Math. Program. 84: 1–24 (1999) / DOI 10.1007/s10107980004a

A.S. Lewis*

Nonsmooth analysis of eigenvalues

Received October 28, 1996 / Revised version received January 28, 1998
Published online October 9, 1998

Abstract. The eigenvalues of a symmetric matrix depend on the matrix nonsmoothly. This paper describes the nonsmooth analysis of these eigenvalues. In particular, I present a simple formula for the approximate (limiting Fréchet) subdifferential of an arbitrary function of the eigenvalues, subsuming earlier results on convex and Clarke subgradients. As an example I compute the subdifferential of the k'th largest eigenvalue.

Key words. eigenvalue optimization – nonsmooth analysis – approximate subdifferential – Clarke subgradient – generalized derivative – horizon subgradient – semidefinite program

1. Introduction

Two opposite perspectives motivate this work. First, variational properties of the eigenvalues of a symmetric matrix are fundamental in several areas of optimization: since these eigenvalues vary nonsmoothly with the matrix, nonsmooth analysis is the obvious language of study. Secondly, the health of the area of nonsmooth analysis depends ultimately on its power to illuminate interesting examples: eigenvalue optimization is such an example, both in its relevance and challenge.

Sensitivity of the eigenvalues of a matrix with respect to a single parameter is a classical subject. (An excellent standard reference is [13].) Engineering design abounds with eigenvalue optimization problems involving many parameters (see the recent survey paper [19], for example), and over the last few years semidefinite programming in particular has been widely studied, mainly due to its amenability to interior point methods (see the survey [27], for example).

By contrast with the classical approach of [13], one may consider the eigenvalues of a matrix X in the Euclidean space of $n \times n$ real symmetric matrices $S(n)$ directly as functions of X (see for example [5, 22, 23, 9]). Writing these eigenvalues (by multiplicity) $\lambda_1(X) \geq \lambda_2(X) \geq \ldots \geq \lambda_n(X)$, I define the 'eigenvalue' map

$$\lambda : S(n) \rightarrow \mathbf{R}^n ,$$

A.S. Lewis: Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1, e-mail: `aslewis@orion.uwaterloo.ca`, `http://orion.uwaterloo.ca/~aslewis`

Mathematics Subject Classification (1991): primary: 90C31, 15A18, secondary: 49K40, 26B05

* The author wishes to thank the LAO, Paul Sabatier University, Toulouse, France, for their hospitality during the writing of this article. Research partially supported by the Natural Sciences and Engineering Research Council of Canada

having components $\lambda_1, \lambda_2, \ldots, \lambda_n$. The variational analysis of this map is the subject of this paper.

Any extended-real-valued function of the eigenvalues of the matrix X can be written in the composite form

$$f \circ \lambda : S(n) \to [-\infty, +\infty], \tag{1}$$

where the function $f : \mathbf{R}^n \to [-\infty, +\infty]$ is invariant under coordinate permutations. My main result (Theorem 6) identifies the subdifferential $\partial(f \circ \lambda)(X)$ as the set of matrices

$$\{U^T(\mathrm{Diag}\,\mu)U : U \text{ orthogonal},\ U^T(\mathrm{Diag}\,\lambda(X))U = X, \mu \in \partial f(\lambda(X))\}\ ,$$

where $\mathrm{Diag}\,\mu$ denotes the diagonal matrix with diagonal entries $\mu_1, \mu_2, \ldots, \mu_n$. Thus the subgradients of the underlying function f characterize the subgradients of the matrix function $f \circ \lambda$.

By 'subdifferential' I mean the 'approximate subdifferential' first investigated by Mordukhovich [20], and studied also by Kruger [14] and Ioffe [11]. Of the many notions introduced in recent years, this subdifferential is the smallest satisfying reasonable calculus rules, and is therefore often the object of choice in nonconvex optimization. The book [25] revolves around this idea. The subdifferential formula above remains valid for the Fréchet subdifferential, and, by taking convex hulls, for the Clarke subdifferential (when f is locally Lipschitz). When f is convex, the approximate and usual subdifferentials coincide. This paper therefore unifies, simplifies, and refines the subdifferential results in [15] and [16]. The corresponding results for Hermitian matrices are completely analogous.

Nonsmooth analysts will ask why I have not simply applied a standard chain rule to the composite function (1). Such a rule (see for example [12, Cor 5.3]) shows

$$\partial(f \circ \lambda) \subset \bigcup \{\partial(y^T\lambda)(X) : y \in \partial f(\lambda(X))\}\ .$$

This formula has a number of drawbacks. First, it is only an inclusion, and the usual conditions for equality do not obviously hold (see for example [25, Thm 10.49]). Secondly, calculating $\partial(y^T\lambda)(X)$ does not seem essentially any easier than the original calculation. Lastly, even after circumventing the first two difficulties, using this result to derive the elegant subdifferential formula above does not appear immediate. I have therefore taken a direct approach.

The subdifferential formula provides some pretty results. For example, the approximate subdifferential of the k'th largest eigenvalue at the zero matrix consists of those positive semidefinite matrices with trace 1 and rank at most $n - k + 1$. Omitting the rank condition gives the Clarke subdifferential, a striking illustration of the distinction between the two notions.

To summarize, functions of eigenvalues provide an important testing-ground for nonsmooth analysis. I hope these results also prove useful to practitioners of eigenvalue optimization.

2. The approximate subdifferential

This paper concerns the first order behaviour of functions of the eigenvalues of a symmetric matrix variable. Optimization theorists study first order behaviour via generalized gradients: the convex subdifferential is prototypical [24]. For nonconvex functions, various generalized gradients have been studied, inspired originally by the ideas of Clarke [3]. More recently, a notion of subdifferential originating with ideas of Ioffe, Mordhukovich and Kruger has received increasing attention. This notion, commonly called the 'approximate subdifferential', is the one I study in this work.

I follow the terminology and notation of [25]. Given a Euclidean space E (by which I mean a finite-dimensional real inner-product space), a function $f : E \to [-\infty, \infty]$, and a point x in E at which f is finite, an element y of E is a *regular subgradient* of f at x if it satisfies

$$f(x+z) \geq f(x) + \langle y, z\rangle + o(z) \text{ as } z \to 0 \text{ in } E\ .$$

As usual, $o(\cdot)$ denotes a real-valued function defined on a neighbourhood of the origin in E, and satisfying $\lim_{z\to 0} \|z\|^{-1} o(z) = 0$. The set of regular subgradients is denoted $\hat{\partial} f(x)$: it is always closed and convex.

This definition is just a one-sided version of the classical (Fréchet) derivative. Indeed, it is also known as the 'Fréchet subdifferential'. Without modificiation, this natural concept of subdifferential has some disadvantages: even for well-behaved functions f, it may, for example, be empty. The idea of the approximate subdifferential resolves these difficulties by a process of 'stabilization'. An element y of E is an *(approximate) subgradient* if there is a sequence of points x^r in E approaching x with values $f(x^r)$ approaching the finite value $f(x)$, and a sequence of regular subgradients y^r in $\hat{\partial} f(x^r)$ approaching y. The set of all subgradients is the *(approximate) subdifferential* $\partial f(x)$. Common alternative names are 'limiting Fréchet subgradient (subdifferential)'. If, instead of y^r approaching y in this definition, there is a sequence of reals t_r decreasing to 0 for which $t_r y^r$ approaches y, then y is a *horizon subgradient*: the set of horizon subgradients is denoted $\partial^\infty f(x)$. If $f(x)$ is infinite then the sets $\partial f(x)$ and $\hat{\partial} f(x)$ are defined to be empty, and $\partial^\infty f(x)$ to be $\{0\}$. If f is proper and convex, both $\partial f(x)$ and $\hat{\partial} f(x)$ coincide with the usual convex subdifferential. Finally, if the function f is finite at the point x with at least one subgradient there then it is *(Clarke) regular* at x if it is lower semicontinuous near x, every subgradient is regular, and furthermore

$$\partial^\infty f(x) = (\hat{\partial} f(x))^\infty$$

(where C^∞ denotes the recession cone of a closed convex set C) — see [25, Cor 8.11].

For a function f which is locally Lipschitz around x, convex combinations of subgradients are called *Clarke subgradients*. The set of Clarke subgradients is the *Clarke subdifferential* $\partial^c f(x)$. (This definition is equivalent to the standard one in [3] — see for example [12, Thm 2].)

Let L be a subset of the space E, and fix a point x in E. An element d of E belongs to the *contingent cone* to L at x, written $K(L|x)$, if either $d = 0$ or there is a sequence

(x^r) in L approaching x with $\|x^r - x\|^{-1}(x^r - x)$ approaching $\|d\|^{-1}d$. The *(negative) polar* of a subset H of E is the set

$$H^- = \{y \in E : \langle x, y\rangle \leq 0 \ \forall x \in H\} .$$

I use the following easy and standard result later.

