PERGAMON

Nonlinear Analysis 42 (2000) 813–820

Nonlinear Analysis

www.elsevier.nl/locate/na

Convex analysis on Cartan subspaces

A.S. Lewis [*,1]

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Received 1 July 1997; accepted 22 May 1998

Keywords: Convexity; Fenchel conjugate; Subgradient; Unitarily invariant norm; Spectral function; Eigenvalue optimization; Semidefinite program; von Neumann's Lemma; Cartan decomposition; Kostant convexity theorem

1. Unitarily invariant norms and convex spectral functions

In 1937, von Neumann [31] gave a famous characterization of *unitarily invariant* matrix norms (that is, norms f on $\mathbf{C}^{p\times q}$ satisfying $f(uxv) = f(x)$ for all unitary matrices u and v and matrices x in $\mathbf{C}^{p\times q}$). His result states that such norms are those functions of the form $g \circ \sigma$, where the map

$$x \in \mathbf{C}^{p\times q} \mapsto \sigma(x) \in \mathbf{R}^p$$

has components the singular values $\sigma_1(x) \geq \sigma_2(x) \geq \cdots \geq \sigma_p(x)$ of x (assuming $p \leq q$) and g is a norm on $\mathbf{R}^p$, invariant under sign changes and permutations of components. Furthermore, he showed the respective dual norms satisfy

$$(g \circ \sigma)^D = g^D \circ \sigma \tag{1.1}$$

(where we regard $\mathbf{C}^{p\times q}$ as a Euclidean space with inner product $\langle x, y\rangle = \operatorname{Re}\operatorname{tr} x^* y$ for matrices x and y in $\mathbf{C}^{p\times q}$).

* Fax: +519-725-5441.

E-mail address: aslewis@orion.uwaterloo.ca (A.S. Lewis)

[1] Research partially supported by the Natural Sciences and Engineering Research Council of Canada.

PII: S0362-546X(99)00126-1

This simple duality relationship helps us calculate the subdifferentials of these norms (in the sense of convex analysis [24]). Specifically,

$$y \in \partial(g \circ \sigma)(x)$$

holds if and only if there are vectors w and z in $\mathbf{R}^p$ satisfying $z \in \partial g(w)$, $x = u(\mathrm{Diag}\, w)v$ and $y = u(\mathrm{Diag}\, z)v$ (see [32,33]). Such analysis also helps us understand the geometry of the corresponding unit balls: for example, x is an extreme (exposed, smooth) point of the unit ball of $g \circ \sigma$ if and only if $\sigma(x)$ is an extreme (exposed, smooth) point of the unit ball of g (see [1] and also [34,35,7,6,8]).

Of interest in the current intensive study of 'semidefinite programming' (see for example [29]), although considerably less well known than von Neumann's result, is a 1957 theorem of Davis [4] giving an analogous characterization of *weakly unitarily invariant* convex functions on the space $\mathbf{H}^n$ of $n \times n$ Hermitian matrices (that is, functions f on $\mathbf{H}^n$ satisfying $f(u^*xu) = f(x)$ for all unitary matrices u and Hermitian matrices x). Davis's result states that such functions are those of the form $g \circ \lambda$, where the map

$$x \in \mathbf{H}^n \mapsto \lambda(x) \in \mathbf{R}^n$$

has as components the eigenvalues $\lambda_1(x) \geq \lambda_2(x) \geq \cdots \geq \lambda_n(x)$ of x and g is a convex function, invariant under coordinate permutations. Functions $g \circ \lambda$ are sometimes called *spectral* [11].

It transpires that a duality relationship analogous to (1.1) also exists in this setting (with the same matrix inner product):

$$(g \circ \lambda)^* = g^* \circ \lambda, \tag{1.2}$$

where $*$ denotes the Legendre–Fenchel conjugate [24]. Furthermore, an analogous characterization of subgradients holds, and we can analyze the geometry of 'spectral' convex sets (sets determined by eigenvalue properties) in a parallel manner to our previous summary for unitarily invariant unit balls. These results appear in [19].

It seems clear that these important families of results are closely related: a unifying framework should help our understanding and perhaps lead to interesting generalizations. That unifying framework is the aim of this paper. The result is a surprising and elegant interplay between semisimple Lie theory and classical convex analysis.

2. The Kostant convexity theorem

In both von Neumann's characterization of unitarily invariant norms and Davis's characterization of weakly unitarily invariant convex functions, a key feature is the relationship between a convex matrix function and its restriction to the diagonal matrices. This suggests that a Cartan subspace of a semisimple Lie algebra is the natural setting, and in this framework the central result about convexity is the Kostant convexity theorem outlined below.

