SIAM J. MATRIX ANAL. APPL.
Vol. 17, No. 4, pp. 927–949, October 1996

013

GROUP INVARIANCE AND CONVEX MATRIX ANALYSIS*

A. S. LEWIS†

Abstract. Certain interesting classes of functions on a real inner product space are invariant under an associated group of orthogonal linear transformations. This invariance can be made explicit via a simple decomposition. For example, rotationally invariant functions on $\mathbf{R}^2$ are just even functions of the Euclidean norm, and functions on the Hermitian matrices (with trace inner product) which are invariant under unitary similarity transformations are just symmetric functions of the eigenvalues. We develop a framework for answering geometric and analytic (both classical and nonsmooth) questions about such a function by answering the corresponding question for the (much simpler) function appearing in the decomposition. The aim is to understand and extend the foundations of eigenvalue optimization, matrix approximation, and semidefinite programming.

Key words. convexity, group invariance, nonsmooth analysis, semidefinite program, eigenvalue optimization, Fenchel conjugate, subdifferential, spectral function, unitarily invariant norm, Schur convex, extreme point, von Neumann's lemma

AMS subject classifications. Primary, 15A45, 90C25; Secondary, 49J52, 65K05

1. Introduction.

Why is there such a strong parallel between, on the one hand, semidefinite programming and other eigenvalue optimization problems, and on the other hand, ordinary linear programming and related problems? Why are there close analogies between many important matrix norms on the one hand, and associated vector norms on the other? This paper aims to explain the simple algebraic symmetries which drive these parallels.

A simple example may be illustrative. Suppose that we wish to understand convex functions $f : \mathbf{R}^n \to \mathbf{R}$ which are "orthogonally invariant." By this we mean that $f(x) = f(Ux)$ for any point x in $\mathbf{R}^n$ and any orthogonal matrix U. What can we say about such functions?

We might observe first that, since f is determined by its behaviour on the half-line $\{\beta e^1 \mid \beta \geq 0\}$, where $e^1 = (1, 0, 0, \ldots, 0)$, we can write $f(x) = h(\|x\|)$, where the function $h : \mathbf{R}_+ \to \mathbf{R}$ is defined by $h(\beta) = f(\beta e^1)$. What conditions on h are equivalent to the convexity of f? Clearly h must be convex (being the restriction of f to a half-line), but this is not sufficient.

After some more thought we might arrive at the following answer: h must be convex and nondecreasing at the origin. But this obscures the essential symmetry of f. A simple trick allows us to preserve this in our answer. Instead of examining the restriction of f to the half-line $\mathbf{R}_+ e^1$ we consider the restriction to the whole subspace $\mathbf{R}e^1$. We then arrive at the following much more satisfactory answer: $h(\|\cdot\|)$ is convex if and only if the function $h : \mathbf{R} \to \mathbf{R}$ is *even* and convex.

This easy example illustrates the fundamental technique of this paper—analyzing the consequences of the symmetries of a function by analyzing its symmetries on a "transversal" (or defining) subspace. von Neumann's famous 1937 characterization of unitarily invariant matrix norms [27] is precisely of this mold. One statement of this result is that a unitarily invariant matrix function f (one satisfying $f(x) = f(uxv)$

* Received by the editors March 17, 1995; accepted for publication (in revised form) by M. L. Overton November 20, 1995. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada.

† Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada (aslewis@orion.uwaterloo.ca).

for any unitary u and v) is a norm exactly when its restriction to the subspace of real diagonal matrices is a symmetric gauge function.

What algebraic structure underlies von Neumann's result? There are three essential ingredients: first, a real inner product space X (in this case $X = \mathbf{C}^{m\times n}$ with $\langle x, w\rangle = \operatorname{Re}\operatorname{tr} x^*w$); second, a (closed) group $\mathcal{G}$ of orthogonal linear transformations (in this case those of the form $x \mapsto uxv$ for unitary u and v); third, a map γ from X to a transversal subspace (in this case $\gamma(x)$ is the diagonal matrix with diagonal entries the singular values of x arranged in nonincreasing order). The map γ should be $\mathcal{G}$-invariant and should satisfy the following conditions.

AXIOM 1.1 (decomposition). *Any element x of X can be decomposed as $x = A\gamma(x)$ for some operator A in $\mathcal{G}$.*

AXIOM 1.2 (angle contraction). *Any elements x and w in X satisfy the inequality $\langle x, w\rangle \le \langle \gamma(x), \gamma(w)\rangle$.*

In von Neumann's case, Axiom 1.1 is just the singular value decomposition, and Axiom 1.2 is "von Neumann's lemma" (see, for example, [7]).

This structure $(X, \mathcal{G}, \gamma)$ (which we call a *normal decomposition system*) is the focus of this paper. Our aim is to analyze $\mathcal{G}$-invariant functions on X via their restriction on the range of γ. For this to be of much interest we would hope that the range of γ has lower dimension than X. Our other main example, of fundamental interest in matrix optimization, also has this property:

$$\begin{aligned} X &= \{n \times n \text{ symmetric matrices}\}, \quad \text{with } \langle x, w\rangle = \operatorname{tr} xw, \\ \mathcal{G} &= \{x \mapsto u^T x u \mid u \text{ orthogonal}\}, \quad \text{and} \\ \gamma(x) &= \operatorname{Diag}\lambda(x), \quad \text{where} \\ &\qquad \lambda_1(x) \ge \lambda_2(x) \ge \cdots \ge \lambda_n(x) \quad \text{are the eigenvalues of } x. \end{aligned}$$

In a later paper [22] a broad family of examples generated from the theory of semisimple Lie groups will be discussed. In this paper we concentrate on outlining how the idea of a normal decomposition system provides a simple yet powerful unifying framework in which to study a wide variety of important results. Examples include Schur convexity (see, for example, [23]), the convexity of eigenvalue functions [10, 6, 11, 3, 13, 18], calculations of Fenchel conjugates and subdifferentials of convex eigenvalue functions [26, 5, 12, 34, 31, 28, 29, 30, 15, 16, 1, 17, 18, 24, 33], von Neumann's original result [27] and generalizations (for example, [4, 19]), subdifferentials of unitarily invariant norms [37, 38, 39, 40, 41, 8, 7, 9, 19], and characterizations of extreme, exposed, and smooth points of unit balls [2, 40, 41, 8, 7, 9, 19].

This paper concentrates on convexity and its ramifications.

2. Group invariant normal forms. Underlying all the work in this paper is a rather simple algebraic structure. We therefore begin by fixing our notation and formally defining this structure.

We will work in a real inner product space X. For simplicity we will assume that X is finite dimensional, although many of our results extend easily. The *adjoint* of a linear operator $A : X \to X$ is the linear operator $A^* : X \to X$ defined by $\langle A^*w, x\rangle = \langle w, Ax\rangle$ for all points w and x in X. We denote the identity operator by $\mathrm{id} : X \to X$, and if $A^*A = \mathrm{id}$ then we say that A is *orthogonal.* In fact, A is orthogonal if and only if it is norm preserving: $\|Ax\| = \|x\|$ for all x in X, where the norm is defined by $\|x\| = \sqrt{\langle x, x\rangle}$. In this case, $A^{-1} = A^*$.

We denote the group of all orthogonal linear operators on X (with composition) by $O(X)$, which we endow with the natural topology. Thus A_r approaches A in $O(X)$

if and only if $A_r x$ approaches Ax in X for all x in X. Given a subgroup $\mathcal{G}$ of $O(X)$, a function f on X is *$\mathcal{G}$-invariant* if $f(Ax) = f(x)$ for all x in X and A in $\mathcal{G}$. We can now describe our fundamental structure—this structure is an underlying assumption throughout the paper.

DEFINITION 2.1. *Given a real inner product space X and a closed subgroup $\mathcal{G}$ of the orthogonal group $O(X)$, the map $\gamma : X \to X$ induces a $\mathcal{G}$-*invariant normal form *on X if*

(a) *γ is $\mathcal{G}$-invariant,*

(b) *for any point x in X there is an operator A in $\mathcal{G}$ with $x = A\gamma(x)$, and*

(c) *any points x and w in X satisfy the inequality $\langle x, w\rangle \le \langle \gamma(x), \gamma(w)\rangle$.*

In this case $(X, \mathcal{G}, \gamma)$ is called a normal decomposition system.

Notice two immediate consequences of this definition: the map γ must be *idempotent*, since for any point x in X, properties (a) and (b) imply $\gamma(\gamma(x)) = \gamma(A^*x) = \gamma(x)$, and furthermore γ must be norm preserving, since $\|\gamma(x)\| = \|A^*x\| = \|x\|$. Our first result, which is somewhat less trivial, has the following important corollary.

- The condition for equality in property (c) is the existence of an operator A in $\mathcal{G}$ with $x = A\gamma(x)$ and $w = A\gamma(w)$.

THEOREM 2.2. *A subset K of X has the property that $\langle x, w\rangle = \langle \gamma(x), \gamma(w)\rangle$ for every pair of elements x and w of K if and only if there is an operator A in $\mathcal{G}$ satisfying $x = A\gamma(x)$ for all x in K.*

Proof. The "if" direction is easy, so consider the "only if" direction. Without loss of generality, K is nonempty, so choose a point z in $\mathrm{ri}\,(\mathrm{conv}K)$ and an A in $\mathcal{G}$ for which $z = A\gamma(z)$. If there is a point x in K with $x \ne A\gamma(x)$ then the Cauchy–Schwartz inequality implies that $\langle x, A\gamma(x)\rangle < \|x\|^2$. It is easy to write z as a convex combination $z = \alpha_0 x + \sum_{i>0} \alpha_i x^i$ for strictly positive α_i's with sum 1 and points x^i in K. But now we have

$$\begin{aligned}
\langle \gamma(x), \gamma(z)\rangle &= \langle A\gamma(x), A\gamma(z)\rangle = \langle A\gamma(x), z\rangle \\
&< \alpha_0\|x\|^2 + \langle A\gamma(x), \textstyle\sum_{i>0} \alpha_i x^i\rangle \\
&\le \alpha_0\|x\|^2 + \sum_{i>0} \alpha_i \langle \gamma(x), \gamma(x^i)\rangle \\
&= \alpha_0\|x\|^2 + \sum_{i>0} \alpha_i \langle x, x^i\rangle \\
&= \langle x, z\rangle \le \langle \gamma(x), \gamma(z)\rangle,
\end{aligned}$$

which is a contradiction. $\square$

We defer a systematic discussion of examples until the end of the paper. However, for the sake of concreteness the reader may wish to keep in mind the following extremely simple example: $X = \mathbf{R}$ (with $\langle x, w\rangle = xw$), $\mathcal{G} = \{\pm \mathrm{id}\}$, and $\gamma(x) = |x|$. The properties are easily verified.

We will only use the closedness of $\mathcal{G}$ very rarely (specifically, in Theorem 3.3), but it does not rule out much of interest. We think of the formula $x = A\gamma(x)$ in property (b) as being a "normal form decomposition" of x. Property (c) expresses the fact that γ contracts the angle between the vectors x and w unless they have a simultaneous normal form decomposition (in which case the angle remains constant). If we write

$$\mathcal{G}^x = \{A \in \mathcal{G} \mid x = A\gamma(x)\}, \tag{2.1}$$

then (b) says that $\mathcal{G}^x$ is nonempty, while the condition for equality in (c) is that $\mathcal{G}^x \cap \mathcal{G}^w$ be nonempty.

