.

Theorem 5.1 (Fan, 1949) *Any matrix X in S_n and reals $d_1 \geq d_2 \geq \ldots \geq d_n$ satisfy*

$$d^T \lambda(X) = \max\{\mathrm{tr}\, U^T X U \mathrm{Diag}\, d : U \in O_n\}.$$

I summarize the diagonal case below.

Theorem 5.2 (Diagonal Subgradients) *For any permutation-invariant function f, any vectors x and y in $\mathbf{R}^n$ satisfy the following relationship (in any of the Fréchet, limiting Fréchet, or locally Lipschitz Clarke cases):*

$$\mathrm{Diag}\, y \in \partial^{\#}(f \circ \lambda)(\mathrm{Diag}\, x) \quad \Leftrightarrow \quad y \in \partial^{\#} f(x).$$

Combining this result with the diagonal reduction argument of the previous section, I immediately obtain my main result.

Theorem 5.3 (Subgradients) *Consider the Fréchet, limiting Fréchet, or (in the locally Lipschitz case) Clarke subgradients of an eigenvalue function $f \circ \lambda$ (where the function $f : \mathbf{R}^n \to [-\infty, +\infty]$ is permutation-invariant). In any of these cases (so with # denoting F, L, or C), matrices X and Y in S_n satisfy*

$$Y \in \partial^{\#}(f \circ \lambda)(X)$$

if and only if there are vectors x and y in $\mathbf{R}^n$ satisfying

$$y \in \partial^{\#} f(x)$$

and a matrix U in O_n satisfying

$$U^T X U = \mathrm{Diag}\, x \quad \textit{and} \quad U^T Y U = \mathrm{Diag}\, y.$$

Similar arguments show the analogous result for the 'horizon' subgradients described in [6].

Example 5.4 (Eigenvalues) The k'th largest eigenvalue λ_k is exactly the composite function $\phi_k \circ \lambda$, where ϕ_k is the k'th order satistic (whose subgradient I calculated in Section 3). Applying the main result, I obtain the following subdifferentials at the origin. (The cone of positive semidefinite matrices in S_n I denote S_n^+.)

$$\begin{aligned}
\partial^F \lambda_k(0) &= \emptyset \quad (\text{if } k > 1). \\
\partial^L \lambda_k(0) &= \{Y \in S_n^+ : \operatorname{tr} Y = 1,\ \operatorname{rank} Y \le n - k + 1\}. \\
\partial^C \lambda_k(0) &= \{Y \in S_n^+ : \operatorname{tr} Y = 1\}.
\end{aligned}$$

Notice once again that the Clarke subdifferential does not depend on k, and is much larger than the limiting Fréchet subdifferential. The calculations of these subdifferentials at an arbitrary point closely resemble those at the origin: see [4] for details.

6 Concluding remarks

The various subdifferentials I consider in this paper illustrate a broader observation: many first order properties of eigenvalue functions depend only on their behaviour on the diagonal matrices. The reader may find more details in [5, 4]. I hope my main result, characterizing subgradients of eigenvalue functions, advances our understanding of eigenvalue optimization, both in theory and in practice. In any case, I believe my results demonstrate the elegance of eigenvalue functions as a class of examples for nonsmooth analysis.

References

[1] M.S. Ashbaugh and R.D. Benguria. Proof of the Payne-Pólya-Weinberger conjecture. *Bulletin of the American Mathematical Society*, 25:19–29, 1991.

[2] F.H. Clarke. *Optimization and nonsmooth analysis.* Wiley, New York, 1983.

[3] R.A. Horn and C. Johnson. *Matrix analysis.* Cambridge University Press, Cambridge, U.K., 1985.

[4] A.S. Lewis. Nonsmooth analysis of eigenvalues. Forthcoming.

[5] A.S. Lewis and M.L. Overton. Eigenvalue optimization. *Acta Numerica*, 1996. To appear.

[6] R.T. Rockafellar and R.J.B. Wets. Variational analysis. To appear, 1996.