We begin by fixing some notation. Consider a real semisimple Lie algebra $\mathbf{g}$ with a fixed Cartan decomposition

$$\mathbf{g} = \mathbf{t} \oplus \mathbf{p}.$$

In this decomposition, $\mathbf{t}$ is a subalgebra of $\mathbf{g}$ tangent to a maximal compact subgroup K of the adjoint group $\mathrm{Int}(\mathbf{g})$, and the vector space sum is direct with respect to the Killing form (denoted $(\cdot,\cdot)$). We will follow the notation and basic results of Onishchik and Vinberg [23, Chapter 5, Sections 3,4].

We write $\mathrm{Ad} : K \to GL(\mathbf{p})$ and $\mathrm{ad} : \mathbf{g} \to GL(\mathbf{p})$ for the natural adjoint representations. A subalgebra $\mathbf{a}$ of $\mathbf{g}$ is $\mathbf{R}$-*diagonalizable* if $\mathbf{g}$ has a basis with respect to which every operator $\mathrm{ad}\, x$ (for x in $\mathbf{a}$) is represented as a diagonal matrix. Given such a subalgebra $\mathbf{a}$ contained in the Cartan subspace $\mathbf{p}$, we define the normalizer and centralizer subgroups

$$N_K(\mathbf{a}) = \{k \in K \mid (\mathrm{Ad}\, k)\mathbf{a} = \mathbf{a}\},$$

$$Z_K(\mathbf{a}) = \{k \in K \mid (\mathrm{Ad}\, k)x = x, \text{ for all } x \text{ in } \mathbf{a}\}.$$

The subalgebra $\mathbf{a}$ with inner product $(\cdot,\cdot)$ is a Euclidean space, and (with a slight abuse of notation) the restricted homomorphism $\mathrm{Ad} : N_K(\mathbf{a}) \to GL(\mathbf{a})$ has kernel $Z_K(\mathbf{a})$ and range the associated Weyl group

$$W \simeq \frac{N_K(\mathbf{a})}{Z_K(\mathbf{a})}$$

(see [23, p. 278]).

Let us fix a closed Weyl chamber D in $\mathbf{a}$. Since for any element x of $\mathbf{p}$, the orbit $(\mathrm{Ad}\, K)x$ intersects $\mathbf{a}$ in a W-orbit (see for example [14, p. 285]), we can make the following definition.

Definition 2.1. For any element x of the Cartan subspace $\mathbf{p}$, let $\gamma(x)$ denote the unique element of the closed Weyl chamber D which is conjugate to x under $\mathrm{Ad}\, K$.

Let $P_{\mathbf{a}} : \mathbf{p} \to \mathbf{a}$ denote the orthogonal projection with respect to the Killing form.

Theorem 2.2 (Kostant [17]). *Let x be an element of $\mathbf{p}$. Then the projected orbit $P_{\mathbf{a}}((\mathrm{Ad}\, K)x)$ is the convex hull of the W-orbit of $\gamma(x)$.*

For clarity, consider the following example.

Example 2.3 (*Spectral Decomposition*). Let $\mathbf{g} = \mathbf{sl}(n, \mathbf{R})$, the special linear Lie algebra of all $n \times n$ real matrices with trace zero. The usual Cartan decomposition has $\mathbf{t} = \mathbf{so}(n)$, the skew-symmetric matrices, and $\mathbf{p} = \mathbf{p}_n$, the symmetric matrices with trace zero, which corresponds to the maximal compact subgroup $K = SO(n)$, the special orthogonal group. The Killing form is given by $(x, z) = 2n\, \mathrm{tr}(xz)$ (see for example [28, p. 268] or [23, p. 310]). We can choose the maximal subalgebra $\mathbf{a}$ as the diagonal matrices with trace zero.

Now standard calculations [28, p. 268] give the normalizer $N_K(\mathbf{a})$ as the group of matrices with determinant 1 and each row and column having exactly one nonzero entry of ± 1, and the centralizer $Z_K(\mathbf{a})$ as the group of diagonal matrices with determinant 1 and diagonal entries ± 1. Hence the Weyl group acts on $\mathbf{a}$ by permuting diagonal entries: W is isomorphic to $\mathscr{P}_n$, the group of $n \times n$ permutation matrices.

Let the closed Weyl chamber D be the set of diagonal matrices with trace zero and nonincreasing diagonal entries. Then, in the notation of Definition 2.1, for any matrix x in $\mathbf{p}_n$, $\gamma(x)$ is just $\mathrm{Diag}\,\lambda(x)$, the diagonal matrix with diagonal entries the eigenvalues of x arranged in nonincreasing order.