PROPOSITION 2.3. *For points x and w in X,*

$$\max_{A \in \mathcal{G}} \langle x, Aw \rangle = \langle \gamma(x), \gamma(w) \rangle.$$

Proof. Note that $\langle x, Aw \rangle \leq \langle \gamma(x), \gamma(Aw) \rangle = \langle \gamma(x), \gamma(w) \rangle$ for any operator A in $\mathcal{G}$. On the other hand, since there exist operators B and C in $\mathcal{G}$ with $x = B\gamma(x)$ and $w = C\gamma(w)$, we have that $\langle \gamma(x), \gamma(w) \rangle = \langle B^*x, C^*w \rangle = \langle x, BC^*w \rangle$, so the maximum is attained by $A = BC^*$. □

Given a subset K of X, the *dual cone* of K is defined to be the closed, convex cone

$$K^+ = \{w \in X \mid \langle x, w \rangle \geq 0 \text{ for all } x \text{ in } K\}.$$

The set K is a closed, convex cone if and only if $K = K^{++}$ [32, Thm. 14.1]. The function γ is K^+-*convex* if the real function $\langle \gamma(\cdot), w \rangle$ is convex for all vectors w in K, and a function $f : K \to [-\infty, +\infty]$ is K^+-*isotone* if $f(x) \geq f(w)$ for any x and w in K satisfying $x - w \in K^+$.

It transpires that Definition 2.1 has strong implications for possible maps γ.

THEOREM 2.4. *The range $R(\gamma)$ of the map γ is a closed, convex cone. Furthermore, γ is norm preserving, positively homogeneous, and $R(\gamma)^+$-convex with global Lipschitz constant* 1.

Proof. For any point x in X it follows from Definition 2.1 that $\langle x, w \rangle \leq \langle \gamma(x), w \rangle$ for all points w in $R(\gamma)$, and hence $\gamma(x) - x \in R(\gamma)^+$. If in particular x lies in $R(\gamma)^{++}$ then

$$0 \leq \langle x, \gamma(x) - x \rangle = \langle x, \gamma(x) \rangle - \|x\|^2 \leq 0$$

since, as we have seen, $\|\gamma(x)\| = \|x\|$. It follows that $x = \gamma(x) \in R(\gamma)$, so $R(\gamma)^{++} \subset R(\gamma)$, and hence $R(\gamma)$ is a closed, convex cone.

Supposing once more that x lies in X and that the scalar λ is nonnegative, we have

$$\begin{aligned}
\|\gamma(\lambda x) - \lambda\gamma(x)\|^2 &= \|\gamma(\lambda x)\|^2 + \lambda^2\|\gamma(x)\|^2 - 2\lambda\langle \gamma(\lambda x), \gamma(x) \rangle \\
&\leq \|\lambda x\|^2 + \lambda^2\|x\|^2 - 2\lambda\langle \lambda x, x \rangle \\
&= 0;
\end{aligned}$$

whence $\gamma(\lambda x) = \lambda\gamma(x)$. Thus γ is positively homogeneous.

By Proposition 2.3, for any w in $R(\gamma)$ we have

$$\langle \gamma(x), w \rangle = \max_{A \in \mathcal{G}} \langle x, Aw \rangle,$$

and hence $\langle \gamma(\cdot), w \rangle$ is convex, being a pointwise maximum of linear functions. Thus γ is $R(\gamma)^+$-convex.

Finally, for any x and w in X,

$$\begin{aligned}
\|\gamma(x) - \gamma(w)\|^2 &= \langle \gamma(x), \gamma(x) \rangle + \langle \gamma(w), \gamma(w) \rangle - 2\langle \gamma(x), \gamma(w) \rangle \\
&\leq \|x\|^2 + \|w\|^2 - 2\langle x, w \rangle \\
&= \|x - w\|^2,
\end{aligned}$$

whence the Lipschitz constant 1. □

Various algebraic ideas can be applied naturally to the concept of a normal decomposition system. For example, we say that two normal decomposition systems $(X_1, \mathcal{G}_1, \gamma_1)$ and $(X_2, \mathcal{G}_2, \gamma_2)$ are *isomorphic* if there is an inner product space isomorphism $\alpha : X_1 \to X_2$ and a group isomorphism $\beta : \mathcal{G}_1 \to \mathcal{G}_2$ such that for all points x in X_1 and operators A in $\mathcal{G}_1$ we have $\gamma_2(\alpha(x)) = \alpha(\gamma_1(x))$ and $(\beta(A))(\alpha(x)) = \alpha(Ax)$. There is also a natural notion of the Cartesian product of two normal decomposition systems. Observe finally that, given any inner product space X and subgroup $\mathcal{G}$ of $O(X)$, easy examples show that there may be no map γ with $(X, \mathcal{G}, \gamma)$ a normal decomposition system.

3. $\mathcal{G}$-invariant functions and sets. The main aim of this paper is to study functions $f : X \to [-\infty, +\infty]$ on the inner product space X which are $\mathcal{G}$-invariant: $f(Ax) = f(x)$ for all points x in X and operators A in the group $\mathcal{G}$. As usual, we assume that the map γ induces a $\mathcal{G}$-invariant normal form on X in the sense of Definition 2.1.

We will be particularly interested in *convex* functions f, which we define by requiring that the *epigraph*

$$\operatorname{epi} f = \{(x, \alpha) \in X \times \mathbf{R} \mid \alpha \geq f(x)\}$$

be a convex set. The function f is *closed* if its epigraph is closed and is *proper* if it never takes the value $-\infty$ and has nonempty *domain*,

$$\operatorname{dom} f = \{x \in X \mid f(x) < +\infty\}.$$

The *(Fenchel) conjugate* of f is the closed, convex function $f^* : X \to [-\infty, +\infty]$ defined by

$$f^*(w) = \sup\{\langle x, w \rangle - f(x) \mid x \in X\}.$$

For proper, convex f, the conjugate f^* is also proper with $f^{**} = f$ providing that f is also closed. For proper f we can define the *(convex) subdifferential* at a point x in $\operatorname{dom} f$ by

$$\partial f(x) = \{w \in X \mid f(x) + f^*(w) = \langle x, w \rangle\}.$$

Elements of the subdifferential are called *subgradients*. For all of these ideas the standard reference is [32].

The following result is rather reminiscent of the discussion in [32, pp. 110–111]. It shows that conjugacy preserves $\mathcal{G}$-invariance.

PROPOSITION 3.1. *If the function $f : X \to [-\infty, +\infty]$ is $\mathcal{G}$-invariant then so is the conjugate function f^*, and*

$$f^*(w) = \sup\{\langle x, \gamma(w) \rangle - f(x) \mid x \in R(\gamma)\}$$

for any point x in X.

Proof. For any operator A in $\mathcal{G}^w$ (whence $w = A\gamma(w)$),

$$\begin{aligned} f^*(w) &= \sup\{\langle z, w \rangle - f(z) \mid z \in X\} \\ &= \sup\{\langle Bx, A\gamma(w) \rangle - f(Bx) \mid x \in R(\gamma),\ B \in \mathcal{G}\} \\ &= \sup\{\sup\{\langle x, B^* A\gamma(w) \rangle \mid B \in \mathcal{G}\} - f(x) \mid x \in R(\gamma)\} \\ &= \sup\{\langle x, \gamma(w) \rangle - f(x) \mid x \in R(\gamma)\} \end{aligned}$$

by Proposition 2.3. Since γ is $\mathcal{G}$-invariant, so is f^*. □

LEMMA 3.2. *A $\mathcal{G}$-invariant function $f : X \to [-\infty, +\infty]$ is (Frechet) differentiable at the point x in X if and only if it is differentiable at $\gamma(x)$.*

Proof. For any operator B in $\mathcal{G}$, we know that $f(Bw) = f(w)$ for all points w in X, and hence by the chain rule, if f is differentiable at Bw then it is differentiable at w. Choosing an operator A in $\mathcal{G}^x$ (so that $x = A\gamma(x)$), the result follows by setting $w = x$, $B = A^*$ and $w = \gamma(x)$, $B = A$ in turn. □

The next result is our first rather nontrivial observation. A consequence, for example, is that symmetric, convex functions on $\mathbf{R}^n$ are "Schur convex" (see Example 7.1).

THEOREM 3.3. *If the $\mathcal{G}$-invariant function $f : X \to [-\infty, +\infty]$ is convex then it is $R(\gamma)^+$-isotone on $R(\gamma)$: if points x and w lie in $R(\gamma)$ with $x - w$ in $R(\gamma)^+$ then $f(x) \geq f(w)$.*

Proof. The coset $\mathcal{G}x$ is compact (since $\mathcal{G}$ is compact). If w lay outside its convex hull then there would exist a separating hyperplane defined by a vector v in X with

$$\langle \gamma(v), w\rangle \geq \langle v, w\rangle > \max_{A \in \mathcal{G}} \langle v, Ax\rangle = \langle \gamma(v), x\rangle,$$

by Proposition 2.3, and then $\langle \gamma(v), x - w\rangle < 0$, contradicting the assumption that $x - w$ lies in $R(\gamma)^+$. Hence there exist positive scalars $\lambda_1, \lambda_2, \ldots, \lambda_r$ with sum 1 and operators $A_1, A_2, \ldots, A_r$ in $\mathcal{G}$ with $w = \sum_1^r \lambda_i A_i x$.

Suppose that $f(x) < f(w)$. Then we can choose a real number α in the interval $(f(x), f(w))$, and since f is $\mathcal{G}$-invariant, $f(A_i x) = f(x) < \alpha$ for each $i = 1, 2, \ldots, r$. Now since f is convex,

$$f(w) = f\left(\sum_1^r \lambda_i A_i x\right) < \sum_1^r \lambda_i \alpha = \alpha$$

(see [32, Thm. 4.2]), which is a contradiction. □

We will also be interested in $\mathcal{G}$-invariant subsets of X, so we will conclude this section with some simple observations illustrating how various algebraic and topological constructions preserve $\mathcal{G}$-invariance. Notice first that the class of $\mathcal{G}$-invariant sets is easily seen to be closed under arbitrary unions, intersections, and complements. Suppose that the subset D of X is $\mathcal{G}$-invariant (that is, $x \in D$, $A \in \mathcal{G}$ implies $Ax \in D$). Then the interior of D is quickly seen to be $\mathcal{G}$-invariant; whence the closure and boundary of D are also $\mathcal{G}$-invariant.

For each $r = 1, 2, \ldots$, suppose that Λ_r is a subset of $\mathbf{R}^r$ and define a subset of X,

$$\left\{ \sum_{i=1}^r \lambda_i x_i \;\middle|\; r \in \mathbf{N},\ \lambda \in \Lambda_r,\ x_1, x_2, \ldots, x_r \in D \right\}.$$

It is immediate that this set is $\mathcal{G}$-invariant. By taking Λ_r to be $\mathbf{R}^r$, $\mathbf{R}^r_+$, $\{\lambda \in \mathbf{R}^r \mid \sum \lambda_i = 1\}$, and $\{\lambda \in \mathbf{R}^r_+ \mid \sum \lambda_i = 1\}$ in turn we see that the *linear hull*, the *conical hull*, the *affine hull*, and the *convex hull* of D are all $\mathcal{G}$-invariant.