Proposition 1 (Normal cone). *Given a function $f : E \to [-\infty, +\infty]$ and a point x^0 in E, any regular subgradient of f at x^0 is polar to the level set $L = \{x \in E : f(x) \leq f(x^0)\}$ at x^0: that is,*

$$\hat{\partial} f(x^0) \subset (K(L|x^0))^- .$$

Proof. For any regular subgradient y and any element d of $K(L|x^0)$, I want to show $\langle y, d\rangle \leq 0$. I can assume f is finite at x^0 and d is nonzero, so there is a sequence (x^r) approaching x^0 with $f(x^r) \leq f(x^0)$ for each index r such that if I define vectors $z^r = x^r - x^0$ then $\|z^r\|^{-1}z^r$ approaches $\|d\|^{-1}d$. By definition,

$$f(x^0) \geq f(x^r) \geq f(x^0) + \langle y, z^r\rangle + o(z^r)$$

for each r; dividing through by $\|z^r\|$ and taking the limit completes the proof.

□

The functions in this paper have important invariance properties which I continually exploit. A linear transformation g on the space E is *orthogonal* if it preserves the inner product: that is

$$\langle gx, gy\rangle = \langle x, y\rangle \ \text{ for all elements } x \text{ and } y \text{ of } E \ ,$$

Such linear transformations comprise a group $O(E)$. A function f on E is *invariant* under a subgroup G of $O(E)$ if $f(gx) = f(x)$ for all points x in E and transformations g in G.

In the following result, $f'(\cdot;\cdot)$ denotes the usual directional derivative:

$$f'(x; z) = \lim_{t \downarrow 0} \frac{f(x + tz) - f(x)}{t}, \quad \text{(when well-defined)}$$

for elements x and z of E.

Proposition 2 (Subgradient invariance). *If the function $f : E \to [-\infty, +\infty]$ is invariant under a subgroup G of $O(E)$, then any point x in E and transformation g in G satisfy $\partial f(gx) = g\partial f(x)$. Corresponding results hold for regular, horizon, and (if f is Lipschitz around x) Clarke subgradients, and f is regular at the point gx if and only if it is regular at x. Furthermore, for any element z of E, the directional derivative $f'(gx; gz)$ exists if and only if $f'(x; z)$ does, and in this case the two are equal.*

Proof. The result about directional derivatives follows immediately from the definition. Turning to the claims about subgradients, suppose first $y \in \hat{\partial} f(x)$, so small elements z of E satisfy $f(x+z) \geq f(x) + \langle y, z\rangle + o(z)$. The invariance of f shows

$$\begin{aligned} f(gx+z) &= f(x+g^{-1}z) \\ &\geq f(x) + \langle y, g^{-1}z\rangle + o(g^{-1}z) \\ &= f(gx) + \langle gy, z\rangle + o(z) \ , \end{aligned}$$

whence $gy \in \hat{\partial} f(gx)$.

Now suppose $y \in \partial f(x)$, so for a sequence of points x^r in E approaching x with $f(x^r)$ approaching $f(x)$, there is a sequence of regular subgradients y^r in $\hat{\partial} f(x^r)$ approaching y. Consequently, gx^r approaches gx with $f(gx^r) = f(x^r) \to f(x) = f(gx)$, and gy^r approaches gy with, by the above, $gy^r \in \hat{\partial} f(gx^r)$. Hence $gy \in \partial f(gx)$, as I claimed. The converse is immediate, and the horizon subgradient case is almost identical. The Clarke case follows from the observation

$$\partial^c f(gx) = \operatorname{conv} \partial f(gx) = \operatorname{conv} g\partial f(x) = g\operatorname{conv} \partial f(x) = g\partial^c f(x) \ .$$

The regularity claim follows quickly from these results: regularity of f at x implies $\partial f(gx) = g\partial f(x) = g\hat{\partial} f(x) = \hat{\partial} f(gx) \neq \emptyset$, and it is also easy to check $\partial^\infty f(gx) = g\partial^\infty f(x) = g(\hat{\partial} f(x))^\infty = (g\hat{\partial} f(x))^\infty = (\hat{\partial} f(gx))^\infty$. The result follows.

□

This section ends with a lemma which is useful in the later analysis of regularity.

Lemma 1 (Recession). *For any nonempty closed convex subset C of E, closed subgroup H of $O(E)$, and transformation g in $O(E)$, the set gHC is closed, and if it is also convex then its recession cone is $gH(C^\infty)$.*

Proof. Assume g is the identity: the general case follows easily. Given sequences (x^r) in C and (h_r) in H, suppose $h_r x^r$ approaches z. Since H is compact, without loss of generality assume h_r approaches a transformation h in H. Thus $h_r(x^r - h^{-1}z)$ approaches zero, and since the transformations h_r preserve the norm, x^r approaches the point $h^{-1}z$, which therefore lies in C. Thus $z \in HC$, so HC is closed.

Fix a point x in C. Given any vector d in C^∞ and transformation h in H, I know $x + td \in C$ for all positive real t, and hence $hx + t(hd) \in HC$. Thus hd belongs to $(HC)^\infty$, and so I deduce $H(C^\infty) \subset (HC)^\infty$.

Conversely, given any vector d in $(HC)^\infty$, I know $x + td \in HC$ for all positive real t. Since H is compact, there is a sequence of positive reals t_r approaching $+\infty$, and a sequence of transformations h_r approaching h in H, satisfying $x + t_r d \in h_r C$ for each r. I deduce, for each r

$$h_r^{-1}d + t_r^{-1}(h_r^{-1}x - x) \in t_r^{-1}(C - x) \ ,$$

and letting r approach ∞ shows $h^{-1}d \in C^\infty$, so $d \in H(C^\infty)$, as I claimed.

□

3. The normal space

This section describes a fundamental result for variational properties of eigenvalues. The result is not new, and it is possible to give an elementary (though somewhat long) proof. However, since some basic differential geometry greatly enhances understanding of this key result, and since a self-contained proof for nonspecialists seems not readily accessible, I sketch the proof from this perspective.

Consider the group of $n \times n$ real orthogonal matrices, $O(n)$. This set is a 'submanifold' of the Euclidean space of all $n \times n$ real matrices, $M(n)$. All we need to know here about a 'manifold' in Euclidean space is that, locally, it is the solution set of a smooth equation with surjective Jacobian, and hence the 'tangent space' coincides with the contingent cone. For example, an easy calculation shows that the tangent space to $O(n)$ at the identity matrix I, denoted $T_I(O(n))$, is just the subspace of skew-symmetric matrices, $A(n)$.

Consider the 'adjoint' action of the group $O(n)$ on the Euclidean space of $n \times n$ real symmetric matrices $S(n)$ (with the inner product $\langle X, Y \rangle = \mathrm{tr}\, XY$), defined by $U.X = U^T XU$, for all U in $O(n)$ and X in $S(n)$. Notice that any vector x in $\mathbf{R}^n$ and matrix P in the group of $n \times n$ permutation matrices $P(n)$ satisfy $\mathrm{Diag}\,(Px) = P^T.\mathrm{Diag}\, x$.

For a fixed matrix X in $S(n)$, the orbit $O(n).X = \{U^T XU : U \in O(n)\}$ is just the set of symmetric matrices with the same eigenvalues (and multiplicities) as X. Here, then, is the key fact (c.f. [1, p. 243] and [7, p. 150]).

Theorem 1 (Normal space). *The orbit $O(n).X$ is a submanifold of the space $S(n)$, with tangent space*

$$T_X(O(n).X) = \{XZ - ZX : Z \in A(n)\}, \tag{2}$$

and normal space

$$(T_X(O(n).X))^{\perp} = \{Y \in S(n) : XY = YX\}. \tag{3}$$

(Sketch proof). Consider the 'stabilizer'

$$O(n)_X = \{U \in O(n) : U^T XU = X\} .$$

The idea of the proof is to relate the orbit $O(n).X$ with the quotient group $O(n)/O(n)_X$. This quotient group itself can be given the structure of a manifold (turning it into a 'homogeneous space'). All we need to know about this construction is that the map

$$\begin{aligned} \phi : O(n)/O(n)_X &\to O(n).X, \quad \text{defined by} \\ U(O(n)_X) &\mapsto U^T XU, \quad \text{for } U \text{ in } O(n), \end{aligned}$$

is then a diffeomorphism, and hence its differential $d\phi$ is an isomorphism between the corresponding tangent spaces

$$T_{O(n)_X}(O(n)/O(n)_X) \text{ and } T_X(O(n).X) .$$

(See for example [7, p. 150, C5] and [2, p. 108] for details.)

Consider, on the other hand, the quotient map

$$\begin{aligned} \pi : O(n) &\to O(n)/O(n)_X, \quad \text{defined by} \\ U &\mapsto U(O(n)_X), \quad \text{for all } U \text{ in } O(n). \end{aligned}$$

The differential of this map,

$$d\pi : T_I(O(n)) \to T_{O(n)_X}(O(n)/O(n)_X)$$

has kernel $T_I(O(n)_X)$ (see for instance [2, p. 165]), and hence, by counting dimensions, is onto.

Now consider a third map

$$\begin{aligned} \psi : O(n) &\to O(n).X, \quad \text{defined by} \\ U &\mapsto U^T XU, \quad \text{for } U \text{ in } O(n). \end{aligned}$$

Since $\psi = \phi \circ \pi$, the chain rule (at I) gives

$$(d\psi)T_I(O(n)) = T_X(O(n).X) \ .$$

But as I noted above, $T_I(O(n))$ is just $A(n)$, and an easy calculation shows that any matrix Z in $A(n)$ satisfies $(d\psi)Z = XZ - ZX$. Equation (2) now follows.