An easy calculation shows the projection $P_\mathbf{a}$ acts on a matrix in $\mathbf{p}_n$ by setting off-diagonal entries to zero. Thus Kostant's theorem specializes to the following result of Horn [5], related to earlier work of Schur [25] – see [21, pp. 22, 218]. (The map $\mathrm{diag} : \mathbf{p}_n \to \mathbf{R}^n$ extracts the diagonal of a matrix.)

Corollary 2.4 (Horn and Johnson [15]). *For any real symmetric matrix x with trace zero,*

$$\mathrm{diag}\{u^{\mathrm{T}}xu \,|\, u \in SO(n)\} = \mathrm{conv}(\mathscr{P}_n\lambda(x)).$$

Returning to our goal of unifying the results in the introduction, the convexity characterizations in both cases are immediate consequences of the following general result. We say a function g on $\mathbf{a}$ is *W-invariant* if g is constant on any W-orbit, and we make an analogous definition for $(\mathrm{Ad}\,K)$-invariant functions on $\mathbf{p}$: clearly such functions are exactly those of the form $g \circ \gamma$, where g is W-invariant. The next result characterizes convexity for such functions (see [18] and [30, Theorem 1.2], extending earlier results in [3]).

Theorem 2.5 (Invariant convex functions). *Let $g : \mathbf{a} \to [-\infty, +\infty]$ be a W-invariant function. Then g is convex if and only if $g \circ \gamma$ is convex on $\mathbf{p}$.*

The proof is an easy consequence of Kostant's theorem: if g is convex we deduce $g \circ \gamma \circ P_\mathbf{a} \leq g \circ \gamma$, and the convexity of $g \circ \gamma$ follows easily. (The converse is immediate.)

Davis's characterization of weakly unitarily invariant convex functions follows immediately by applying this theorem to the Hermitian version of the Spectral Decomposition Example 2.3). The substantial part of von Neumann's characterization (namely the convexity of $g \circ \sigma$) follows by considering the usual Cartan decomposition of $\mathbf{su}(p, q)$, the Lie algebra of trace zero, $(p+q) \times (p+q)$ complex matrices of the form

$$\begin{pmatrix} u & v \\ v^* & w \end{pmatrix}$$

with u and w, respectively, $p \times p$ and $q \times q$ skew-Hermitian, into the subalgebra with $v = 0$ and the Cartan subspace $\mathbf{p}$ with $u = 0$ and $w = 0$.

What about the duality relationships (1.1) and(1.2), the subgradient characterizations, and the extremal geometric structure we outlined in the introduction? We see in the next section that the framework for Kostant's theorem also suffices to explain these results.

More recent work has extended Kostant's fundamental framework in various directions [2,12,13,16]. However, since the original setting suffices to explain the examples of primary interest in optimization and matrix analysis, we do not pursue these refinements.

3. Group invariant convex analysis

The more sophisticated convex-analytic results outlined in the introduction arise elegantly by combining the framework of the Kostant convexity theorem with a straightforward algebraic structure introduced in [20].

Definition 3.1. Given a Euclidean space $(E, \langle\cdot,\cdot\rangle)$, a closed subgroup G of the orthogonal group $O(E)$, and a G-invariant map $\phi : E \to E$, we say (E, G, ϕ) is a *normal decomposition system* if

(i) for any point x in E there is a transformation ψ in G satisfying $\psi(x) = \phi(x)$, and

(ii) any points d and e in E satisfy the inequality

$$\langle d, e\rangle \leq \langle \phi(d), \phi(e)\rangle. \tag{3.2}$$

This structure in fact corresponds exactly with the idea of a *group-induced cone preordering* [22, Definition 2.2], and an *Eaton triple* [27]: the range of ϕ is the cone inducing the ordering.

We return to the notation of the previous section.

Proposition 3.3. $(\mathbf{a}, W, \gamma|_{\mathbf{a}})$ *is a normal decomposition system.*

Proof. We just need to check inequality (3.2), and this is [17, Lemma 3.2]. □

In fact, as proved in [9], (E, W, γ) is a normal decomposition system whenever W is a finite reflection group on E and γ maps W-orbits to their (unique) intersection with a fixed closed Weyl chamber: see also [18]. Indeed, all normal decomposition systems (E, G, ϕ) with the group G finite must have this form, by [26, Theorem 4.1]. The characterization problem for general groups G is studied in [22].