We say that a point x lies in the *intrinsic core* of D, written $\operatorname{icr} D$, if for any point w in the affine hull $\operatorname{aff} D$, $x + \delta(w - x)$ lies in D for all real δ sufficiently small. When D is convex its intrinsic core coincides with its *relative interior* $\operatorname{ri} D$, the interior of D relative to its affine hull (see [32]), and in this case the *relative boundary* $\operatorname{rb} D$ is just $\operatorname{cl} D \setminus \operatorname{ri} D$. Since it is easy to check that $\operatorname{icr} D$ is $\mathcal{G}$-invariant, it follows that for convex, $\mathcal{G}$-invariant D, the relative interior and boundary of D are also $\mathcal{G}$-invariant. Finally, for any $\mathcal{G}$-invariant set D the dual cone D^+ and the orthogonal complement $D^\perp$ are both $\mathcal{G}$-invariant.

4. Reduction. Let us assume once more that $(X, \mathcal{G}, \gamma)$ is a normal decomposition system in the sense of Definition 2.1. If the function $f : X \to [-\infty, +\infty]$ is $\mathcal{G}$-invariant then since $f(x) = f(\gamma(x))$ for all points x in X, the behaviour of f is determined by its behaviour on $R(\gamma)$, the range of γ. The key idea of this paper is then rather simple—we reduce questions about f to corresponding questions about the restriction of f to a subspace Y containing $R(\gamma)$: typically, $Y = R(\gamma) - R(\gamma)$.

Given a subspace Y of X, we denote the *stabilizer* of Y in $\mathcal{G}$ by

$$\mathcal{G}_Y = \{A \in \mathcal{G} \mid AY = Y\}.$$

We will frequently abuse notation and write $\mathcal{G}_Y$ for the group of restricted operators

$$\mathcal{G}_Y|_Y = \{A|_Y \mid A \in \mathcal{G}_Y\}.$$

In other words, we think of operators in $\mathcal{G}_Y$ as orthogonal transformations on Y (as well as on X). When Y contains $R(\gamma)$ we can consider the restricted map $\gamma|_Y : Y \to Y$: we will frequently write γ in place of $\gamma|_Y$.

The following central assumption will remain in force throughout the remainder of the paper.

ASSUMPTION 4.1. *In the sense of Definition* 2.1, $(X, \mathcal{G}, \gamma)$ *is a normal decomposition system. The inner product space* Y *is a subspace of* X *(with the inherited inner product) and contains the range of* γ. *Furthermore,* $(Y, \mathcal{G}_Y, \gamma)$ *is also a normal decomposition system.*

This amounts to the additional assumption that if, in Definition 2.1, the point x in fact lies in Y then the operator A in property (b) can actually be chosen to leave Y invariant. Of course a trivial example is $Y = X$. Once again, since we are deferring examples until the end of the paper, it may be helpful to keep a simple (although nontrivial) example in mind. We take X to be $\mathbf{R}^n$ with the standard inner product, $\mathcal{G} = \mathcal{O}_n$, the orthogonal group on $\mathbf{R}^n$, and let e^1 be the vector $(1, 0, 0, \ldots, 0)$. Then it is easily verified that if we define $\gamma(x) = \|x\| e^1$ for all x in $\mathbf{R}^n$ then we obtain a normal decomposition system, and that if $Y = \operatorname{span}\{e^1\}$ then Assumption 4.1 holds.

An interesting general framework in which Assumption 4.1 holds is developed in [22]. In summary, suppose that G is a real, semisimple Lie group with a maximal compact subgroup K, and that $\mathbf{g} = \mathbf{k} \oplus \mathbf{t}$ is the corresponding Cartan decomposition (where $\mathbf{g}$ and $\mathbf{k}$ are the tangent algebras of G and K, respectively). Now let $X = \mathbf{t}$, let the group $\mathcal{G}$ consist of the adjoint actions of elements of K on $\mathbf{t}$, and let Y be a maximal $\mathbf{R}$-diagonalizable subspace of $\mathbf{t}$. Then $\mathcal{G}_Y$ is (essentially) the associated Weyl group. Finally, fix a closed Weyl chamber $D \subset Y$ and for any point x in $\mathbf{t}$ define $\gamma(x)$ to be the (singleton) intersection of the $\mathcal{G}$-orbit of x with the chamber D. Then Assumption 4.1 holds; see [22] for details. In fact, all of the concrete examples which we develop later fall into this framework.

In what follows, $\circ$ denotes composition. Thus $(h \circ \gamma)(x) = h(\gamma(x))$.

PROPOSITION 4.2. *A function* f *is* $\mathcal{G}$*-invariant on* X *if and only if it can be written in the form* $f = h \circ \gamma$ *for some* $\mathcal{G}_Y$*-invariant function* h *on* Y.

Proof. Any function of the form $h \circ \gamma$ is $\mathcal{G}$-invariant since γ is. If, on the other hand, f is $\mathcal{G}$-invariant then it is immediate that $f = f|_Y \circ \gamma$, and clearly $f|_Y$ is $\mathcal{G}_Y$-invariant. $\square$

Thus henceforth we will restrict attention to $\mathcal{G}$-invariant functions $h \circ \gamma$, where the function h is $\mathcal{G}_Y$-invariant. We now follow two distinct approaches to the elegant fact that such an extended real-valued function $h \circ \gamma$ is convex on X if and only if h is convex on Y. The first approach is direct, using Theorem 3.3, and hence relies on the underlying assumption that the group $\mathcal{G}$ is closed. The second approach does not require this assumption, but instead assumes that the function h is closed and employs an attractive Fenchel conjugacy argument.

THEOREM 4.3 (convex and closed functions). *Suppose that the function $h : Y \to [-\infty, +\infty]$ is $\mathcal{G}_Y$-invariant. Then the function $h \circ \gamma$ is convex (respectively, closed) on X if and only if h is convex (respectively, closed) on Y. Hence, a $\mathcal{G}$-invariant function on X is convex (respectively, closed) if and only if its restriction to Y is convex (respectively, closed).*

Proof. Since $h = (h \circ \gamma)|_Y$, one direction is clear. Conversely, suppose that h is convex. For any points x and w in X and real λ in $(0, 1)$, we know by Theorem 2.4 that $\lambda\gamma(x) + (1 - \lambda)\gamma(w)$ and $\gamma(\lambda x + (1 - \lambda)w)$ both lie in $R(\gamma)$, and that $(\lambda\gamma(x) + (1 - \lambda)\gamma(w)) - \gamma(\lambda x + (1 - \lambda)w)$ lies in $R(\gamma)^+$. Hence by Theorem 3.3 (applied to the system $(Y, \mathcal{G}_Y, \gamma)$), we have

$$h(\lambda\gamma(x) + (1 - \lambda)\gamma(w)) \geq h(\gamma(\lambda x + (1 - \lambda)w)).$$

Now for any real numbers $\alpha > h(\gamma(x))$ and $\beta > h(\gamma(w))$, since h is convex we deduce that $h(\gamma(\lambda x + (1 - \lambda)w)) < \lambda\alpha + (1 - \lambda)\beta$; whence $h \circ \gamma$ is convex [32, Thm. 4.2].

Turning now to the closed case, since $h = (h \circ \gamma)|_Y$ we have that

$$\operatorname{epi} h = \operatorname{epi}(h \circ \gamma) \cap (Y \times \mathbf{R}), \tag{4.1}$$

so that if $h \circ \gamma$ is closed, then so is h. Suppose on the other hand that h is closed. If $\{(x_i, r_i)\}$ is a sequence of points in $\operatorname{epi}(h \circ \gamma)$ approaching the point (x, r), then since the sequence $((\gamma(x_i), r_i))$ lies in the closed set $\operatorname{epi} h$ and approaches $(\gamma(x), r)$ (as γ is continuous by Theorem 2.4), it follows that $(\gamma(x), r) \in \operatorname{epi} h$, and so $(x, r) \in \operatorname{epi}(h \circ \gamma)$. □

The second approach to convexity is rather more transparent once we have derived the following elegant formula.

THEOREM 4.4 (conjugacy). *Suppose that the function $h : Y \to [-\infty, +\infty]$ is $\mathcal{G}_Y$-invariant. Then on the space X,*

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

Proof. By Proposition 3.1 applied in turn to the systems $(X, \mathcal{G}, \gamma)$ and $(Y, \mathcal{G}|_Y, \gamma)$, we see that for any point w in X,

$$\begin{aligned}(h \circ \gamma)^*(w) &= \sup\{\langle x, \gamma(w)\rangle - h(\gamma(x)) \mid x \in R(\gamma)\} \\ &= \sup\{\langle x, \gamma(w)\rangle - h(x) \mid x \in R(\gamma)\} \\ &= h^*(\gamma(w)). \qquad \square\end{aligned}$$

It is an immediate consequence of this conjugacy formula that a $\mathcal{G}_Y$-invariant function $h : Y \to (-\infty, +\infty]$ (note that we exclude $-\infty$) is closed and convex exactly when the function $h \circ \gamma$ is closed and convex on X. One direction is clear from equation (4.1). On the other hand, if h is closed and convex then $h = h^{**}$, so that $h \circ \gamma = (h \circ \gamma)^{**}$ by Theorem 4.4, and hence $h \circ \gamma$ is also closed and convex.

Proposition 4.2 shows that the restriction operation which maps an extended real-valued function h on X to its restriction $h|_Y$ gives a one-to-one correspondence between $\mathcal{G}$-invariant functions on X and $\mathcal{G}_Y$-invariant functions on Y. Theorem 4.3 (convex and closed functions) shows that this correspondence preserves convexity and closedness, and Theorem 4.4 (conjugacy) shows that it also preserves the conjugacy operation. We shall see in §6 that restriction also preserves essential strict convexity and smoothness (Corollary 6.2).

The next result provides perhaps a more compelling motivation for the conjugacy approach. Recall that for a point x in X, the set $\mathcal{G}^x$ describes the possible decompositions of x: $\mathcal{G}^x = \{A \in \mathcal{G} \mid x = A\gamma(x)\}$.

THEOREM 4.5 (subdifferentials). *Given a function $h : Y \to (-\infty, +\infty]$ which is $\mathcal{G}_Y$-invariant, suppose that the point x in X satisfies $\gamma(x) \in \operatorname{dom}(h)$. Then the element w of X is a subgradient of the function $h \circ \gamma$ at x if and only if $\gamma(w)$ is a subgradient of h at $\gamma(x)$ with x and w having simultaneous decompositions: $\mathcal{G}^x \cap \mathcal{G}^w \neq \emptyset$. In fact, the following "chain rule" holds:*

$$\partial(h \circ \gamma)(x) = \mathcal{G}^x \partial h(\gamma(x)). \tag{4.2}$$

Proof. By definition, $w \in \partial(h \circ \gamma)(x)$ if and only if

$$\langle x, w\rangle = (h \circ \gamma)(x) + (h \circ \gamma)^*(w) = h(\gamma(x)) + h^*(\gamma(w)),$$

using Theorem 4.4 (conjugacy). But then, since

$$h(\gamma(x)) + h^*(\gamma(w)) \geq \langle \gamma(x), \gamma(w)\rangle \geq \langle x, w\rangle,$$

equality holds throughout, and the first part of the result follows using Theorem 2.2.