Finally, if a matrix Y in $S(n)$ commutes with X then any matrix Z in $A(n)$ satisfies

$$\text{tr}\,(XZ - ZX)Y = \text{tr}\,(YX)Z - \text{tr}\,Z(YX) = 0 \ .$$

Conversely, suppose

$$\text{tr}\,(XZ - ZX)Y = 0 \ \text{ for all } Z \text{ in } A(n) \ .$$

Choosing $Z = XY - YX$ gives

$$\begin{aligned} 0 &= \text{tr}\,(X(XY - YX) - (XY - YX)X)Y \\ &= \text{tr}\,XY^2X + \text{tr}\,YX^2Y - \text{tr}\,XYXY - \text{tr}\,YXYX \\ &= -\text{tr}\,(XY - YX)(XY - YX), \end{aligned}$$

whence $XY = YX$. Equation (3) follows.

□

I support my claim that the Normal Space Theorem (1) is fundamental in the following sections. However, as an immediate application, I next use it to derive an important inequality, essentially due to von Neumann [21] (the condition for equality coming from [26]).

The following rather standard combinatorial lemma is an essential tool in the proof (see [15]). I denote the cone of vectors x in $\mathbf{R}^n$ satisfying $x_1 \geq x_2 \geq \ldots \geq x_n$ by $\mathbf{R}^n_{\geq}$.

Lemma 2. *Any vectors x and y in $\mathbf{R}^n_{\geq}$ and any matrix P in $P(n)$ satisfy the inequality $x^T Py \leq x^T y$; equality holds if and only if some matrix Q in $P(n)$ satisfies $Qx = x$ and $Qy = Py$.*

Two matrices X and Y have a *simultaneous ordered spectral decomposition* if there is a matrix U in $O(n)$ satisfying $X = U.\mathrm{Diag}\,\lambda(X)$ and $Y = U.\mathrm{Diag}\,\lambda(Y)$.

Theorem 2 (Von Neumann–Theobald). *Any matrices X and Y in $S(n)$ satisfy the inequality* $\mathrm{tr}\,XY \le \lambda(X)^T\lambda(Y)$*; equality holds if and only if X and Y have a simultaneous ordered spectral decomposition.*

Proof. For fixed X and Y, consider the optimization problem

$$\alpha = \sup_{Z \in O(n).X} \mathrm{tr}\,YZ. \tag{4}$$

There is a matrix U in $O(n)$ satisfying $Y = U.\mathrm{Diag}\,\lambda(Y)$, and then choosing $Z = U.\mathrm{Diag}\,\lambda(X)$ shows $\alpha \ge \lambda(X)^T\lambda(Y)$.

On the other hand, since the orbit $O(n).X$ is compact, problem (4) has an optimal solution, $Z = Z_0$ say, which by stationarity must satisfy

$$Y \perp T_{Z_0}(O(n).X) \ \ (= T_{Z_0}(O(n).Z_0))\ .$$

The Normal Space Theorem now shows that the matrices Y and Z_0 commute and hence there is a matrix U in $O(n)$ simultaneously diagonalizing them:

$$Y = U.\mathrm{Diag}\,(P\lambda(Y)) \ \text{ and } \ Z_0 = U.\mathrm{Diag}\,\lambda(Z_0), \tag{5}$$

for some matrix P in $P(n)$. Hence

$$\alpha = \mathrm{tr}\,YZ_0 = \lambda(Z_0)^T P\lambda(Y) \le \lambda(Z_0)^T\lambda(Y) = \lambda(X)^T\lambda(Y) \le \alpha\ ,$$

using Lemma 2, whence $\alpha = \lambda(X)^T\lambda(Y)$, and some matrix Q in $P(n)$ satisfies $Q\lambda(Z_0) = \lambda(Z_0)$ and $Q\lambda(Y) = P\lambda(Y)$: combining this with the decompositions (5) gives a simultaneous ordered spectral decomposition of Y and Z_0. The result now follows.

□

This section ends with another simple linear-algebraic result which is useful later.

Proposition 3 (Simultaneous conjugacy). *Given vectors x, y, u and v in $\mathbf{R}^n$, there is a matrix U in $O(n)$ with* $\mathrm{Diag}\,x = U.\mathrm{Diag}\,u$ *and* $\mathrm{Diag}\,y = U.\mathrm{Diag}\,v$ *if and only if there is a matrix P in $P(n)$ with $x = Pu$ and $y = Pv$.*

Proof. The ‘only if’ case with $x = u$ is Lemma 3.5 in [18], and the general case reduces to this case by a simple trick. Considering eigenvalues shows there is a matrix Q in $P(n)$ with $u = Qx$. Hence I can write

$$\mathrm{Diag}\,x = (Q^TU).\mathrm{Diag}\,x \ \text{ and } \ \mathrm{Diag}\,y = (Q^TU).\mathrm{Diag}\,(Q^Tv)\ .$$

By the result I quoted above, there is a matrix R in $P(n)$ with

$$x = Rx = RQ^Tu \ \text{ and } \ y = RQ^Tv\ ,$$

so I can choose $P = RQ^T$. Conversely, for the ‘if’ direction, I can choose $U = P^T$.

□

4. Simultaneous diagonalization

Functions of the eigenvalues of a symmetric matrix are the subject of this paper. Two slightly different perspectives help. The first views such functions as those which are invariant under orthogonal similarity transformations, while the second directly considers functions of the eigenvalues $\lambda_i(\cdot)$. The following trivial result shows these two perspectives are equivalent.

Proposition 4 (Invariant functions). *The following two properties of a matrix function* $F : S(n) \to [-\infty, +\infty]$ *are equivalent:*

(i) F *is* **orthogonally invariant***; that is, any matrices* X *in* $S(n)$ *and* U *in* $O(n)$ *satisfy* $F(U.X) = F(X)$.
(ii) $F = f \circ \lambda$ *for some* **permutation-invariant** *function* $f : \mathbf{R}^n \to [-\infty, +\infty]$ *(that is, any vector* x *in* $\mathbf{R}^n$ *and matrix* P *in* $P(n)$ *satisfy* $f(Px) = f(x)$*).*

Proof. The implication $(ii) \Rightarrow (i)$ follows from the invariance of eigenvalues under orthogonal similarity. To see the converse, define $f(x) = F(\operatorname{Diag} x)$.

□

Definition 1. *An* **eigenvalue function** *is an extended-real-valued function on* $S(n)$ *of the form* $f \circ \lambda$ *for a function* $f : \mathbf{R}^n \to [-\infty, +\infty]$*: the function* f *is understood to be permutation-invariant.*

Many elementary but important properties of eigenvalue functions $f \circ \lambda$ follow from the facts that the function λ is globally Lipschitz (with constant 1 — see [17, Thm 2.4]) and that $f = (f \circ \lambda) \circ \operatorname{Diag}$. In particular, notice that $f \circ \lambda$ is continuous (respectively lower semicontinuous, Lipschitz) at a matrix X in $S(n)$ if and only if f is continuous (respectively lower semicontinuous, Lipschitz) at $\lambda(X)$.

My analysis of eigenvalue functions depends heavily on the following fundamental fact.

Theorem 3 (Commutativity). *If a matrix* Y *in* $S(n)$ *is a subgradient or horizon subgradient of an eigenvalue function at a matrix* X *in* $S(n)$ *then* X *and* Y *commute. Furthermore, if the eigenvalue function is Lipschitz around* X*, and* Y *is a Clarke subgradient there, then* X *and* Y *commute.*

Proof. Call the eigenvalue function F, and assume first that the subgradient Y is regular. By the Normal Cone Proposition (1), the constancy of F on the orbit $O(n).X$ shows

$$\begin{aligned} Y &\in (K(\{Z : F(Z) \le F(X)\} \mid X))^- \\ &\subset (K(O(n).X \mid X))^- \\ &= (T_X(O(n).X))^{\perp} \end{aligned}$$

(since, as I mentioned, the tangent space coincides with the contingent cone). The result follows from the Normal Space Theorem (1).

Now suppose Y is a subgradient of F at X, so for some sequence of matrices X_r in $S(n)$ approaching X there is a corresponding sequence of regular subgradients Y_r in $\hat{\partial} F(X_r)$ approaching Y. By the first part,

$$XY = \lim_r X_r Y_r = \lim_r Y_r X_r = YX \ .$$

If Y is a horizon subgradient then instead of Y_r approaching Y, we have $t_r Y_r$ approaching Y, where the reals t_r decrease to 0. Thus

$$XY = \lim_r X_r t_r Y_r = \lim_r t_r Y_r X_r = YX \ .$$

If the eigenvalue function is locally Lipschitz then any Clarke subgradient is a convex combination of subgradients, and since every subgradient commutes with X, so must any convex combination.

□

Commuting symmetric matrices are simultaneously digonalizable. Hence if the matrix Y in $S(n)$ is a subgradient of some eigenvalue function at the matrix X in $S(n)$ then some vectors x and y in $\mathbf{R}^n$ and some matrix U in $O(n)$ satisfy

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

Consequently, by the Subgradient Invariance Proposition (2) applied to the space $S(n)$ with the adjoint action of the group $O(n)$, Diag y must be a subgradient at Diag x. Thus to characterize when a matrix Y is a subgradient of an eigenvalue function at a matrix X, it suffices to consider the case when X and Y are both diagonal. In one direction this is easy, as I show below.