Proposition 3.4. $(\mathbf{p}, \mathrm{Ad}\,K, \gamma)$ *is a normal decomposition system.*

Proof. Again we just need to check inequality (3.2), and this is an easy consequence of Kostant's theorem and the previous proposition. A more direct proof simply considers the stationarity condition at an optimal solution of the variational problem

$$\max(x, (\mathrm{Ad}\,K)y)$$

for fixed points x and y in $\mathbf{p}$: see [18]. □

A similar result holds for compact Lie groups: see [27, Example 2].

In the Spectral Decomposition Example (2.3), inequality (3.2) becomes the fundamental inequality, for $n \times n$ symmetric matrices x and y,

$$\mathrm{tr}(xy) \le \lambda(x)^{\mathrm{T}}\lambda(y),$$

(see for example [15]) while in the case $\mathbf{g} = \mathbf{su}(p,q)$ discussed at the end of the previous section we obtain von Neumann's famous inequality

$$\mathrm{Re}\,\mathrm{tr}(x^* y) \le \sigma(x)^{\mathrm{T}}\sigma(y) \quad \text{for all } x, y \in \mathbf{C}^{p\times q}.$$

With these two propositions in place we can now apply all the consequences of [20, Assumption 4.1]. The following list is not exclusive but covers the results outlined in the introduction. By the remark above, a similar result holds for compact Lie groups.

Theorem 3.5 (Conjugates and subgradients). *Any W-invariant function $g : \mathbf{a} \to [-\infty, +\infty]$ satisfies*

$$(g \circ \gamma)^* = g^* \circ \gamma.$$

For any elements x and y of $\mathbf{p}$ with $g(\gamma(x))$ finite, $y \in \partial(g\circ\gamma)(x)$ holds if and only if $\gamma(y) \in \partial g(\gamma(x))$ holds and there exists an element k of the group K such that x and y both lie in $(\mathrm{Ad}\,k)D$. Furthermore, $g\circ\gamma$ is convex, or essentially strictly convex, or essentially smooth [24] *if and only if g is likewise.*

Proof. Apply Theorems 4.4, 4.5 and Corollary 6.2 in [20]. □

We say a subset C of a Euclidean space is *invariant* under a group G of orthogonal linear transformations if $\psi C = C$ for any transformation ψ in G.

Theorem 3.6 (Invariant convex sets). *An $(\mathrm{Ad}\,K)$-invariant subset C of $\mathbf{p}$ is convex if and only if $C \cap \mathbf{a}$ is convex. In this case an element x of $\mathbf{p}$ is an extreme (exposed) point of C if and only if $\gamma(x)$ is an extreme (exposed) point of $C \cap \mathbf{a}$.*

Proof. The convexity is an immediate consequence of Kostant's theorem. (The symmetric matrix case for closed sets C appeared in [10].) The extremal properties follow from Theorem 5.5 and Corollary 6.3 in [20]. □

Theorem 3.7 (Invariant norms). *$(\mathrm{Ad}\,K)$-invariant norms on $\mathbf{p}$ are those functions of the form $g\circ\gamma$, where g is a W-invariant norm on $\mathbf{a}$. In this case the dual norms satisfy*

$$(g\circ\gamma)^D = g^D \circ \gamma,$$

and an element x of $\mathbf{p}$ is an extreme (exposed, smooth) point of the unit ball of $g\circ\gamma$ if and only if $\gamma(x)$ is an extreme (exposed, smooth) point of the unit ball of g: in particular, $g\circ\gamma$ is a strict (smooth) norm if and only if g is likewise.

Proof. Apply [20, Theorem 6.5]. □

The results in the introduction follow by considering the Lie algebras $\mathbf{sl}(n,\mathbf{C})^{\mathbf{R}}$ and $\mathbf{su}(p,q)$. Other interesting cases arise from the other classical simple Lie algebras: see [18].

Acknowledgements

Many thanks to Professor D. Djoković for his patient help and in particular for pointing out the unifying possibilities of the Cartan decomposition, to Professors J. Hilgert and K.-H. Neeb for helpful discussions about Kostant's theorem, to Professor J.-B. Hiriart-Urruty for bringing [30,3] to his attention, and to Professor P. Diaconis for pointing out [9].