Suppose that $w \in \partial(h \circ \gamma)(x)$. Then by the above, $\gamma(w) \in \partial h(\gamma(x))$ and we can choose an operator A in $\mathcal{G}^x \cap \mathcal{G}^w$. Then

$$w = A\gamma(w) \in A\partial h(\gamma(x)) \subset \mathcal{G}^x \partial h(\gamma(x)).$$

Conversely, suppose that $y \in \partial h(\gamma(x))$ and that $A \in \mathcal{G}^x$. Then

$$\begin{aligned}(h \circ \gamma)(x) + (h \circ \gamma)^*(Ay) &= h(\gamma(x)) + h^*(\gamma(Ay)) \\ &= h(\gamma(x)) + h^*(\gamma(y)) \\ &= h(\gamma(x)) + h^*(y) \\ &= \langle \gamma(x), y\rangle = \langle x, Ay\rangle,\end{aligned}$$

using Theorem 4.4 (conjugacy) and the $\mathcal{G}_Y$-invariance of h^*. Thus Ay lies in $\partial(h \circ \gamma)(x)$, as required. □

Notice that this result is the first point at which we have used the condition for equality in property (c) of Definition 2.1.

COROLLARY 4.6. *Suppose that the function $f : X \to (-\infty, +\infty]$ is $\mathcal{G}$-invariant and that the point x lies in $\operatorname{dom} f$. Then the element w of X is a subgradient of f at x if and only if $\gamma(w)$ is a subgradient of f at $\gamma(x)$ with x and w having simultaneous decompositions: $\mathcal{G}^x \cap \mathcal{G}^w \neq \emptyset$. In fact, $\partial f(x) = \mathcal{G}^x \partial f(\gamma(x))$.*

Proof. Take $Y = X$ in Theorem 4.5 (subdifferentials). □

COROLLARY 4.7. *Suppose that the function $f : X \to (-\infty, +\infty]$ is $\mathcal{G}$-invariant and convex. If f is differentiable at the point x then $\nabla f(\gamma(x)) = \gamma(\nabla f(x))$.*

Proof. By Lemma 3.2, f is differentiable at $\gamma(x)$. By Corollary 4.6, since $\nabla f(x) \in \partial f(x)$ it follows that $\gamma(\nabla f(x)) \in \partial f(\gamma(x)) = \{\nabla f(\gamma(x))\}$. $\square$

Given functions $h, p : Y \to (-\infty, +\infty]$, we define the *infimal convolution* $h \square p : Y \to [-\infty, +\infty]$ by

$$(h \square p)(y) = \inf_{w \in Y} \{h(w) + p(y - w)\}.$$

An analogous definition holds on X.

THEOREM 4.8 (infimal convolution). *Suppose that the functions $h, p : Y \to (-\infty, +\infty]$ are $\mathcal{G}_Y$-invariant and convex. Then*

$$(h \square p) \circ \gamma = (h \circ \gamma) \square (p \circ \gamma).$$

Proof. Given any two points x and z in X, define two compact convex subsets of Y by

$$\begin{aligned} C &= \operatorname{conv} \mathcal{G}_Y \gamma(z), \quad \text{and} \\ D &= \gamma(x) - \operatorname{conv} \mathcal{G}_Y \gamma(x - z). \end{aligned}$$

These two sets are not disjoint, since a separating hyperplane would give an element u of Y and a scalar β with

$$\begin{aligned} \langle \gamma(u), \gamma(z) \rangle &= \max_{A \in \mathcal{G}_Y} \langle u, A\gamma(z) \rangle \\ &\leq \beta < \min_{A \in \mathcal{G}_Y} \langle u, \gamma(x) - A\gamma(x - z) \rangle \\ &= \langle u, \gamma(x) \rangle - \langle \gamma(u), \gamma(x - z) \rangle \\ &\leq \langle \gamma(u), \gamma(x) - \gamma(x - z) \rangle, \end{aligned}$$

which contradicts the convexity and positive homogeneity of $\langle \gamma(u), \gamma(\cdot) \rangle$ (Theorem 2.4). Thus there is a point w in $C \cap D$, and this point must satisfy $h(w) \leq h(\gamma(z))$ and $p(\gamma(x) - w) \leq p(\gamma(x - z))$.

Now consider any fixed point x in X. By the above argument we see that

$$\begin{aligned} (h \square p)(\gamma(x)) &= \inf_{w \in Y} \{h(w) + p(\gamma(x) - w)\} \\ &\leq \inf_{z \in X} \{h(\gamma(z)) + p(\gamma(x - z))\} \\ &= ((h \circ \gamma) \square (p \circ \gamma))(x). \end{aligned}$$

On the other hand, we can choose an operator A in $\mathcal{G}^x$, and then

$$\begin{aligned} ((h \circ \gamma) \square (p \circ \gamma))(x) &= \inf_{z \in X} \{h(\gamma(z)) + p(\gamma(x - z))\} \\ &\leq \inf_{w \in Y} \{h(\gamma(Aw)) + p(\gamma(A(\gamma(x) - w)))\} \\ &= \inf_{w \in Y} \{h(w) + p(\gamma(x) - w)\} \\ &= (h \square p)(\gamma(x)). \end{aligned}$$

The result follows. $\square$

This result strengthens and generalizes those in [33, 22].

5. Invariant sets. In the last section we studied $\mathcal{G}$-invariant functions on the space X. In this section we consider analogous questions for $\mathcal{G}$-invariant subsets of X. As usual, $(X, \mathcal{G}, \gamma)$ is a normal decomposition system with a subsystem $(Y, \mathcal{G}_Y, \gamma)$ (where Y contains the range of γ). In other words, Assumption 4.1 holds.

PROPOSITION 5.1. *A subset D of X is $\mathcal{G}$-invariant if and only if it has the form $D = \gamma^{-1}(C)$ for some $\mathcal{G}_Y$-invariant subset C of Y.*

Proof. Clearly any set of the form $\gamma^{-1}(C) = \{x \in X \mid \gamma(x) \in C\}$ is $\mathcal{G}$-invariant because γ is. On the other hand, if D is $\mathcal{G}$-invariant then it is easily checked that we can write $D = \gamma^{-1}(D \cap Y)$, which has the required form. □

Thus henceforth we will restrict our attention to $\mathcal{G}$-invariant sets $\gamma^{-1}(C)$ (or, equivalently, $\mathcal{G}C$), where the set $C \subset Y$ is $\mathcal{G}_Y$-invariant. Sets can be effectively studied via their indicator functions,

$$\delta_C(y) = \begin{cases} 0, & y \in C, \\ +\infty, & y \notin C. \end{cases}$$

Notice, for example, that $\delta_{\gamma^{-1}(C)} = \delta_C \circ \gamma$ for any subset C of Y.

COROLLARY 5.2 (closed and convex sets). *Suppose that the subset C of Y is $\mathcal{G}_Y$-invariant. Then the set $\gamma^{-1}(C)$ is closed (respectively, convex) if and only if C is closed (respectively, convex).*

Proof. Apply Theorem 4.3 to the function δ_C. □

A fundamental idea in optimization is the *(convex) normal cone* to a subset C of Y at a point y in C, defined by

$$N(y|C) = \{w \in Y \mid \langle w, z - y\rangle \leq 0 \text{ for all } z \in C\}.$$

It is easily checked that $N(y|C) = \partial\delta_C(y)$, whence the following useful formula.

COROLLARY 5.3 (normal cones). *Suppose that the subset C of Y is $\mathcal{G}_Y$-invariant and that the point x in X satisfies $\gamma(x) \in C$. Then the element w of X lies in the normal cone $N(x|\gamma^{-1}(C))$ if and only if $\gamma(w)$ lies in $N(\gamma(x)|C)$ with x and w having simultaneous decompositions ($\mathcal{G}^x \cap \mathcal{G}^w \neq \emptyset$). In fact,*

$$N(x|\gamma^{-1}(C)) = \mathcal{G}^x N(\gamma(x)|C).$$

Proof. Apply Theorem 4.5 (subdifferentials) to the function $h = \delta_C$. □

Other convex-analytic formulae follow easily from Theorem 4.4 (conjugacy). For convenience, we collect some similar-looking results in a single theorem. The *polar set* of a subset C of Y is defined by

$$C^\circ = \{z \in Y \mid \langle z, y\rangle \leq 1 \text{ for all } y \in C\},$$

while the *polar cone* is

$$C^- = \{z \in Y \mid \langle z, y\rangle \leq 0 \text{ for all } y \in C\}.$$

THEOREM 5.4. *Suppose that the subset C of Y is $\mathcal{G}_Y$-invariant. Then*
(i) $(\gamma^{-1}(C))^- = \gamma^{-1}(C^-)$,
(ii) $(\gamma^{-1}(C))^\circ = \gamma^{-1}(C^\circ)$, *and*
(iii) $\operatorname{int}_X \gamma^{-1}(C) = \gamma^{-1}(\operatorname{int}_Y C)$.
Furthermore, $C = \gamma^{-1}(C) \cap Y$, and if C is also convex then
(iv) $\operatorname{ri} \gamma^{-1}(C) = \gamma^{-1}(\operatorname{ri} C)$, *and*

(v) $\operatorname{aff}\gamma^{-1}(C) = \gamma^{-1}(\operatorname{aff} C)$.

Proof. An element w of X lies in $(\gamma^{-1}(C))^-$ if and only if

$$0 \geq \delta^*_{\gamma^{-1}(C)}(w) = (\delta_C \circ \gamma)^*(w) = \delta^*_C(\gamma(w)),$$

whence (i), and (ii) is similar.

To see (iii), note that since γ may be regarded as a map from X to Y and is continuous by Theorem 2.4, $\gamma^{-1}(\operatorname{int}_Y C)$ is an open subset of $\gamma^{-1}(C)$, and hence $\gamma^{-1}(\operatorname{int}_Y C) \subset \operatorname{int}_X \gamma^{-1}(C)$. Conversely, suppose that the point x lies in $\operatorname{int}_X \gamma^{-1}(C)$, and yet $\gamma(x) \notin \operatorname{int}_Y(C)$. Then there is a sequence of points (y_n) in $Y \setminus C$ approaching $\gamma(x)$. Each point has a decomposition $y_n = A_n \gamma(y_n)$ for some operator A_n in $\mathcal{G}_Y$, and since $\mathcal{G}_Y$ is compact there is a convergent subsequence $A_{n'} \to A \in \mathcal{G}_Y$. Now notice that the sequence $\gamma(y_{n'}) = A^*_{n'} y_{n'}$ approaches $A^*\gamma(x)$. Since $\gamma^{-1}(C)$ is $\mathcal{G}$-invariant, so is $\operatorname{int}_X \gamma^{-1}(C)$, and hence since x lies in $\operatorname{int}_X \gamma^{-1}(C)$, so does $A^*\gamma(x)$. Thus for sufficiently large n' we have $\gamma(y_{n'}) \in \gamma^{-1}(C)$; whence $\gamma(y_{n'}) \in C$. Now since C is $\mathcal{G}_Y$-invariant, $y_{n'} \in C$, which is a contradiction.

For any point y in C, y lies in Y and there is an operator A in $\mathcal{G}_Y$ with $y = A\gamma(y)$. Since C is $\mathcal{G}_Y$-invariant it follows that $\gamma(y) \in C$, so that $y \in \gamma^{-1}(C) \cap Y$. Conversely, if $y \in \gamma^{-1}(C) \cap Y$ then again there exists A in $\mathcal{G}_Y$ with $y = A\gamma(y)$; whence $y \in C$ since C is $\mathcal{G}_Y$-invariant. Thus $C = \gamma^{-1}(C) \cap Y$.