Proposition 5. *Any vectors x and y in $\mathbf{R}^n$ and eigenvalue function $f \circ \lambda$ satisfy*

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

Corresponding results hold for regular and horizon subgradients.

Proof. Suppose first that Diag y is a regular subgradient. For a small vector z in $\mathbf{R}^n$ I obtain

$$\begin{aligned} f(x+z) &= (f \circ \lambda)(\mathrm{Diag}\, x + \mathrm{Diag}\, z) \\ &\geq (f \circ \lambda)(\mathrm{Diag}\, x) + \mathrm{tr}\,(\mathrm{Diag}\, y)(\mathrm{Diag}\, z) + o(\mathrm{Diag}\, z) \\ &= f(x) + y^T z + o(z), \end{aligned}$$

whence $y \in \hat{\partial} f(x)$.

Now assume Diag $y \in \partial(f \circ \lambda)(\mathrm{Diag}\, x)$, so there is a sequence of matrices X_r in $S(n)$ approaching Diag x, with $f(\lambda(X_r))$ approaching $f(x)$, and a sequence of regular subgradients Y_r in $\hat{\partial}(f \circ \lambda)(X_r)$ approaching Diag y. By the Commutativity Theorem (3) there are sequences of vectors x^r and y^r in $\mathbf{R}^n$ and matrices U_r in $O(n)$ with

$$X_r = U_r.\mathrm{Diag}\, x^r \ \text{ and } \ Y_r = U_r.\mathrm{Diag}\, y^r \tag{6}$$

for each index r. The Subgradient Invariance Proposition (2) now shows $\mathrm{Diag}\, y^r \in \hat{\partial}(f \circ \lambda)(\mathrm{Diag}\, x^r)$, whence by the first part, $y^r \in \hat{\partial} f(x^r)$.

Since the group $O(n)$ is compact, I can assume U_r approaches a matrix U in $O(n)$. Hence, from equations (6), there must be vectors u and v in $\mathbf{R}^n$ with

$$U^T.\mathrm{Diag}\, x = \mathrm{Diag}\, u \quad \text{and} \quad U^T.\mathrm{Diag}\, y = \mathrm{Diag}\, v, \tag{7}$$

and u and v must be the limits of the sequences (x^r) and (y^r) respectively. Since the first equation of (7) demonstrates $f(u) = f(x)$, v belongs to $\partial f(u)$. But equations (7) also guarantee the existence of a matrix P in $P(n)$ with $x = Pu$ and $y = Pv$, by the Simultaneous Conjugacy Proposition (3). Applying the Subgradient Invariance Proposition (2) again, this time to the space $\mathbf{R}^n$ with the group $P(n)$, y belongs to $\partial f(x)$, as I claimed. The horizon subgradient argument is almost identical.

□

5. Directional derivatives of eigenvalues

There is one missing ingredient for characterizing subgradients of eigenvalue functions. I need, for vectors x and y in $\mathbf{R}^n$ and an eigenvalue function $f \circ \lambda$, to show

$$y \in \hat{\partial} f(x) \quad \Rightarrow \quad \mathrm{Diag}\, y \in \hat{\partial}(f \circ \lambda)(\mathrm{Diag}\, x). \tag{8}$$

This implication, which is the aim of this section, rests on somewhat deeper properties of the eigenvalue map λ. I begin with some easy preliminary results. For a vector x in $\mathbf{R}^n$, I write $\bar{x}$ for the vector (in $\mathbf{R}^n_{\geq}$) with the same components arranged in nonincreasing order.

Lemma 3. *For any vector w in $\mathbf{R}^n_{\geq}$, the function $w^T\lambda$ is convex, and any vector x in $\mathbf{R}^n_{\geq}$ satisfies* $\mathrm{Diag}\, w \in \partial(w^T\lambda)(\mathrm{Diag}\, x)$.

Proof. The permutation-invariant function $f : \mathbf{R}^n \to \mathbf{R}$ defined by $f(z) = w^T\bar{z}$ is convex, since

$$f(z) = \max\{w^T Pz : P \in P(n)\}\ ,$$

by Lemma 2. It suffices to show that any matrix Z in $S(n)$ satisfies

$$\mathrm{tr}\,(\mathrm{Diag}\, w)(Z - \mathrm{Diag}\, x) \leq w^T\lambda(Z) - w^T x\ ,$$

or in other words, $\mathrm{tr}\,(\mathrm{Diag}\, w)Z \leq w^T\lambda(Z)$. This inequality follows from the Von Neumann-Theobald Theorem (2).

□

For a vector x in $\mathbf{R}^n$, I write $P(n)_x$ for the stabilizer of x in the group $P(n)$:

$$P(n)_x = \{P \in P(n) : Px = x\}\ .$$

The next result is [16, Lemma 2.2].

Lemma 4. *For v in $\mathbf{R}^n$ and x in $\mathbf{R}^n_{\geq}$, if the stabilizer $P(n)_x$ is a subgroup of $P(n)_v$ then $v^T\lambda$ is differentiable at* $\mathrm{Diag}\, x$ *with* $(v^T\lambda)'(\mathrm{Diag}\, x) = \mathrm{Diag}\, v$.

I can now derive the main tool of this section, which gives directional derivative information about the eigenvalue map λ. The adjoint of the map $\mathrm{Diag} : \mathbf{R}^n \to S(n)$ is the map $\mathrm{diag} : S(n) \to \mathbf{R}^n$ taking a matrix Z to a vector with components its diagonal entries.

Theorem 4 (Eigenvalue derivatives). *Any vector x in $\mathbf{R}^n_{\geq}$ and matrix Z in $S(n)$ satisfy*

$$\mathrm{diag}\, Z \in \mathrm{conv}\,(P(n)_x \lambda'(\mathrm{Diag}\, x;\, Z)). \tag{9}$$

Proof. Partition the set $\{1, 2, \ldots, n\}$ into consecutive blocks of integers $I_1, I_2, \ldots, I_k$, so that $x_i = x_j$ if and only if the indices i and j belong to the same block. Correspondingly, write any vector y in $\mathbf{R}^n$ in the form

$$y = \bigoplus_{r=1}^{k} y^r, \quad \text{where } y^r \in \mathbf{R}^{|I_r|} \text{ for each } r\ .$$

Notice also that the stabilizer $P(n)_x$ consists of matrices of permutations fixing each block I_r.

Now suppose relation (9) fails. From the existence of a separating hyperplane, some vector y in $\mathbf{R}^n$ satisfies

$$y^T \mathrm{diag}\, Z > y^T P\lambda'(\mathrm{Diag}\, x;\, Z) \ \text{ for all } P \text{ in } P(n)_x. \tag{10}$$

Let $\hat{y}$ denote the vector $\oplus_r \overline{y^r}$. There is a vector v in $\mathbf{R}^n$ with equal components within every block I_r (or in other words with $P(n)_x$ a subgroup of $P(n)_v$) so that $v + \hat{y}$ lies in $\mathbf{R}^n_{\geq}$. Lemma 3 shows

$$\mathrm{Diag}\,(v + \hat{y}) \in \partial((v + \hat{y})^T\lambda)(\mathrm{Diag}\, x)\ ,$$

or equivalently, any matrix T in $S(n)$ satisfies

$$\mathrm{tr}\,(T(\mathrm{Diag}\,(v + \hat{y}))) \leq (v + \hat{y})^T\lambda'(\mathrm{Diag}\, x;\, T). \tag{11}$$

On the other hand, Lemma 4 shows

$$\mathrm{tr}\,(T(\mathrm{Diag}\, v)) = v^T\lambda'(\mathrm{Diag}\, x;\, T). \tag{12}$$

Subtracting equation (12) from inequality (11) gives

$$\mathrm{tr}\,(T(\mathrm{Diag}\, \hat{y})) \leq \hat{y}^T\lambda'(\mathrm{Diag}\, x;\, T). \tag{13}$$

Writing $\mathrm{diag}\, Z = z = \oplus_r z^r$, there is a matrix Q in $P(n)_x$ satisfying

$$\mathrm{diag}\,(Q.Z) = \bigoplus_r \overline{z^r}\ .$$

Now choosing $T = Q.Z$ in inequality (13) and using Lemma 2 repeatedly shows

$$\begin{aligned} y^T z &\le \left(\bigoplus_r \overline{y^r}\right)^T \left(\bigoplus_r \overline{z^r}\right) \\ &= \mathrm{tr}\,((Q.Z)\mathrm{Diag}\,\hat{y}) \\ &\le \hat{y}^T \lambda'(\mathrm{Diag}\,x;\, Q.Z) \\ &= \hat{y}^T \lambda'(\mathrm{Diag}\,x;\, Z), \end{aligned}$$

using the Subgradient Invariance Proposition (2). But now choosing the matrix P in inequality (10) so that $P^T y = \hat{y}$ gives a contradiction.

□

The directional derivative of the eigenvalue map λ gives a first order estimate which is uniform in the direction of perturbation. The next result makes this precise.