References

[1] J. Arazy, On the geometry of the unit ball of unitary matrix spaces, Integral Equations Operator Theory 4 (1981) 151–171.
[2] M.F. Atiyah, Convexity and commuting Hamiltonians, Bull. London Math. Soc. 14 (1982) 1–15.
[3] B. Dacorogna, H. Koshigoe, On the different notions of convexity for rotationally invariant functions, Ann. Fac. Sci. Toulouse 2 (1993) 163–184.
[4] C. Davis, All convex invariant functions of hermitian matrices, Arch. Math. 8 (1957) 276–278.
[5] R.A. Horn, Doubly stochastic matrices and the diagonal of a rotation matrix, Amer. J. Math. 76 (1954) 620–630.
[6] E.M. de Sá, Exposed faces and duality for symmetric and unitarily invariant norms, Linear Algebra Appl. 197,198 (1994) 429–450.
[7] E.M. de Sá, Faces and traces of the unit ball of a symmetric gauge function, Linear Algebra Appl. 197,198 (1994) 349–395.
[8] E.M. de Sá, Faces of the unit ball of a unitarily invariant norm, Linear Algebra Appl. 197,198 (1994) 451–493.
[9] M.L. Eaton, M.D. Perlman, Reflection groups, generalized Schur functions, and the geometry of majorization, Ann. Probab. 5 (1977) 829–860.
[10] P.A. Fillmore, J.P. Williams, Some convexity theorems for matrices, Glasgow Math. J. 12 (1971) 110–117.
[11] S. Friedland, Convex spectral functions, Linear and Multilinear Algebra 9 (1981) 299–316.
[12] V. Guillemin, S. Sternberg, Convexity properties of the moment mapping, Invent. Math. 67 (1982) 491–513.
[13] V. Guillemin, S. Sternberg, Convexity properties of the moment mapping II, Invent. Math. 77 (1984) 533–546.
[14] S. Helgason, Differential Geometry, Lie Groups, and Symmetric Spaces, Academic Press, New York, 1978.
[15] R.A. Horn, C. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 1985.
[16] F. Kirwan, Convexity properties of the moment mapping, III, Invent. Math. 77 (1984) 547–552.
[17] B. Kostant, On convexity, the Weyl group and the Iwasawa decomposition, Ann. Sci. l'Ecole Norm. Sup. 6 (1973) 413–455.
[18] A.S. Lewis, Von Neumann's lemma and a Chevalley-type theorem for convex functions on Cartan subspaces, Technical Report CORR 95-19, University of Waterloo, 1995.
[19] A.S. Lewis, Convex analysis on the Hermitian matrices, SIAM J. Optim. 6 (1996) 164–177.
[20] A.S. Lewis, Group invariance and convex matrix analysis, SIAM J. Matrix Anal. Appl. 17 (1996) 927–949.
[21] A.W. Marshall, I. Olkin, Inequalities: Theory of Majorization and its Applications, Academic Press, New York, 1979.

[22] M. Niezgoda, Group majorization and Schur type inequalities, Linear Algebra Appl. 268 (1998) 9–30.
[23] A.L. Onishchik, E.B. Vinberg, Lie Groups and Algebraic Groups, Springer, Berlin, 1990.
[24] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[25] I. Schur, Uber eine Klasse von Mittelbildungen mit Anwendungen auf der Determinantentheorie, Sitzungsber. Berliner Math. Gesellschaft 22 (1923) 9–20.
[26] A.G.M. Steerneman, G-majorization, group-induced cone orderings and reflection groups, Linear Algebra Appl. 127 (1990) 107–119.
[27] B.T.-Y. Tam, An extension of a result of Lewis, Technical report, Auburn University, 1998.
[28] A. Terras, Harmonic Analyis on Symmetic Spaces and Applications II, Springer, New York, 1988.
[29] L. Vandenberghe, S. Boyd, Semidefinite programming, SIAM Rev. 38 (1996) 49–95.
[30] F. Vincent, Une note sur les fonctions invariantes, Ann. Fac. Sci. Toulouse 6 (1997) 357–363.
[31] J. von Neumann, Some matrix inequalities and metrization of matric-space, Tomsk Univ. Rev. 1 (1937) 286–300. In: Collected Works, Pergamon, Oxford, 1962, vol. IV, pp. 205–218.
[32] G.A. Watson, Characterization of the subdifferential of some matrix norms, Linear Algebra Appl. 170 (1992) 33–45.
[33] G.A. Watson, On matrix approximation problems with Ky Fan *k* norms, Numer. Algorithms 5 (1993) 263–272.
[34] K. Ziętak, On the characterization of the extremal points of the unit sphere of matrices, Linear Algebra Appl. 106 (1988) 57–75.
[35] K. Ziętak, Subdifferentials, faces and dual matrices, Linear Algebra Appl. 185 (1993) 125–141.