Now suppose that C is convex and, without loss of generality, nonempty. By Corollary 5.2, $\gamma^{-1}(C)$ is a nonempty, $\mathcal{G}$-invariant, convex set, so there exists a point x in $\operatorname{ri}\gamma^{-1}(C)$. Since $\operatorname{ri}\gamma^{-1}(C)$ is $\mathcal{G}$-invariant, $\gamma(x)$ lies in $Y \cap \operatorname{ri}\gamma^{-1}(C)$. Hence the relative interiors of the convex sets $\gamma^{-1}(C)$ and Y intersect, so $\operatorname{ri} C = \operatorname{ri}(Y \cap \gamma^{-1}(C)) = Y \cap \operatorname{ri}\gamma^{-1}(C)$, by [32, Cor. 6.5.1], and it is elementary to check that $\operatorname{aff}(Y \cap \gamma^{-1}(C)) = Y \cap \operatorname{aff}\gamma^{-1}(C)$. Now, a point z belongs to $\operatorname{ri}\gamma^{-1}(C)$ if and only if $\gamma(z) \in Y \cap \operatorname{ri}\gamma^{-1}(C) = \operatorname{ri} C$, by the $\mathcal{G}$-invariance of $\gamma^{-1}(C)$, and (iv) follows. Equation (v) is similar. □

The pattern of these results is clear. *If the convex subset C of Y is $\mathcal{G}_Y$-invariant then for many set operations '#' the following metaformula holds*:

$$\boxed{\#(\gamma^{-1}(C)) = \gamma^{-1}(\#(C)).} \tag{5.1}$$

The utility of this formula lies in expressing the result of an operation in the larger space X on a complicated set, $\gamma^{-1}(C)$, in terms of the result of the same operation in a smaller space Y on the simpler set C. A straightforward deduction (in light of Theorem 5.4) is that if the convex subset D of X is $\mathcal{G}$-invariant then for many set operations '#' the following metaformula holds:

$$\boxed{Y \cap \#D = \#(Y \cap D).} \tag{5.2}$$

We will see another example of this pattern in the next section—we will show that $\exp(\gamma^{-1}(C)) = \gamma^{-1}(\exp C)$ for a closed, convex set C, where $\exp C$ denotes the set of exposed points of C. To end this section we prove the analogous result for the set of *extreme points* of C, denoted $\operatorname{ext} C$, which are those points x in C for which $C \setminus \{x\}$ is convex.

THEOREM 5.5 (extreme points). *If the subset C of Y is convex and $\mathcal{G}_Y$-invariant then*

$$\operatorname{ext}(\gamma^{-1}(C)) = \gamma^{-1}(\operatorname{ext} C).$$

Proof. Suppose first that the point x in X does not belong to $\gamma^{-1}(\operatorname{ext} C)$, so that $\gamma(x) \notin \operatorname{ext} C$. If $\gamma(x) \notin C$ then clearly $x \notin \operatorname{ext}(\gamma^{-1}(C))$, so suppose that for some points u and v in C distinct from $\gamma(x)$ and some scalar α in $(0,1)$ we have $\gamma(x) = \alpha u + (1-\alpha)v$. For any operator A in $\mathcal{G}^x$ we now have $x = A\gamma(x) = \alpha Au + (1-\alpha)Av$, and since the points Au and Av are distinct from x in the set $\gamma^{-1}(C)$, it follows that x is not extreme in this set.

On the other hand, suppose that $\gamma(x)$ is extreme in C and yet x is not extreme in $\gamma^{-1}(C)$—we will derive a contradiction. Pick points u_1 and v_1 distinct from x in $\gamma^{-1}(C)$ and a scalar α_1 in $(0,1)$ with $x = \alpha_1 u_1 + (1-\alpha_1)v_1$. Now define a new point

$$u = \begin{cases} \frac{1}{2}(x+u_1) & \text{if } \gamma(x) = \gamma(u_1), \\ u_1 & \text{otherwise.} \end{cases}$$

Since $\gamma^{-1}(C)$ is convex, u lies in $\gamma^{-1}(C)$, and if $\gamma(x) = \gamma(u_1)$ then $\|x\| = \|u_1\|$ by Theorem 2.4 with

$$\|\gamma(u)\| = \|u\| < \frac{\|x\| + \|u_1\|}{2} = \|x\| = \|\gamma(x)\|.$$

Hence in either case $\gamma(u) \neq \gamma(x)$. By defining a point v in an analogous fashion we arrive at a representation $x = \alpha u + (1-\alpha)v$ for a scalar α in $(0,1)$ with $\gamma(u)$ and $\gamma(v)$ distinct from $\gamma(x)$ in C.

Now certainly $\gamma(x)$ does not belong to either of the cosets $\mathcal{G}_Y\gamma(u)$ or $\mathcal{G}_Y\gamma(v)$. For example, if $\gamma(x) = A\gamma(u)$ for some operator A in $\mathcal{G}_Y$ then applying γ gives a contradiction. Thus, since $\gamma(x)$ is extreme,

$$\gamma(x) \notin \operatorname{conv}(\mathcal{G}_Y\gamma(u) \cup \mathcal{G}_Y\gamma(v)).$$

Since the set on the right-hand side is compact, we can choose an element y of Y (defining a separating hyperplane) so that

$$\begin{aligned} \langle \gamma(y), \gamma(x)\rangle \geq \langle y, \gamma(x)\rangle &> \max \langle y, \mathcal{G}_Y\gamma(u) \cup \mathcal{G}_Y\gamma(v)\rangle \\ &= \max\{\langle \gamma(y), \gamma(u)\rangle, \langle \gamma(y), \gamma(v)\rangle\}, \end{aligned}$$

using Proposition 2.3. But this contradicts the fact that $x = \alpha u + (1-\alpha)v$ and the function $\langle \gamma(y), \gamma(\cdot)\rangle$ is convex (Theorem 2.4). □

6. Smoothness, strict convexity, and invariant norms. Our aim in this section is to investigate the dual concepts of smoothness and strict convexity for $\mathcal{G}$-invariant convex functions. Once again, we assume throughout that $(X, \mathcal{G}, \gamma)$ is a normal decomposition system, and that Assumption 4.1 holds, which is to say that $(Y, \mathcal{G}_Y, \gamma)$ is a subsystem where the space Y contains the range of γ.

The first result shows that a $\mathcal{G}$-invariant convex function $h \circ \gamma$ (where the function h is $\mathcal{G}_Y$-invariant—see Proposition 4.2) is differentiable at a point x in X if and only if h is differentiable at $\gamma(x)$.

THEOREM 6.1 (differentiability). *Let the function $h : Y \to (-\infty, +\infty]$ be $\mathcal{G}_Y$-invariant. If $h \circ \gamma$ is differentiable at a point x in X then h is differentiable at $\gamma(x)$, and the following chain rule holds:*

$$(6.1) \qquad \nabla(h \circ \gamma)(x) = A\nabla h(\gamma(x)) \quad \text{for any operator } A \in \mathcal{G}^x.$$

Conversely, if h is in addition convex, and differentiable at $\gamma(x)$, then $h \circ \gamma$ is differentiable at x and, furthermore, $\gamma(\nabla(h \circ \gamma)(x)) = \nabla h(\gamma(x))$.

Proof. For any operator A in $\mathcal{G}^x$ we have $x = A\gamma(x)$, and for all points y in Y

$$(h \circ \gamma)(Ay) = h(\gamma(Ay)) = h(\gamma(y)) = h(y),$$

since h is $\mathcal{G}_Y$-invariant. The left-hand side is differentiable at $y = \gamma(x)$, by the chain rule, hence so is the right-hand side with $\nabla h(\gamma(x)) = A^*\nabla(h \circ \gamma)(x)$. The first part of the result follows.

On the other hand, if h is also convex, and differentiable at $\gamma(x)$, then $\partial h(\gamma(x)) = \{\nabla h(\gamma(x))\}$ by [32, Thm. 25.1]. Now by Theorem 4.5 (subdifferentials), if an element w of X belongs to $\partial(h \circ \gamma)(x)$ then $\gamma(w) \in \partial h(\gamma(x))$, and so $\gamma(w) = \nabla h(\gamma(x))$. In particular, since γ is norm preserving (Theorem 2.4), any such subgradient has norm $\|\nabla h(\gamma(x))\|$. Since $\partial(h \circ \gamma)(x)$ is a convex set and $\|\cdot\|$ is a strict norm, $\partial(h \circ \gamma)(x)$ has at most one element. However, it is nonempty by the chain rule (4.2). Thus it is a singleton, whence $h \circ \gamma$ is differentiable at x by [32, Thm. 25.1], and the result follows. □

We say that a proper, closed, convex function $h : Y \to (-\infty, +\infty]$ is *essentially smooth* if it is differentiable at any point where it has a subgradient, and is *essentially strictly convex* if it is strictly convex on any convex set on which the subdifferential is everywhere nonempty. These two concepts are dual to each other: h is essentially smooth if and only if its conjugate is essentially strictly convex and vice versa [32, Thm. 26.3].

COROLLARY 6.2 (essential smoothness and strict convexity). *Suppose that the function $h : Y \to (-\infty, +\infty]$ is $\mathcal{G}_Y$-invariant, closed, proper, and convex. Then the function $h \circ \gamma$ is essentially smooth (respectively, essentially strictly convex) if and only if h is essentially smooth (respectively, essentially strictly convex).*

Proof. Suppose first that $h \circ \gamma$ is essentially smooth. If h has a subgradient $v \in Y$ at the point $y \in Y$ then by Corollary 4.6 we have $\gamma(v) \in \partial h(\gamma(y))$. Since the identity operator lies in $\mathcal{G}^{\gamma(y)}$ it follows from the subdifferential formula (4.2) that $\gamma(v) \in \partial(h \circ \gamma)(\gamma(y))$. Thus because $h \circ \gamma$ is essentially smooth, it is differentiable at $\gamma(y)$, and hence by Theorem 6.1 (differentiability), h is differentiable at $\gamma(y)$, and therefore also at y by Lemma 3.2. Thus h must be essentially smooth.

Conversely, suppose that h is essentially smooth. If $h \circ \gamma$ has a subgradient at a point x in X then the subdifferential formula (4.2) implies that $\partial h(\gamma(x))$ is nonempty. Hence h is differentiable at $\gamma(x)$, and therefore $h \circ \gamma$ is differentiable at x by Theorem 6.1 (differentiability). Thus $h \circ \gamma$ is essentially smooth.

The essentially strictly convex case follows by taking conjugates. □

The following result is another example of the pattern (5.1) that we observed in the last section: $\#(\gamma^{-1}(C)) = \gamma^{-1}(\#(C))$. If the subset C of Y is closed and convex then we say that a point y in C is *exposed* if there is an element z of Y with $\langle z, y\rangle > \langle z, u\rangle$ for all points u in $C \setminus \{y\}$. Equivalently, a point y in C is exposed if and only if it lies in the range of $\nabla\delta_C^*$ [32, Cor. 25.1.3]. We denote the set of exposed points by $\exp(C)$. A generalization of this result to exposed *faces* appears in [21].

COROLLARY 6.3 (exposed points). *Suppose that the subset C of Y is $\mathcal{G}_Y$-invariant, closed, and convex. Then*

$$\exp(\gamma^{-1}(C)) = \gamma^{-1}(\exp(C)).$$

Proof. If the point x in X satisfies $\gamma(x) \in \exp(C)$ then for some element v of Y we have $\gamma(x) = \nabla\delta_C^*(v)$, and by Corollary 4.7 it follows that $\gamma(x) = \nabla\delta_C^*(\gamma(v))$. Notice that $\delta_{\gamma^{-1}(C)} = \delta_C \circ \gamma$, and hence by Theorem 4.4 (conjugacy) we have $\delta^*_{\gamma^{-1}(C)} = \delta^*_C \circ \gamma$. Choose an operator A in $\mathcal{G}^x$, so that $x = A\gamma(x)$, and observe that $A \in \mathcal{G}^{A\gamma(v)}$. Thus, applying the chain rule (6.1),

$$\nabla \delta^*_{\gamma^{-1}(C)}(A\gamma(v)) = \nabla(\delta^*_C \circ \gamma)(A\gamma(v)) = A\nabla\delta^*_C(\gamma(v)) = A\gamma(x) = x,$$

so that $x \in \exp(\gamma^{-1}(C))$.