Lemma 5. *Given a matrix X in $S(n)$, small matrices Z in $S(n)$ satisfy*

$$\lambda(X + Z) = \lambda(X) + \lambda'(X; Z) + o(Z)\ .$$

Proof. This relationship relies only on the Lipschitzness and directional differentiability of the map λ (see for example [8, Lem 2.1.1]). These two properties follow easily from writing each component function λ_i as the difference of the two finite convex functions, $\sum_{j=1}^{i} \lambda_j$ and $\sum_{j=1}^{i-1} \lambda_j$ (see Lemma 3).

□

I can now prove the implication (8).

Theorem 5. *For any vectors x in $\mathbf{R}^n_{\ge}$ and y in $\mathbf{R}^n$, and any eigenvalue function $f \circ \lambda$,*

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

Proof. By the Subgradient Invariance Proposition (2), every element of the finite set $P(n)_x y$ is a regular subgradient of f at x. The convex hull of this set, which I denote Λ, has support function given by

$$\delta^*_\Lambda(z) = \max\{z^T P y : P \in P(n)_x\}, \quad \text{for all } z \text{ in } \mathbf{R}^n\ .$$

This function is sublinear, with global Lipschitz constant $\|y\|$.

Fix a real $\epsilon > 0$. The definition of regular subgradients implies, for small vectors z in $\mathbf{R}^n$,

$$f(x + z) \ge f(x) + \delta^*_\Lambda(z) - \epsilon\|z\|. \tag{14}$$

On the other hand, using the previous lemma (5), small matrices Z in $S(n)$ must satisfy

$$\|\lambda(\mathrm{Diag}\,x + Z) - x - \lambda'(\mathrm{Diag}\,x;\, Z)\| \le \epsilon\|Z\|\ ,$$

and hence, by inequality (14),

$$
\begin{aligned}
&f(\lambda(\operatorname{Diag} x + Z)) \\
&= f(x + (\lambda(\operatorname{Diag} x + Z) - x)) \\
&\geq f(x) - \epsilon\|\lambda(\operatorname{Diag} x + Z) - x\| \\
&\quad + \delta_\Lambda^*(\lambda'(\operatorname{Diag} x;\, Z) + [\lambda(\operatorname{Diag} x + Z) - x - \lambda'(\operatorname{Diag} x;\, Z)]) \\
&\geq f(x) + \delta_\Lambda^*(\lambda'(\operatorname{Diag} x;\, Z)) - (1 + \|y\|)\epsilon\|Z\|,
\end{aligned}
$$

using the Lipschitz property of λ.

The Eigenvalue Derivatives Theorem (4) states

$$
\operatorname{diag} Z \in \operatorname{conv}(P(n)_x \lambda'(\operatorname{Diag} x;\, Z)). \tag{15}
$$

Since the polytope Λ is obviously invariant under the group $P(n)_x$, so is its support function, whence

$$
\delta_\Lambda^*(P\lambda'(\operatorname{Diag} x;\, Z)) = \delta_\Lambda^*(\lambda'(\operatorname{Diag} x;\, Z)) \ ,
$$

for any matrix P in $P(n)_x$. This, combined with the convexity of δ_Λ^* and relation (15), demonstrates

$$
\delta_\Lambda^*(\operatorname{diag} Z) \leq \delta_\Lambda^*(\lambda'(\operatorname{Diag} x;\, Z)) \ .
$$

So the argument above shows

$$
\begin{aligned}
f(\lambda(\operatorname{Diag} x + Z)) &\geq f(x) + \delta_\Lambda^*(\operatorname{diag} Z) - (1 + \|y\|)\epsilon\|Z\| \\
&\geq f(x) + y^T \operatorname{diag} Z - (1 + \|y\|)\epsilon\|Z\| \\
&= f(x) + \langle \operatorname{Diag} y, Z\rangle - (1 + \|y\|)\epsilon\|Z\|,
\end{aligned}
$$

and since ϵ was arbitrary, the result follows.

□

Corollary 1 (Diagonal subgradients). *For any vectors x and y in $\mathbf{R}^n$ and any eigenvalue function $f \circ \lambda$,*

$$
y \in \partial f(x) \quad \Leftrightarrow \quad \operatorname{Diag} y \in \partial(f \circ \lambda)(\operatorname{Diag} x) \ .
$$

Corresponding results hold for regular and horizon subgradients. If f is Lipschitz around $\lambda(X)$ then the implication '$\Rightarrow$' also holds for for Clarke subgradients.

Proof. By virtue of Proposition 5, I need only prove the implications '$\Rightarrow$'. Suppose first that y is a regular subgradient. Fixing a matrix P in $P(n)$ satisfying $\bar{x} = Px$, the assumption $y \in \hat{\partial} f(x)$ implies $Py \in \hat{\partial} f(Px)$, by the Subgradient Invariance Proposition (2). Hence the previous result shows

$$
P^T.\operatorname{Diag} y = \operatorname{Diag}(Py) \in \hat{\partial}(f \circ \lambda)(\operatorname{Diag} Px) = \hat{\partial}(f \circ \lambda)(P^T.\operatorname{Diag} x) \ ,
$$

and the result follows by applying the Subgradient Invariance Proposition again.

Now suppose $y \in \partial f(x)$, so there is a sequence of vectors x^r in $\mathbf{R}^n$ approaching x, with $f(x^r)$ approaching $f(x)$, and a sequence of regular subgradients $y^r \in \hat{\partial} f(x^r)$ approaching y. Hence $\mathrm{Diag}\, x^r$ approaches $\mathrm{Diag}\, x$ with $f(\lambda(\mathrm{Diag}\, x^r))$ approaching $f(\lambda(\mathrm{Diag}\, x))$, and by the above argument, each matrix $\mathrm{Diag}\, y^r$ is a regular subgradient of $f \circ \lambda$ at $\mathrm{Diag}\, x^r$. Since $\mathrm{Diag}\, y^r$ approaches $\mathrm{Diag}\, y$, the result follows. The horizon subgradient case is amost identical.

If the function f is Lipschitz around $\lambda(X)$ and y is a Clarke subgradient at x, then y is a convex combination of subgradients $y^i \in \partial f(x)$. Since, by the above argument, each matrix $\mathrm{Diag}\, y^i$ is a subgradient of $f \circ \lambda$ at X, and $\mathrm{Diag}\, y$ is a convex combination of these matrices, $\mathrm{Diag}\, y$ must be a Clarke subgradient.

□

Note 1. I prove the converse implication '⇐' in the Clarke case in §8.

6. The main result

My main result, characterizing the subgradients of an arbitrary eigenvalue function, combines the diagonal case developed in the previous section with the diagonal reduction argument of §4. Recall for matrices X in $S(n)$ and U in $O(n)$ the notation $U.X = U^T XU$.

Theorem 6 (Subgradients). *The (approximate) subdifferential of any eigenvalue function $f \circ \lambda$ at a matrix X in $S(n)$ is given by the formula*

$$\partial(f \circ \lambda)(X) = O(n)^X.\mathrm{Diag}\,(\partial f(\lambda(X))), \tag{16}$$

where

$$O(n)^X = \{U \in O(n) : U.\mathrm{Diag}\,\lambda(X) = X\}\ .$$

The sets of regular and horizon subgradients satisfy corresponding formulae.

Proof. For any vector y in $\partial f(\lambda(X))$, the Diagonal Subgradients Corollary (1) shows

$$\mathrm{Diag}\, y \in \partial(f \circ \lambda)(\mathrm{Diag}\,\lambda(X))\ ,$$

and now, for any matrix U in $O(n)^X$, the Subgradient Invariance Proposition (2) implies

$$U.\mathrm{Diag}\, y \in \partial(f \circ \lambda)(U.\mathrm{Diag}\,\lambda(X)) = \partial(f \circ \lambda)(X)\ ,$$

as required.

On the other hand, any subgradient Y in $\partial(f \circ \lambda)(X)$ commutes with X, by the Commutativity Theorem (3). Hence X and Y diagonalize simultaneously: there is a matrix U in $O(n)^X$ and a vector y in $\mathbf{R}^n$ with $Y = U.\mathrm{Diag}\, y$. The Subgradient Invariance Proposition shows

$$\mathrm{Diag}\, y \in \partial(f \circ \lambda)(\mathrm{Diag}\,\lambda(X))\ ,$$

whence $y \in \partial f(\lambda(X))$, by the Diagonal Subgradients Corollary. Thus the matrix Y belongs to the right-hand-side of equation (16), as required. The arguments for regular and horizon subgradients are completely analogous.

□

Note 2. The Clarke subdifferential satisfies the corresponding formula in the locally Lipschitz case — see §8.

Corollary 2 (Unique regular subgradients). *An eigenvalue function $f \circ \lambda$ has a unique regular subgradient at a matrix X in $S(n)$ if and only if f has a unique regular subgradient at $\lambda(X)$.*

Proof. Suppose f has a unique regular subgradient y at $\lambda(X)$. Then the subdifferential formula (16) shows that every matrix in the nonempty convex set $\hat{\partial}(f \circ \lambda)(X)$ has the same norm, namely $\|y\|$, and hence this set is a singleton. The converse is immediate. □

The same proof, in the Lipschitz case, works for Clarke subgradients, which shows that $f \circ \lambda$ is strictly differentiable at X if and only if f is strictly differentiable at $\lambda(X)$ (see [16]). Another proof of this fact follows later in the section. A more direct proof of the following result appears in [16].

