SIAM J. MATRIX ANAL. APPL.
Vol. 21, No. 2, pp. 379–381


# LIDSKII'S THEOREM VIA NONSMOOTH ANALYSIS*

A. S. LEWIS†

**Abstract.** Lidskii's theorem on eigenvalue perturbation is proved via a nonsmooth mean value theorem.

**Key words.** Lidskii's theorem, eigenvalue optimization, nonsmooth analysis, Clarke generalized gradient

**AMS subject classifications.** Primary, 90C31, 15A18; Secondary, 49K40, 26B05

**PII.** S0895479898338676

One of the central tools for studying perturbation theory for the eigenvalues of symmetric matrices is Lidskii's theorem. This states that any matrices $Z$ and $Y$ in $\mathbf{S}^n$, the Euclidean space of $n \times n$ real symmetric matrices (with trace inner product), satisfy

$$\lambda(Z+Y) - \lambda(Z) \prec \lambda(Y), \tag{0.1}$$

where $\lambda(Z) \in \mathbf{R}^n$ is the column vector of eigenvalues of $Z$ written by multiplicity and in decreasing order. The symbol $\prec$ denotes *majorization*: $u \prec v$ for vectors $u$ and $v$ in $\mathbf{R}^n$ means

$$u \in \operatorname{conv}\{Pv : P \in \mathbf{P}^n\},$$

where $\mathbf{P}^n$ is the group of $n \times n$ permutation matrices and "conv" denotes convex hull. A standard separation argument (see [6], for example) shows this inclusion is equivalent to the condition

$$x^T u \le \bar{x}^T \bar{v} \quad \text{for all } x \in \mathbf{R}^n,$$

where $\bar{x}$ is the vector in $\mathbf{R}^n$ with the same components as $x$ arranged in decreasing order. Lidskii's theorem is, for example, one of the unifying themes of the recent book [1].

This note approaches Lidskii's theorem via nonsmooth analysis, using the *Clarke generalized gradient* (see [2]). For a real, locally Lipschitz function $g$ defined close to a point $p$ in a Euclidean space, we can define the generalized gradient $\partial g(p)$ as the convex hull of the set of cluster points of gradients of $g$ at points near $p$ in a set of full measure [2, Thm. 2.5.1]. Thus for a smooth function $g$, the generalized gradient coincides with the usual gradient $g'(p)$.

For a smooth function $g$, the classical mean value theorem states that, given points $q$ and $r$, there is a point $p$ on the line segment between $q$ and $r$ satisfying the equation

$$g(q) - g(r) = \langle g'(p), q - r \rangle.$$

*Received by the editors May 13, 1998; accepted for publication (in revised form) by M. Overton December 16, 1998; published electronically October 26, 1999. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada.
http://www.siam.org/journals/simax/21-2/33867.html
†Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1 (aslewis@orion.uwaterloo.ca, http://orion.uwaterloo.ca/∼aslewis).

If $g$ is merely locally Lipschitz, the Lebourg mean value theorem [2, Thm. 2.3.7] generalizes this result, asserting the existence of an element $s$ of $\partial g(p)$ satisfying the equation

$$g(q) - g(r) = \langle s, q - r \rangle.$$

We apply this mean value theorem to derive Lidskii's theorem from a powerful variational result about eigenvalues. This variational result calculates the Clarke generalized gradient of the function $f \circ \lambda$ for an arbitrary locally Lipschitz *permutation-invariant* function $f : \mathbf{R}^n \to \mathbf{R}$ (that is, $f(Px) = f(x)$ for all matrices $P$ in $\mathbf{P}^n$ and vectors $x$ in $\mathbf{R}^n$). Specifically, the result (see [4]) considers diagonalizations of any matrix $X$ in $\mathbf{S}^n$,

$$X = U^T(\operatorname{Diag} x)U, \quad U \in \mathbf{O}^n, \quad x \in \mathbf{R}^n, \tag{0.2}$$

where $\mathbf{O}^n$ is the group of $n \times n$ orthogonal matrices, and states

$$\partial(f \circ \lambda)(X) = \{U^T(\operatorname{Diag} y)U : (0.2) \text{ holds and } y \in \partial f(x)\}. \tag{0.3}$$

The approach of this note is certainly not the simplest proof of Lidskii's theorem, formula (0.3) being a relatively difficult result (see also [5]). However, the approach is delightfully transparent, and it reveals the increasing possibilities of applying nonsmooth analytic techniques in matrix analysis.

Lidskii's theorem (0.1) is equivalent to the inequality

$$w^T(\lambda(Z+Y) - \lambda(Z)) \le \bar{w}^T \lambda(Y) \quad \text{for all } w \in \mathbf{R}^n. \tag{0.4}$$

Fix $w$ and consider the (nonconvex) locally Lipschitz, permutation-invariant function defined by

$$f(x) = w^T \bar{x}.$$

Note the relationship

$$(f \circ \lambda)(X) = w^T \lambda(X) \quad \text{for all } X \in \mathbf{S}^n.$$

Whenever $x$ has distinct components (a subset of $\mathbf{R}^n$ of full measure), $f$ is differentiable at $x$ with gradient $f'(x) = Pw$ for some matrix $P$ in $\mathbf{P}^n$. Hence at *any* point $x$ in $\mathbf{R}^n$ we have the inclusion

$$\partial f(x) \subset \operatorname{conv}\{Pw : P \in \mathbf{P}^n\}, \tag{0.5}$$

by our definition of the generalized gradient. (There are more elementary although less concise ways to see this inclusion.)

By the Lebourg mean value theorem applied to $f \circ \lambda$, there is a matrix $X$ in $\mathbf{S}^n$ (between $Z$ and $Z + Y$) and a matrix $V$ in $\partial(f \circ \lambda)(X)$ satisfying

$$w^T(\lambda(Z+Y) - \lambda(Z)) = \operatorname{tr}(VY) \le \lambda(V)^T \lambda(Y),$$

the last inequality following from von Neumann's trace theorem (see [3, Eq. (7.4.14)]). But now we can apply the generalized gradient formulae (0.3) and (0.5) to deduce

$$\lambda(V) \in \operatorname{conv}\{Pw : P \in \mathbf{P}^n\}$$

(or in other words, $\lambda(V) \prec \bar{w}$). Hence $\lambda(V)^T \lambda(Y) \le \bar{w}^T \lambda(Y)$, and Lidskii's theorem (0.4) follows.

**Acknowledgment.** The author thanks the School of Operations Research and Industrial Engineering at Cornell University, where this work was completed.

## REFERENCES

[1] R. Bhatia, *Matrix Analysis*, Grad. Texts in Math. 169, Springer-Verlag, New York, 1997.
[2] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley, New York, 1983; reprinted as Classics Appl. Math. 5, SIAM, Philadelphia, 1990.
[3] R. A. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[4] A. S. Lewis, *Derivatives of spectral functions*, Math. Oper. Res., 6 (1996), pp. 576–588.
[5] A. S. Lewis, *Nonsmooth analysis of eigenvalues*, Math. Programming, 84 (1999), pp. 1–24.
[6] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Application*, Academic Press, New York, 1979.