Conversely, if $x \in \exp(\gamma^{-1}(C))$ then for some element w of X we have $x = \nabla\delta^*_{\gamma^{-1}(C)}(w) = \nabla(\delta^*_C \circ \gamma)(w)$. It follows by Theorem 6.1 (differentiability) that $\gamma(x) = \nabla\delta^*_C(\gamma(w))$, whence $\gamma(x) \in \exp(C)$. $\square$

To end this section we examine our results for the special case of invariant norms. If p is a norm on Y then we denote the *dual norm* on Y by p^D, where for an element z of Y,

$$p^D(z) = \max\{\langle y, z\rangle \mid y \in Y,\ p(y) = 1\}.$$

We relate the dualizing operation for norms with conjugacy by the following standard and straightforward trick.

LEMMA 6.4. *If p is a norm on Y then $(\frac{1}{2}p(\cdot)^2)^* = \frac{1}{2}(p^D(\cdot))^2$.*

A norm p on Y is *smooth* if it is differentiable except at the origin. Equivalently, the proper, closed, convex function $p^2/2$ is essentially smooth. Furthermore, p is *strict* if $p(u+v) < 2$ for all distinct points u and v in the unit ball for p, namely, $\{y \in Y \mid p(y) \le 1\}$. Equivalently, $p^2/2$ is essentially strictly convex. A point y in Y is a *smooth point* of the unit ball if $p(y) = 1$ and p is differentiable at y.

THEOREM 6.5 (norms). *The $\mathcal{G}$-invariant norms on X are those functions of the form $p \circ \gamma$, where p is a $\mathcal{G}_Y$-invariant norm on Y. The dual of such a norm is $p^D \circ \gamma$. The norm $p \circ \gamma$ is smooth (respectively, strict) if and only if p is smooth (respectively, strict). A point x in X is an extreme (respectively, exposed, smooth) point of the unit ball for $p \circ \gamma$ if and only if $\gamma(x)$ is an extreme (respectively, exposed, smooth) point of the unit ball for p.*

Proof. By Proposition 4.2, the $\mathcal{G}$-invariant functions on X are those of the form $p \circ \gamma$ with p a $\mathcal{G}_Y$-invariant function on Y. If $p \circ \gamma$ is actually a norm on X then p is a norm on Y, since by $\mathcal{G}_Y$-invariance, p agrees with $p \circ \gamma$ on Y. Conversely, suppose that p is a $\mathcal{G}_Y$-invariant norm. Then certainly $(p \circ \gamma)(x) = p(\gamma(x)) \ge 0$ for all points x in X with equality if and only if $\gamma(x) = 0$ or, equivalently, $x = 0$. Positive homogeneity of $p \circ \gamma$ follows from that of γ (Theorem 2.4). Finally, $p \circ \gamma$ is convex by Theorem 4.3, and hence is a norm.

By Lemma 6.4 we have

$$\begin{aligned}((p \circ \gamma)^D)^2/2 = ((p \circ \gamma)^2/2)^* &= (p^2/2 \circ \gamma)^* \\ &= (p^2/2)^* \circ \gamma = (p^D)^2/2 \circ \gamma = (p^D \circ \gamma)^2/2,\end{aligned}$$

using Theorem 4.4 (conjugacy). Hence $(p \circ \gamma)^D = p^D \circ \gamma$. The norm $p \circ \gamma$ is smooth if and only if $(p \circ \gamma)^2/2 = p^2/2 \circ \gamma$ is essentially smooth, which by Corollary 6.2 is equivalent to the essential smoothness of $p^2/2$, and hence to the smoothness of p. The strict case is analogous.

The last statement follows by applying Theorem 5.5 (extreme points) and Corollary 6.3 (exposed points) to the unit ball for p, and by applying Theorem 6.1 (differentiability) to p. $\square$

7. Examples. The idea of a normal decomposition system that we introduced in Definition 2.1 works well as an abstract mechanism. Its real significance, however, is in the variety of examples that it models. In this section we discuss these examples. They fall into two distinct categories: "discrete" examples, where the group $\mathcal{G}$ is a reflection group (in fact a "Weyl group") and the range of the map γ has full dimension in the underlying inner product space X, and "continuous" examples, where γ maps X into a strictly smaller space Y. Both categories are important for our purposes. Further discussion of the role of Weyl groups in this construction may be found in [22].

First we explain some notation for various sets of matrices. The trace of a matrix w is denoted by $\mathrm{tr}\,(w)$ and the Hermitian conjugate by w^*.

$\mathcal{O}_n$: The (multiplicative) group of $n \times n$ real orthogonal matrices.
$\mathcal{U}_n$: The (multiplicative) group of $n \times n$ complex unitary matrices.
$\mathcal{P}_n$: The (multiplicative) group of $n \times n$ permutation matrices.
$\mathcal{P}_n^{\pm}$: The (multiplicative) group of $n \times n$ "signed" permutation matrices (having exactly one nonzero entry, ± 1, in each row and each column.
S_n: The inner product space of $n \times n$ real symmetric matrices with $\langle w, v\rangle = \mathrm{tr}\,(wv)$.
H_n: The (*real*) inner product space of $n \times n$ complex Hermitian matrices with $\langle w, v\rangle = \mathrm{tr}\,(wv)$.
$M_{m,n}(\mathbf{R})$: The inner product space of $m \times n$ real matrices with $\langle w, v\rangle = \mathrm{tr}\,(w^T v)$.
$M_{m,n}(\mathbf{C})$: The (*real*) inner product space of $m \times n$ complex matrices with $\langle w, v\rangle = \mathrm{Re}\,\mathrm{tr}\,(w^* v)$.

For a matrix w in S_n or H_n, the vector $\lambda(w) \in \mathbf{R}^n$ has components the eigenvalues of w, arranged in nonincreasing order. For a matrix w in $M_{m,n}(\mathbf{R})$ or $M_{m,n}(\mathbf{C})$, the vector $\sigma(w) \in \mathbf{R}^l_+$ (where $l = \min\{m, n\}$) has components the singular values of w, arranged in nonincreasing order.

Recall that a normal decomposition system consists of a real inner product space X, a subgroup $\mathcal{G}$ of the orthogonal group on X, $O(X)$, and a map $\gamma : X \to X$ satisfying Definition 2.1.

EXAMPLE 7.1 (reordering on $\mathbf{R}^n$). We take $X = \mathbf{R}^n$ (with the standard inner product), $\mathcal{G} = \mathcal{P}_n$ (considered as a subgroup of $O(\mathbf{R}^n) = \mathcal{O}_n$ in the natural way), and $\gamma(x) = \bar{x}$, where the vector $\bar{x} \in \mathbf{R}^n$ has components $\{x_1, x_2, \ldots, x_n\}$ arranged in nonincreasing order. The conditions in Definition 2.1 are immediate except for (c), which states that

$$\langle x, z\rangle \leq \langle \bar{x}, \bar{z}\rangle \quad \textit{for all } x \textit{ and } z \textit{ in } \mathbf{R}^n \tag{7.1}$$

(with equality if and only if $x = A\bar{x}$ and $z = A\bar{z}$ for some permutation matrix A). Inequality (7.1) is classical; see, for example, [14, Thm. 368] and [18, Lem. 2.1].

The range of γ is $\mathbf{R}^n_{\geq} = \{x \in \mathbf{R}^n \mid (x_i) \text{ nonincreasing}\}$. The dual cone is straightforward to compute. In fact, a vector z lies in $(\mathbf{R}^n_{\geq})^+$ if and only if

$$\sum_1^j z_i \geq 0 \quad \text{for } j = 1, 2, \ldots, n$$

with equality for $j = n$. We say that a real function f on $\mathbf{R}^n_{\geq}$ is *Schur convex* if $f(x) \geq f(w)$ whenever x and w lie in $\mathbf{R}^n_{\geq}$ with $x - w$ in $(\mathbf{R}^n_{\geq})^{+}$. Theorem 3.3 now shows that any symmetric, convex function is Schur convex [23, Prop. 3.C.2].

EXAMPLE 7.2 (absolute reordering on $\mathbf{R}^n$). We take $X = \mathbf{R}^n$, $\mathcal{G} = \mathcal{P}_n^{\pm}$, and $\gamma(x) = \overline{|x|}$ (where $|x| = (|x_1|, |x_2|, \ldots, |x_n|)$). Thus

$$(\overline{|x|})_1 = \max\{|x_1|, |x_2|, \ldots, |x_n|\},$$

and so forth. The conditions in Definition 2.1 are easy to check: (c) follows from inequality (7.1).

Diagonal matrices will play an important role in our continuous examples. We denote the smaller of the two dimensions m and n by $l = \min\{m, n\}$, and then we define a map Diag : $\mathbf{R}^l \to M_{m,n}(\mathbf{C})$ by

$$(\operatorname{Diag}\alpha)_{ij} = \begin{cases} \alpha_i & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

EXAMPLE 7.3 (symmetric matrices). We take $X = S_n$ and $\mathcal{G}$ to be the group of orthogonal similarity transformations $x \mapsto u^T x u$ for symmetric matrices x and orthogonal matrices u. Finally, we define $\gamma(x) = \operatorname{Diag}\lambda(x)$.

More formally, define the *adjoint representation* of $\mathcal{O}_n$ on S_n, which we write $\operatorname{Ad} : \mathcal{O}_n \to O(S_n)$, by $(\operatorname{Ad}(u))x = u^T x u$ for orthogonal u and symmetric x. Then $\mathcal{G}$ is just the range of this representation, which has kernel $\{\pm\mathrm{id}\}$, and so $\mathcal{G}$ is isomorphic to $\mathcal{O}_n/\{\pm\mathrm{id}\}$.

Let us check the conditions of Definition 2.1. Condition (a), the $\mathcal{G}$-invariance of γ, amounts to the invariance of the set of eigenvalues under orthogonal similarity. Condition (b), the decomposition axiom, follows from the spectral decomposition. Condition (c), the angle contraction axiom, becomes the following inequality:

$$\operatorname{tr}(wx) \leq \langle \lambda(w), \lambda(x) \rangle \quad \text{for all } w, x \in S_n$$

(with equality if and only if there exists an orthogonal matrix u satisfying $x = u^T(\operatorname{Diag}\lambda(x))u$ and $w = u^T(\operatorname{Diag}\lambda(w))u$). The inequality appears in [25], for example, and the conditions for equality may be found in [35], using algebraic techniques. A variational proof is given in [18]. The result is closely connected with earlier work of von Neumann—see Example 7.5.

The natural choice for the subspace Y is the space of diagonal matrices $\operatorname{Diag}\mathbf{R}^n$. A standard calculation shows that an orthogonal u has $\operatorname{Ad}(u)$ in the stabilizer $\mathcal{G}_Y$ if and only if $u \in \mathcal{P}_n^{\pm}$. However, since

$$(\operatorname{Ad}(\operatorname{Diag}(\pm 1, \pm 1, \ldots, \pm 1)))|_Y = \mathrm{id},$$

we see that the group $\mathcal{G}_Y$ acting on the space Y of diagonal matrices is simply the permutation group $\mathcal{P}_n$ acting on the diagonal entries.