Corollary 3 (Fréchet differentiability). *An eigenvalue function $f \circ \lambda$ is Fréchet differentiable at a matrix X in $S(n)$ if and only if f is Fréchet differentiable at $\lambda(X)$.*

Proof. This follows immediately from the preceding result, since a function h is Fréchet differentiable at a point if and only if both h and $-h$ have unique regular subgradients there. □

Corollary 4 (Regularity). *Suppose the permutation-invariant function f is finite at $\lambda(X)$ (for a matrix X in $S(n)$), and has at least one subgradient there. Then the eigenvalue function $f \circ \lambda$ is (Clarke) regular at X if and only if f is regular at $\lambda(X)$.*

Proof. Since f has a subgradient at $\lambda(X)$, the Subgradients Theorem (6) shows $f \circ \lambda$ has a subgradient at X. Furthermore, $f \circ \lambda$ is lower semicontinuous near X if and only if f is lower semicontinuous near $\lambda(X)$.

By definition, f is regular at $\lambda(X)$ if and only if it is lower semicontinuous near $\lambda(X)$ and the conditions

$$\partial f(\lambda(X)) = \hat{\partial} f(\lambda(X)) \neq \emptyset, \quad \text{and} \tag{17}$$

$$(\hat{\partial} f(\lambda(X)))^{\infty} = \partial^{\infty} f(\lambda(X)) \tag{18}$$

hold, whereas $f \circ \lambda$ is regular at X if and only if it is lower semicontinuous near X and the conditions

$$\partial(f \circ \lambda)(X) = \hat{\partial}(f \circ \lambda)(X) \neq \emptyset, \quad \text{and} \tag{19}$$

$$(\hat{\partial}(f \circ \lambda)(X))^{\infty} = \partial^{\infty}(f \circ \lambda)(X), \tag{20}$$

hold. By formula (16) and its regular analogue, condition (17) implies condition (19). Conversely, condition (19) is equivalent to

$$\partial(f \circ \lambda)(\operatorname{Diag} \lambda(X)) = \hat{\partial}(f \circ \lambda)(\operatorname{Diag} \lambda(X)) \neq \emptyset\ ,$$

by the Subgradient Invariance Proposition (2), and condition (17) follows by the Diagonal Subgradients Corollary (1).

Applying the Recession Lemma (1) to the regular version of formula (16), noting that the set of regular subgradients is always closed and convex, shows

$$(\hat{\partial}(f \circ \lambda)(X)))^{\infty} = O(n)^X.[\mathrm{Diag}\,\hat{\partial} f(\lambda(X))]^{\infty} = O(n)^X.\mathrm{Diag}\,[(\hat{\partial} f(\lambda(X)))^{\infty}]\ .$$

Hence condition (18) implies condition (20), by the horizon version of formula (16).

On the other hand, condition (20) is equivalent to

$$(\hat{\partial}(f \circ \lambda)(\mathrm{Diag}\,\lambda(X)))^{\infty} = \partial^{\infty}(f \circ \lambda)(\mathrm{Diag}\,\lambda(X))\ ,$$

by the Subgradient Invariance Proposition, whence

$$\begin{aligned}\mathrm{Diag}\,((\hat{\partial} f(\lambda(X)))^{\infty}) &= (\mathrm{Diag}\,\hat{\partial} f(\lambda(X)))^{\infty}\\ &= (\hat{\partial}(f \circ \lambda)(\mathrm{Diag}\,\lambda(X)) \cap \mathrm{Diag}\,\mathbf{R}^n)^{\infty}\\ &= (\hat{\partial}(f \circ \lambda)(\mathrm{Diag}\,\lambda(X)))^{\infty} \cap \mathrm{Diag}\,\mathbf{R}^n\\ &= \partial^{\infty}(f \circ \lambda)(\mathrm{Diag}\,\lambda(X)) \cap \mathrm{Diag}\,\mathbf{R}^n\\ &= \mathrm{Diag}\,(\partial^{\infty} f(\lambda(X))),\end{aligned}$$

using the Diagonal Subgradients Corollary again. Condition (18) follows.

□

Corollary 5 (Strict differentiability). *An eigenvalue function $f \circ \lambda$ is strictly differentiable at a matrix X in $S(n)$ if and only if the function f is strictly differentiable at $\lambda(X)$.*

Proof. Strict differentiability of f at $\lambda(X)$ is equivalent, by [25, Thm 9.18], to continuity in a neighbourhood and regularity of both f and $-f$ at $\lambda(X)$. The result therefore follows by the Regularity Corollary (4).

□

The Subgradients Theorem (6) is more attractive (although perhaps less practical) when written in a graphical form. The graph of the subdifferential is the set $\mathrm{Graph}\,\partial f = \{(x, y) \in (\mathbf{R}^n)^2 : y \in \partial f(x)\}$. Define a binary operation $* : O(n) \times (\mathbf{R}^n)^2 \to (S(n))^2$ by

$$U * (x, y) = (U.\mathrm{Diag}\,x, U.\mathrm{Diag}\,y)\ .$$

Corollary 6 (Subdifferential graphs). *The graph of the subdifferential of an eigenvalue function $f \circ \lambda$ is given by the formula*

$$\mathrm{Graph}\,\partial(f \circ \lambda) = O(n) * \mathrm{Graph}\,\partial f\ .$$

Analogous formulae hold for the subdifferentials $\hat{\partial}$, ∂^{∞} and (in the locally Lipschitz case) ∂^c.

Proof. A pair of matrices (X, Y) lies in Graph $\partial(f \circ \lambda)$ exactly when Y belongs to $\partial(f \circ \lambda)(X)$, and in this case, by the Subgradients Theorem, there is a vector y in $\partial f(\lambda(X))$ and a matrix U in $O(n)^X$ satisfying $Y = U.\mathrm{Diag}\, y$. Hence $(X, Y) = U * (\lambda(X), y)$.

Conversely, for a pair of vectors (x, y) in Graph ∂f and a matrix U in $O(n)$, y lies in $\partial f(x)$, whence $\mathrm{Diag}\, y \in \partial(f \circ \lambda)(\mathrm{Diag}\, x)$, by the Diagonal Subgradients Corollary (1). The Subgradient Invariance Proposition now implies $U.\mathrm{Diag}\, y \in \partial(f \circ \lambda)(U.\mathrm{Diag}\, x)$, or in other words $U * (x, y) \in$ Graph $\partial(f \circ \lambda)$. The arguments for the other subdifferentials are exactly analogous.

□

An eigenvalue function $f \circ \lambda$ is convex if and only if the function f is convex (see [17]). In this case the regular subgradients are exactly the subgradients in the usual sense of convex analysis. The following result (cf. [17]) is easy to deduce from the regular case of the Subgradients Theorem.

Corollary 7 (Convex subgradients). *Consider an eigenvalue function $f \circ \lambda$, where the function f is convex. A matrix Y in $S(n)$ is a (convex) subgradient of $f \circ \lambda$ at a matrix X in $S(n)$ if and only if X and Y have a simultaneous ordered spectral decomposition and $\lambda(Y)$ is a (convex) subgradient of f at $\lambda(X)$.*

7. Invariance under Young subgroups

The Subgradients Theorem (6) provides one approach to the calculation of Clarke subgradients. To pursue this approach, in this section I first derive a subsidiary result about an important class of convex sets.

A subset C of the Euclidean space E is *invariant* under a subgroup G of $O(E)$ if $gC = C$ for all transformations g in G. If the function $f : \mathbf{R}^n \to [-\infty, +\infty]$ is permutation-invariant then the regular subdifferential of f at a point x in $\mathbf{R}^n$ is a convex set, invariant under the stabilizer $P(n)_x$ (by the Subgradient Invariance Proposition (2)). Subgroups of $P(n)$ of the form $P(n)_x$ are called *Young subgroups*: they are those subgroups corresponding to permutations preserving given partitions of the set $\{1, 2, \ldots, n\}$. In this section I prove a general property of convex sets invariant under Young subgroups.

Given an equivalence relation $\sim$ on the set $\{1, 2, \ldots, n\}$, define a subspace

$$\mathbf{R}^n_\sim = \{x \in \mathbf{R}^n : x_i = x_j \text{ whenever } i \sim j\},$$

and two groups

$$P(n)_\sim = \{P \in P(n) : Px = x \text{ for all } x \in \mathbf{R}^n_\sim\},$$
$$O(n)_\sim = \{U \in O(n) : U.\mathrm{Diag}\, x = \mathrm{Diag}\, x \text{ for all } x \in \mathbf{R}^n_\sim\}.$$

Thus $P(n)_\sim$ consists of those matrices of permutations leaving invariant the equivalence classes of $\sim$.