Notice also that for any vector α in $\mathbf{R}^n$ we have $\gamma(\operatorname{Diag}\alpha) = \operatorname{Diag}\bar{\alpha}$. Hence the subsystem $(Y, \mathcal{G}_Y, \gamma)$ is a normal decomposition system isomorphic to the "reordering" system described in Example 7.1. In particular, Assumption 4.1 holds, so that all of the machinery that we have developed can be applied. We list some consequences in the final section.

EXAMPLE 7.4 (Hermitian matrices). The complex analogue of the previous example is very similar (and there is a quaternionic analogue). We take $X = H_n$, which we consider a *real* inner product space (since we are primarily concerned with properties of real vector spaces, such as convexity). The group $\mathcal{G}$ now consists of unitary similarity transformations $x \mapsto u^*xu$ for Hermitian x and unitary u and, as before $\gamma(x) = \operatorname{Diag}\lambda(x)$.

Formally, we define the adjoint representation of $\mathcal{U}_n$ on H_n, written $\operatorname{Ad} : \mathcal{U}_n \to O(H_n)$, by $(\operatorname{Ad}(u))(x) = u^*xu$ for unitary u and Hermitian x. Then $\mathcal{G}$ is just the range of this representation, which has kernel $\mathbf{T}\mathrm{id}$, where $\mathbf{T}$ is the circle group $\{\tau \in \mathbf{C} \mid |\tau| = 1\}$. Thus $\mathcal{G}$ is isomorphic to $\mathcal{U}_n/\mathbf{T}\mathrm{id}$. Checking Definition 2.1 is entirely analogous to the previous example.

An aside is illustrative at this point. If we choose the subspace Y as S_n then the stabilizer $\mathcal{G}_Y$ acts on Y exactly as $\operatorname{Ad}\mathcal{O}_n$. Thus in this case the subsystem $(Y, \mathcal{G}_Y, \gamma)$ is a normal decomposition system isomorphic to the previous "symmetric matrix" Example 7.3. In particular, Assumption 4.1 holds.

The natural choice, however, is again to choose Y as the subspace of diagonal matrices $\operatorname{Diag}\mathbf{R}^n$. Then it is once again straightforward to identify the action of the stabilizer $\mathcal{G}_Y$ on this subspace with the permutation group $\mathcal{P}_n$ acting on the diagonal entries. Thus the subsystem $(Y, \mathcal{G}_Y, \gamma)$ is a normal decomposition system isomorphic to the "reordering" system, Example 7.1. Again Assumption 4.1 holds, so our machinery applies.

EXAMPLE 7.5 (real matrices). We take $X = M_{m,n}(\mathbf{R})$ and $\mathcal{G}$ to be the group of transformations $x \mapsto u^T xv$ for orthogonal matrices u in $\mathcal{O}_m$ and v in $\mathcal{O}_n$. Then we define $\gamma(x) = \operatorname{Diag}\sigma(x)$.

Formally, we define a representation of $\mathcal{O}_m \times \mathcal{O}_n$ on $M_{m,n}(\mathbf{R})$, written $\operatorname{Ac} : \mathcal{O}_m \times \mathcal{O}_n \to O(M_{m,n}(\mathbf{R}))$, by $(\operatorname{Ac}(u, v))x = u^T xv$. Then $\mathcal{G}$ is the range of this representation: since the kernel of Ac is just $\{\pm(\mathrm{id}, \mathrm{id})\}$, the group $\mathcal{G}$ is isomorphic to $(\mathcal{O}_m \times \mathcal{O}_n)/\{\pm(\mathrm{id}, \mathrm{id})\}$.

Checking Definition 2.1, $\mathcal{G}$-invariance amounts to the invariance of the set of singular values under the transformations we consider. Condition (b), the decomposition axiom, follows from the singular value decomposition, and condition (c), the angle contraction axiom, becomes "von Neumann's lemma" [27]:

$$\operatorname{tr}(w^T x) \le \langle \sigma(w), \sigma(x) \rangle, \quad \text{for all } w, x \in M_{m,n}(\mathbf{R}) \tag{7.2}$$

(with equality if and only if w and x have simultaneous singular value decompositions $w = u^T(\operatorname{Diag}\sigma(w))v$ and $x = u^T(\operatorname{Diag}\sigma(x))v$ for some u in $\mathcal{O}_m$ and v in $\mathcal{O}_n$)—see the discussion in [7].

The natural choice for Y is the space of diagonal matrices $\operatorname{Diag}\mathbf{R}^l$ (where $l = \min\{m, n\}$). A little thought then identifies the action of the stabilizer $\mathcal{G}_Y$ on the space Y with the group of transformations $\operatorname{Diag}(\alpha) \mapsto \operatorname{Diag}(p\alpha)$ for a vector α in $\mathbf{R}^l$ and a matrix p in $\mathcal{P}_l^{\pm}$. To see this, note that any such transformation clearly belongs to $\mathcal{G}_Y$, whereas on the other hand a transformation in $\mathcal{G}_Y$ must preserve diagonality and the singular values.

Notice also that $\gamma(\operatorname{Diag}\alpha) = \overline{|\alpha|}$ for any vector α in $\mathbf{R}^l$. Thus the subsystem $(Y, \mathcal{G}_Y, \gamma)$ is a normal decomposition system isomorphic to the "absolute reordering" system described in Example 7.2. Since Assumption 4.1 holds, our machinery applies. Some consequences appear in the final section.

Two special cases deserve mention. The case $m = 1$ gives exactly the example discussed after Assumption 4.1. The even more special case $m = n = 1$ gives our very first example of a normal decomposition system, discussed after Definition 2.1.

EXAMPLE 7.6 (complex matrices). The complex analogue of the previous example is very similar (and again there is a quaternionic analogue). We take $X = M_{m,n}(\mathbf{C})$

and $\mathcal{G}$ to be the group of transformations $x \mapsto u^*xv$ for a matrix x in $M_{m,n}(\mathbf{C})$ and unitary matrices u in $\mathcal{U}_m$ and v in $\mathcal{U}_n$. Once again we define $\gamma(x) = \operatorname{Diag}\sigma(x)$.

Formally, we define a representation of $\mathcal{U}_m \times \mathcal{U}_n$ on $M_{m,n}(\mathbf{C})$, written $\mathrm{Ac} : \mathcal{U}_m \times \mathcal{U}_n \to O(M_{m,n}(\mathbf{C}))$, by $(\mathrm{Ac}(u,v))x = u^*xv$. Then $\mathcal{G}$ is the range of this representation. Since the kernel of Ac is easily checked to be $\mathbf{T}(\mathrm{id}, \mathrm{id})$ (where $\mathbf{T}$ is once again the circle group), we see that the group $\mathcal{G}$ is isomorphic to $(\mathcal{U}_m \times \mathcal{U}_n)/\mathbf{T}\mathrm{id}$. Checking Definition 2.1 is analogous to the previous example. In fact, if we choose $Y = M_{m,n}(\mathbf{R})$ then the subsystem $(Y, \mathcal{G}_Y, \gamma)$ is isomorphic to the previous example.

If we make the natural choice for Y, namely, the space of real diagonal matrices $\operatorname{Diag}\mathbf{R}^l$, then a similar argument to the previous example identifies the action of the stabilizer $\mathcal{G}_Y$ on the space Y as the group of transformations $\operatorname{Diag}(\alpha) \mapsto \operatorname{Diag}(p\alpha)$ for a vector α in $\mathbf{R}^l$ and a matrix p in $\mathcal{P}_l^{\pm}$. Thus just as in the previous example, the subsystem $(Y, \mathcal{G}_Y, \gamma)$ is isomorphic to the "absolute reordering" system, Example 7.2. Again, all our machinery applies.

8. Consequences for matrix functions. In this concluding section we consider how our results can be applied to the examples in the previous section to derive a variety of interesting results in the literature. We begin with the case of symmetric matrices, Example 7.3. The complex analogue is entirely similar, and we do not pursue it.

Symmetric matrices. A function $h : \mathbf{R}^n \to [-\infty, +\infty]$ is *symmetric* if for any vector α in $\mathbf{R}^n$ the value $h(\alpha)$ is unchanged by permuting the components of α—using the notation of Example 7.1, $h(\alpha) = h(\bar{\alpha})$. Similarly, a subset C of $\mathbf{R}^n$ is *symmetric* when $\alpha \in C$ if and only if $\bar{\alpha} \in C$. A function on the space of symmetric matrices $f : S_n \to [-\infty, +\infty]$ is *weakly orthogonally invariant* if $f(u^Txu) = f(x)$ for any matrices x in S_n and u in $\mathcal{O}_n$. Such functions have also been called *spectral* [13]. Analogously, a subset D of S_n is *weakly orthogonally invariant* if $u^Txu \in D$ whenever $x \in D$ (for orthogonal u).

The following result follows immediately by applying our machinery to Example 7.3. We make no attempt to be exhaustive.

THEOREM 8.1 (convex spectral functions). *Weakly orthogonally invariant extended real-valued functions on S_n are exactly those functions of the form $h \circ \lambda$ for a symmetric function $h : \mathbf{R}^n \to (-\infty, +\infty]$. Such a function on S_n is convex (respectively, closed, essentially strictly convex, essentially smooth) if and only if h is convex (respectively, closed, essentially strictly convex, essentially smooth). For any such symmetric function h we have*

$$(h \circ \lambda)^* = h^* \circ \lambda. \tag{8.1}$$

Suppose further that some symmetric matrix x satisfies $\lambda(x) \in \operatorname{dom} h$. Then the symmetric matrix w is a subgradient of $h \circ \lambda$ at x if and only if $\lambda(w)$ is a subgradient of h at $\lambda(x)$ and x and w have simultaneous spectral decompositions $x = u^T(\operatorname{Diag}\lambda(x))u$ and $w = u^T(\operatorname{Diag}\lambda(w))u$ for some orthogonal matrix u. In fact, the following "chain rule" holds:

$$\partial(h \circ \lambda)(x) = \{u^T(\operatorname{Diag}\mu)u \mid u \in \mathcal{O}_n,\ u^T(\operatorname{Diag}\lambda(x))u = x,\ \mu \in \partial h(\lambda(x))\}.$$

If h is convex then $h \circ \lambda$ is differentiable at x if and only if h is differentiable at $\lambda(x)$.

EXAMPLE 8.2 (the log barrier). Let us define a symmetric function $h : \mathbf{R}^n \to (-\infty, +\infty]$ by

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

Then h is a closed, convex function, essentially smooth, and essentially strictly convex, with conjugate

$$h^*(\mu) = \begin{cases} -n - \sum_{i=1}^n \log(-\mu_i) & \text{if } \mu < 0, \\ +\infty & \text{otherwise.} \end{cases}$$

It follows that the matrix function $h \circ \lambda : S_n \to (-\infty, +\infty]$ defined by

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

is also closed, convex, essentially smooth, and essentially strictly convex with conjugate

$$(h \circ \lambda)^*(w) = \begin{cases} -n - \log(\det(-w)) & \text{if } w \text{ is negative definite,} \\ +\infty & \text{otherwise.} \end{cases}$$

It is easy to check, using the chain rule, that for a positive-definite symmetric matrix x,

$$\nabla(h \circ \lambda)(x) = -x^{-1}.$$

The convexity part of Theorem 8.1 was essentially first proved in [6]. It was rediscovered in [3]. A characterization of convexity in the differentiable case was proved in [13] via Schur convexity, and the closed case was proved via the conjugacy formula (8.1) in [18]. The latter paper also contains the remainder of Theorem 8.1. A proof appears in [36] that $h \circ \lambda$ is analytic at x if and only if h is analytic at $\lambda(x)$. Somewhat related results appear in [20]. Numerous formulae for subgradients of specific matrix functions appear, for example, in [29, 30, 15, 16]. The chain rule in Theorem 8.1 provides a simple unified approach to these.