I can always reorder the basis so that $\sim$ has equivalence classes consecutive blocks of integers $I_1, I_2, \ldots, I_k$: then I write any vector y in $\mathbf{R}^n$ in the form

$$y = \bigoplus_{r=1}^{k} y^r, \quad \text{where } y^r \in \mathbf{R}^{|I_r|} \text{ for each } r\ ,$$

and for matrices U^r in $M(|I_r|)$ for each r, I write $\mathrm{Diag}\,(U^r)$ for the block diagonal matrix

$$\begin{pmatrix} U^1 & 0 & \ldots & 0 \\ 0 & U^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & U^k \end{pmatrix} \ .$$

Then it is easy to calculate the notions above:

$$\begin{aligned} \mathbf{R}^n_\sim &= \oplus_r \mathbf{R}(1, 1, \ldots, 1)^T, \\ P(n)_\sim &= \{\mathrm{Diag}\,(P^r) : P^r \in P(|I_r|) \ \text{ for each } r\}, \quad \text{and} \\ O(n)_\sim &= \{\mathrm{Diag}\,(U^r) : U^r \in O(|I_r|) \ \text{ for each } r\}. \end{aligned}$$

Theorem 7 (Invariant sets). *If $\sim$ is an equivalence relation on the set $\{1, 2, \ldots, n\}$ and the convex set $C \subset \mathbf{R}^n$ is invariant under the group $P(n)_\sim$, then the set of matrices $O(n)_\sim.\mathrm{Diag}\, C$ is convex.*

Proof. Reorder the basis as above. Diagonalizing each diagonal block easily shows the identity

$$O(n)_\sim.\mathrm{Diag}\, C = \{\mathrm{Diag}\,(X^r) : \oplus_r \lambda(X^r) \in C\}. \tag{21}$$

For two matrices $X = \mathrm{Diag}\,(X^r)$ and $Y = \mathrm{Diag}\,(Y^r)$ in this set, and a number α in $[0, 1]$, I wish to show

$$\alpha X + (1-\alpha)Y \in O(n)_\sim.\mathrm{Diag}\, C\ ,$$

or equivalently, by identity (21),

$$\bigoplus_r \lambda(\alpha X^r + (1-\alpha)Y^r) \in C\ .$$

Since (21) shows $\oplus_r \lambda(X^r)$ and $\oplus_r \lambda(Y^r)$ both lie in the convex set C, which is invariant under the group $P(n)_\sim$, it suffices to show

$$\oplus_r \lambda(\alpha X^r + (1-\alpha)Y^r) \in \mathrm{conv}\,(P(n)_\sim\{\oplus_r \lambda(X^r), \ \ \oplus_r \lambda(Y^r)\})\ .$$

If this fails then there is a separating hyperplane: there exists a vector $z = \oplus_r z^r$ in $\mathbf{R}^n$ satisfying

$$\langle z, \oplus_r \lambda(\alpha X^r + (1-\alpha)Y^r)\rangle > \max \langle z, P(n)_\sim\{\oplus_r \lambda(X^r), \ \ \oplus_r \lambda(Y^r)\}\rangle\ .$$

But then Lemmas 2 and 3 show

$$\begin{aligned}
\sum_r \langle z^r, \lambda(\alpha X^r + (1-\alpha)Y^r)\rangle &> \max\left\{\sum_r \langle \overline{z^r}, \lambda(X^r)\rangle\ ,\ \ \sum_r \langle \overline{z^r}, \lambda(Y^r)\rangle\right\}\\
&\geq \alpha \sum_r \langle \overline{z^r}, \lambda(X^r)\rangle + (1-\alpha)\sum_r \langle \overline{z^r}, \lambda(Y^r)\rangle\\
&= \sum_r \langle \overline{z^r}, \alpha\lambda(X^r) + (1-\alpha)\lambda(Y^r)\rangle\\
&\geq \sum_r \langle \overline{z^r}, \lambda(\alpha X^r + (1-\alpha)Y^r)\rangle\\
&\geq \sum_r \langle z^r, \lambda(\alpha X^r + (1-\alpha)Y^r)\rangle,
\end{aligned}$$

which is a contradiction.

□

8. Clarke subgradients

The Subgradients Theorem (6) did not cover the case of the Clarke subdifferential because of the missing converse in the Diagonal Subgradients Corollary. This section remedies this using the Invariant Sets Theorem (7). The main result, which follows, first appeared in [16], via a different approach.

Theorem 8 (Clarke subgradients). *The Clarke subdifferential of a locally Lipschitz eigenvalue function $f \circ \lambda$ at a matrix X in $S(n)$ is given by the formula*

$$\partial^c(f \circ \lambda)(X) = O(n)^X.\mathrm{Diag}\,(\partial^c f(\lambda(X))). \tag{22}$$

Proof. Assume first $X = \mathrm{Diag}\, x$ for a vector x in $\mathbf{R}^n_\geq$: the general case follows easily, using the Subgradient Invariance Proposition (2). Define an equivalence relation $\sim$ on $\{1, 2, \ldots, n\}$ by

$$i \sim j \quad \Leftrightarrow \quad x_i = x_j\ .$$

Then it is easy to verify $O(n)^{\mathrm{Diag}\, x} = O(n)_\sim$, and since x is invariant under the group $P(n)_\sim$, the convex set $\partial^c f(x)$ is also invariant under $P(n)_\sim$ (by the Subgradient Invariance Proposition). The Invariant Set Theorem now shows that the set $O(n)_\sim.\mathrm{Diag}\,\partial^c f(x)$ is convex.

The Subgradients Theorem (6) demonstrates

$$\partial^c(f \circ \lambda)(\mathrm{Diag}\, x) = \mathrm{conv}\,\partial(f \circ \lambda)(\mathrm{Diag}\, x) = \mathrm{conv}\,(O(n)_\sim.\mathrm{Diag}\,\partial f(x))\ .$$

The inclusion

$$O(n)_\sim.\mathrm{Diag}\,\partial f(x) \subset O(n)_\sim.\mathrm{Diag}\,\partial^c f(x)$$

is trivial, so the convexity of the right-hand-side shows

$$\mathrm{conv}\,(O(n)_{\sim}.\mathrm{Diag}\,\partial f(x)) \subset O(n)_{\sim}.\mathrm{Diag}\,\partial^c f(x)\ .$$

But since $\partial^c f(x) = \mathrm{conv}\,\partial f(x)$, the reverse inclusion is immediate, and the result follows.

□

The Diagonal Subgradients Corollary is interesting in its own right, so this section concludes with the Clarke version.

Corollary 8 (Diagonal Clarke subgradients). *For any vectors x and y in $\mathbf{R}^n$ and any eigenvalue function $f \circ \lambda$,*

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

Proof. The Diagonal Subgradients Corollary (1) shows the implication '$\Rightarrow$'. For the converse, assuming $\mathrm{Diag}\,y \in \partial^c(f \circ \lambda)(\mathrm{Diag}\,x)$, the Clarke Subgradients Theorem above shows the existence of a matrix U in $O(n)$ and a vector z in $\partial^c f(\bar{x})$ with $\mathrm{Diag}\,y = U.\mathrm{Diag}\,z$ and $\mathrm{Diag}\,x = U.\mathrm{Diag}\,\bar{x}$. By the Simultaneous Conjugacy Proposition (3), there is a matrix P in $P(n)$ with $y = Pz$ and $x = P\bar{x}$, and the result now follows from the Subgradient Invariance Proposition (2).

□

9. Order statistics and individual eigenvalues

This article ends with an example illustrating the main result. Specifically, I apply the Subgradients Theorem (6) to calculate the approximate subdifferential of the individual eigenvalue $\lambda_k(\cdot)$. In part my aim is to provide a useful tool, while on the other hand giving an elegant and nontrivial illustration of the approximate subdifferential.

The restriction of the function λ_k to the diagonal matrices is the k'th *order statistic* $\phi_k : \mathbf{R}^n \to \mathbf{R}$ (applied to the vector of diagonal entries). More precisely, this (permutation-invariant) function

$$\phi_k(x) = k^{\text{th}} \text{ largest element of } \{x_1, x_2, \ldots, x_n\}$$

(or in other words $\phi_k(x) = (\bar{x})_k$), satisfies the relation $\phi_k(x) = \lambda_k(\mathrm{Diag}\,x)$. To apply the Subgradients Theorem, note $\lambda_k = \phi_k \circ \lambda$: thus I must first compute the subdifferential of ϕ_k. I denote the canonical basis in $\mathbf{R}^n$ by $\{e^1, e^2, \ldots, e^n\}$.

Proposition 6. *At any point x in $\mathbf{R}^n$, the regular subgradients of the k'th order statistic are described by*

$$\hat{\partial}\phi_k(x) = \begin{cases} \mathrm{conv}\,\{e^i : x_i = \phi_k(x)\}, & \textit{if } k = 1 \textit{ or } \phi_{k-1}(x) > \phi_k(x), \\ \emptyset, & \textit{otherwise}, \end{cases} ,$$

and $\partial^\infty \phi_k(x) = \{0\}$.

Proof. Define a set of indices $I = \{i : x_i = \phi_k(x)\}$. If the inequality $\phi_{k-1}(x) > \phi_k(x)$ holds then clearly, close to the point x, the function ϕ_k is given by $w \in \mathbf{R}^n \mapsto \max_{i \in I} w_i$. The subdifferential at x of this second function (which is convex) is just $\text{conv}\,\{e^i : i \in I\}$.

On the other hand, in the case $\phi_{k-1}(x) = \phi_k(x)$, suppose y is a regular subgradient, and so satisfies

$$\phi_k(x+z) \geq \phi_k(x) + y^T z + o(z), \quad \text{as } z \to 0 \ .$$

For any index i in I, all small positive δ satisfy $\phi_k(x + \delta e^i) = \phi_k(x)$, from which I deduce $y_i \leq 0$, but also

$$\phi_k\left(x - \delta \sum_{i \in I} e^i\right) = \phi_k(x) - \delta \ ,$$

which implies the contradiction $\sum_{i \in I} y_i \geq 1$. The horizon subdifferential is easy to check.

□

To simplify notation, for general subgradients I concentrate first on the case $x = 0$. For a vector y in $\mathbf{R}^n$ I write

$$\text{supp}\, y = \{i : y_i \neq 0\};$$

the number of elements in this set is then $|\text{supp}\, y|$. The *simplex* in $\mathbf{R}^n$ is the set described by

$$\Delta = \left\{ y \in \mathbf{R}^n : y \geq 0, \ \sum y_i = 1 \right\} = \text{conv}\,\{e^1, e^2, \ldots, e^n\} \ .$$

Proposition 7. *The Clarke subdifferential of the k'th order statistic ϕ_k at the origin is just the simplex Δ, whereas the (approximate) subdifferential is given by*

$$\partial \phi_k(0) = \{y \in \Delta : |\text{supp}\, y| \leq n - k + 1\} \ .$$

Regularity holds if and only if $k = 1$.

Proof. Since any regular subgradient of ϕ_k at any point always lies in the right-hand-side set (by the previous proposition), so does an arbitrary subgradient, since this set is closed. Conversely, given any vector y belonging to the right-hand-side, choose a subset J of exactly $(k-1)$ indices i for which y_i is zero. The previous proposition shows, for any small $\delta > 0$,

$$y \in \text{conv}\,\{e^i : i \notin J\} = \hat{\partial} \phi_k\left(\delta \sum_{i \in J} e^i\right) \ ,$$

whence, by taking limits, $y \in \partial \phi_k(0)$, as I claimed. The Clarke case follows by taking convex hulls, and the regularity claim is immediate, by the previous proposition.

□

I denote the cone of positive semidefinite matrices in $S(n)$ by $S(n)^+$. The next result follows immediately from the previous proposition by applying the Subgradients Theorem (6).

Corollary 9. *The Clarke subdifferential of the k'th eigenvalue at the origin is given by*

$$\partial^c \lambda_k(0) = \{Y \in S(n)^+ : \operatorname{tr} Y = 1\} ,$$

whereas the (approximate) subdifferential is given by

$$\partial \lambda_k(0) = \{Y \in S(n)^+ : \operatorname{tr} Y = 1,\ \operatorname{rank} Y \leq n - k + 1\} .$$

Regularity holds if and only if $k = 1$.

The arguments for the general cases are completely analogous, leading to the following results.

Theorem 9 (k'th order statistic). *The Clarke subdifferential of the k'th order statistic ϕ_k at a point x in $\mathbf{R}^n$ is given by*

$$\partial^c \phi_k(x) = \operatorname{conv}\{e^i : x_i = \phi_k(x)\} ,$$

whereas the (approximate) subdifferential is given by

$$\begin{aligned} \partial \phi_k(x) &= \{y \in \partial^c \phi_k(x) : |\operatorname{supp} y| \leq \alpha\}, \quad \textit{where} \\ \alpha &= 1 - k + |\{i : x_i \geq \phi_k(x)\}|. \end{aligned}$$

Regularity holds if and only if $\phi_{k-1}(x) > \phi_k(x)$.

Corollary 10 (Eigenvalue subgradients). *The Clarke subdifferential of the k'th eigenvalue at a matrix X in $S(n)$ is given by*

$$\partial^c \lambda_k(X) = \operatorname{conv}\{uu^T : u \in \mathbf{R}^n,\ \|u\| = 1,\ Xu = \lambda_k(X)u\}, \tag{23}$$

whereas the (approximate) subdifferential is given by

$$\begin{aligned} \partial \lambda_k(X) &= \{Y \in \partial^c \phi_k(X) : \operatorname{rank} Y \leq \alpha\}, \quad \textit{where} \\ \alpha &= 1 - k + |\{i : \lambda_i(X) \geq \lambda_k(X)\}|. \end{aligned}$$

Regularity holds if and only if $\lambda_{k-1}(X) > \lambda_k(X)$.

Formula (23) was first observed in [4] (see also [16]), although a weaker version appears in [6].

I conclude with a nice application of formula (23) to a well-known eigenvalue isotonicity result (see [10], p. 181). For matrices X and Y in $S(n)$, I write $X \succeq Y$ or $X \succ Y$ to mean $X - Y$ is positive semidefinite or definite respectively.

Corollary 11 (Eigenvalue isotonicity). *For any matrices X and Y in $S(n)$,*

$$\begin{aligned} X \succeq Y &\Rightarrow \lambda(X) \geq \lambda(Y), \quad \textit{and} \\ X \succ Y &\Rightarrow \lambda(X) > \lambda(Y) . \end{aligned}$$

Proof. For any index k, by the Lebourg Mean Value Theorem [3], Thm 2.3.7, there is a matrix Z in $S(n)$ such that

$$\lambda_k(X) - \lambda_k(Y) \in \langle \partial^c \lambda_k(Z), X - Y \rangle .$$

The result now follows easily from formula (23).

□

Acknowledgements. To Michael Overton for many helpful suggestions, and to Gabor Pataki for discussions leading to Theorem 7 (an analogous result to which appears independently in his Carnegie-Mellon doctoral thesis (1997)).

References

1. Arnold, V.I. (1988): Geometrical Methods in the Theory of Ordinary Differential Equations, second edition. Springer, New York
2. Boothby, W.M. (1988): An Introduction to Differential Manifolds and Riemannian Geometry, second edition. Academic Press, Orlando
3. Clarke, F.H. (1983): Optimization and Nonsmooth Analysis. Wiley, New York
4. Cox, S.J., Overton, M. (1994): Private communication
5. Fletcher, R. (1985): Semi-definite matrix constraints in optimization. SIAM J. Control Optim. **23**, 493–513
6. Gollan, B. (1987): Eigenvalue perturbation and nonlinear parametric optimization. Math. Program. Study **30**, 67–81
7. Helgason, S. (1978): Differential Geometry, Lie Groups, and Symmetric Spaces. Academic Press, New York
8. Hiriart-Urruty, J.-B., Lemaréchal, C. (1993): Convex Analysis and Minimization Algorithms I. Springer, Berlin
9. Hiriart-Urruty, J.-B., Seeger, A., Ye, D. (1992): Sensitivity Analysis for a Class of Convex Functions Defined over a Space of Symmetric Matrices, Vol. 382 of Lecture Notes in Economics and Mathematical Systems, pp. 133–154. Springer
10. Horn, R.A., Johnson, C. (1985): Matrix Analysis. Cambridge University Press, Cambridge, U.K.
11. Ioffe, A.D. (1981): Sous-différentielles approchées de fonctions numériques. Comtes Rendus de l'Académie des Sciences de Paris **292**, 675–678
12. Ioffe, A.D. (1984): Approximate subdifferentials and applications, I: The finite dimensional theory. Trans. Am. Math. Soc. **281**, 389–416
13. Kato, T. (1982): A Short Introduction to Perturbation Theory for Linear Operators. Springer, New York
14. Kruger, A.Y., Mordukhovich, B.S. (1980): Extremal points and the Euler equation in nonsmooth analysis. Doklady Akademia Nauk BSSR (Belorussian Academy of Sciences) **24**, 684–687
15. Lewis, A.S. (1996): Convex analysis on the Hermitian matrices. SIAM J. Optim. **6**, 164–177
16. Lewis, A.S. (1996): Derivatives of spectral functions. Math. Oper. Res. **6**, 576–588
17. Lewis, A.S. (1996): Group invariance and convex matrix analysis. SIAM J. Matrix Anal. Appl. **17**, 927–949
18. Lewis, A.S. (1998): Eigenvalue-constrained faces. Linear Algebra Appl. **269**, 159–181
19. Lewis, A.S., Overton, M.L. (1996): Eigenvalue optimization. Acta Numerica **5**, 149–190
20. Mordukhovich, B.S. (1976): Maximum principle in the problem of time optimal response with nonsmooth constraints. J. Appl. Math. Mech. **40**, 960–969
21. Von Neumann, J. (1962): Some Matrix Inequalities and Metrization of Matric-space. Tomsk University Rev. **1**, 286–300, 1937. In: Collected Works, Vol. IV, pp. 205–218. Pergamon, Oxford
22. Overton, M.L. (1988): On minimizing the maximum eigenvalue of a symmetric matrix. SIAM J. Matrix Anal. Appl. **9**, 256–268
23. Overton, M.L., Womersley, R.S. (1993): Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. Math. Program. **62**, 321–357
24. Rockafellar, R.T. (1970): Convex Analysis. Princeton University Press, Princeton, N.J.
25. Rockafellar, R.T., Wets, R.J.-B. (1998): Variational Analysis. Springer, New York
26. Theobald, C.M. (1975): An Inequality for the Trace of the Product of Two Symmetric Matrices. Math. Proc. Cambridge Philos. Soc. **77**, 265–266
27. Vandenberghe, L., Boyd, S. (1996): Semidefinite programming. SIAM Rev. **38**, 49–95