THEOREM 8.3 (spectral convex sets). *Weakly orthogonally invariant subsets of S_n are exactly those sets of the form $\lambda^{-1}(C)$ for symmetric subsets C of $\mathbf{R}^n$. If the symmetric matrix x has $\lambda(x)$ in the symmetric set C then a symmetric matrix w lies in the normal cone $N(x|\lambda^{-1}(C))$ if and only if $\lambda(w)$ lies in $N(\lambda(x)|C)$ with x and w having simultaneous spectral decompositions, $x = u^T(\operatorname{Diag}\lambda(x))u$ and $w = u^T(\operatorname{Diag}\lambda(w))u$ for some orthogonal matrix u. In fact,*

$$\begin{aligned} &N(x|\lambda^{-1}(C)) \\ &= \{u^T(\operatorname{Diag}\mu)u \mid u \in \mathcal{O}_n,\ u^T(\operatorname{Diag}\lambda(x))u = x,\ \mu \in N(\lambda(x)|C)\}. \end{aligned}$$

Furthermore,

$$\begin{aligned} (\lambda^{-1}(C))^- &= \lambda^{-1}(C^-), \\ (\lambda^{-1}(C))^\circ &= \lambda^{-1}(C^\circ), \\ \operatorname{int}(\lambda^{-1}(C)) &= \lambda^{-1}(\operatorname{int}(C)), \quad \textit{and} \\ \operatorname{Diag} C &= \lambda^{-1}(C) \cap \operatorname{Diag}\mathbf{R}^n. \end{aligned}$$

The set $\lambda^{-1}(C)$ is convex (respectively, closed) if and only if C is convex (respectively, closed). If C is convex then

$$\begin{aligned} \operatorname{ri}(\lambda^{-1}(C)) &= \lambda^{-1}(\operatorname{ri}(C)), \\ \operatorname{aff}(\lambda^{-1}(C)) &= \lambda^{-1}(\operatorname{aff}(C)), \quad \textit{and} \\ \operatorname{ext}(\lambda^{-1}(C)) &= \lambda^{-1}(\operatorname{ext}(C)), \end{aligned}$$

and if C is in addition closed then

$$\exp(\lambda^{-1}(C)) = \lambda^{-1}(\exp(C)).$$

EXAMPLE 8.4 (the simplex). Let us define a symmetric subset of $\mathbf{R}^n$ by

$$C = \left\{ \alpha \in \mathbf{R}^n \;\middle|\; \alpha \geq 0,\ \sum_{i=1}^{n} \alpha_i = 1 \right\}.$$

Then C is a closed, convex set with the standard unit vectors as extreme (in fact, exposed) points. We deduce that the set of symmetric matrices

$$\lambda^{-1}(C) = \{x \in S_n \mid x \text{ positive semidefinite, } \operatorname{tr}(x) = 1\}$$

is closed and convex with extreme (exposed) points yy^T for unit column vectors y in $\mathbf{R}^n$.

The fact that the set $\lambda^{-1}(C)$ is convex if and only if C is convex, for a symmetric closed set C, was proved in [11].

Unitarily invariant norms. A function $h : \mathbf{R}^l \to [-\infty, +\infty]$ is *absolutely symmetric* if the value $h(\alpha)$ at a vector α in $\mathbf{R}^l$ is independent of the order and signs of the components α_i: in the notation of Example 7.2, $h(\alpha) = h(\overline{|\alpha|})$ for all α. In particular, if such a function is also a norm then it is called a *symmetric gauge function.* A matrix function $f : M_{m,n}(\mathbf{C}) \to [-\infty, +\infty]$ is *(strongly) unitarily invariant* if $f(u^*xv) = f(x)$ for any matrix x in $M_{m,n}(\mathbf{C})$ and unitary matrices u and v.

The following result is a consequence of applying our machinery to Example 7.6. The real analogue is entirely similar. For brevity, we restrict ourselves to the norm case.

THEOREM 8.5 (unitarily invariant norms). *Unitarily invariant norms on $M_{m,n}(\mathbf{C})$ are exactly those functions of the form $h \circ \sigma$ for symmetric gauge functions h on $\mathbf{R}^l$ (where $l = \min\{m, n\}$). In this case the dual norm is given by*

$$(p \circ \sigma)^D = p^D \circ \sigma, \tag{8.2}$$

$p \circ \sigma$ is smooth (respectively, strict) if and only if p is smooth (respectively, strict), and a matrix x is an extreme (respectively, exposed, smooth) point of the unit ball for $p \circ \sigma$ if and only if $\sigma(x)$ is an extreme (respectively, exposed, smooth) point of the unit ball for p. Furthermore, a matrix w is a subgradient of $p \circ \sigma$ at x if and only if $\sigma(w)$ is a subgradient of p at $\sigma(x)$ with x and w having simultaneous singular value decompositions $x = u^*(\operatorname{Diag}\sigma(x))v$ and $w = u^*(\operatorname{Diag}\sigma(w))v$ for unitary matrices u and v. In fact,

$$\begin{aligned}&\partial(p \circ \sigma)(x)\\ &= \{u^*(\operatorname{Diag}\mu)v \mid u \in \mathcal{U}_m,\ v \in \mathcal{U}_n,\ u^*(\operatorname{Diag}\sigma(x))v = x,\ \mu \in \partial p(\sigma(x))\}.\end{aligned}$$

The classical examples are the symmetric gauge function $\|\cdot\|_p$ (for $1 \leq p \leq \infty$), which gives the "Schatten p-norm," and the functions

$$p(\alpha) = \sum_{i=1}^{k} (\overline{|\alpha|})_i \quad (\text{for } k = 1, 2, \ldots, l),$$

which give the "Ky Fan k-norms."

The fundamental characterization of unitarily invariant norms is due to von Neumann [27]. He proved the result in an analogous fashion to our conjugacy argument following Theorem 4.4 by proving the duality formula (8.2) via his lemma (7.2). Some interesting analogous results appear in [4]. The characterization of extreme, exposed, and smooth points was proved in [2]; see also [40, 41, 8, 7, 9]. Versions of the subdifferential formula appear in [38, 39].

Acknowledgments. Many thanks to Jean-Pierre Haeberly for simplifying the presentation of Definition 2.1, and for a number of other helpful comments. Many thanks also to Juan-Enrique Martinez-Legaz for helpful discussions concerning the presentation of Definition 2.1.

REFERENCES

[1] F. Alizadeh, *Optimization over the positive definite cone: interior point methods and combinatorial applications*, in Advances in Optimization and Parallel Computing, P. Pardolos, ed., North–Holland, Amsterdam, 1992, pp. 1–25.

[2] J. Arazy, *On the geometry of the unit ball of unitary matrix spaces*, Integral Equations Operator Theory, 4 (1981), pp. 151–171.

[3] J. M. Ball, *Convexity conditions and existence theorems in nonlinear elasticity*, Arch. Rational Mech. Anal., 63 (1976), pp. 337–403.

[4] A. Barbara and J.-P. Crouzeix, *Concave gauge functions and applications*, ZOR—Math. Meth. Oper. Res., 40 (1994), pp. 43–74.

[5] J. Cullum, W. E. Donath, and P. Wolfe, *The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices*, Math. Programming Study, 3 (1975), pp. 35–55.

[6] C. Davis, *All convex invariant functions of hermitian matrices*, Archiv der Mathematik, 8 (1957), pp. 276–278.

[7] E. M. de Sá, *Exposed faces and duality for symmetric and unitarily invariant norms*, Linear Algebra Appl., 197, 198 (1994), pp. 429–450.

[8] ———, *Faces and traces of the unit ball of a symmetric gauge function*, Linear Algebra Appl., 197, 198 (1994), pp. 349–395.

[9] ———, *Faces of the unit ball of a unitarily invariant norm*, Linear Algebra Appl., 197, 198 (1994), pp. 451–493.

[10] K. Fan, *On a theorem of Weyl concerning eigenvalues of linear transformations*, Proc. Nat. Acad. Sci. U.S.A., 35 (1949), pp. 652–655.

[11] P. A. Fillmore and J. P. Williams, *Some convexity theorems for matrices*, Glasgow Math. J., 12 (1971), pp. 110–117.

[12] R. Fletcher, *Semi-definite matrix constraints in optimization*, SIAM J. Control Optim., 23 (1985), pp. 493–513.

[13] S. Friedland, *Convex spectral functions*, Linear and multilinear algebra, 9 (1981), pp. 299–316.

[14] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, UK, 1952.

[15] J.-B. Hiriart-Urruty, A. Seeger, and D. Ye, *Sensitivity Analysis for a Class of Convex Functions Defined Over a Space of Symmetric Matrices*, Lecture Notes in Economics and Mathematical Systems, Vol. 382, Springer, 1992.

[16] J.-B. Hiriart-Urruty and D. Ye, *Sensitivity analysis of all eigenvalues of a symmetric matrix*, Numer. Math., 70 (1995), pp. 45–72.

[17] F. Jarre, *An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices*, SIAM J. Control Optim., 31 (1993), pp. 1360–1377.

[18] A. S. Lewis, *Convex analysis on the Hermitian matrices*, SIAM J. Optim., 6 (1996), pp. 164–177.

[19] ———, *The convex analysis of unitarily invariant functions*, J. Convex Anal., 2 (1995), pp. 173–183.

[20] ———, *Derivatives of spectral functions*, Math. Oper. Res., 1996, to appear.

[21] ———, *Eigenvalue-constrained faces*, Linear Algebra Appl., submitted.

[22] ———, *Von Neumann's lemma and a Chevalley-type theorem for convex functions on Cartan subspaces*, Trans. Amer. Math. Soc., submitted.

[23] A. W. MARSHALL AND I. OLKIN, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979.

[24] J.-E MARTÍNEZ-LEGAZ, *On Convex and Quasi-Convex Spectral Functions*, Tech. report, Universitat Autònoma de Barcelona, Barcelona, Spain, 1995.

[25] L. MIRSKY, *On the trace of matrix products*, Mathematische Nachrichten, 20 (1959), pp. 171–174.

[26] Y. E. NESTEROV AND A. S. NEMIROVSKII, *Interior Point Polynomial Methods in Convex Programming*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1994.

[27] J. VON NEUMANN, *Some matrix inequalities and metrization of matric-space*, Tomsk University Rev., 1 (1937), pp. 286–300. In Collected Works, Vol. IV, Pergamon, Oxford, 1962, pp. 205–218.

[28] M. L. OVERTON, *On minimizing the maximum eigenvalue of a symmetric matrix*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 256–268.

[29] ———, *Large-scale optimization of eigenvalues*, SIAM J. Optim., 2 (1992), pp. 88–120.

[30] M. L. OVERTON AND R. S. WOMERSLEY, *Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices*, Math. Programming, Series B, 62 (1993), pp. 321–357.

[31] E. POLAK AND Y. WARDI, *A nondifferentiable optimization algorithm for structural problems with eigenvalue inequality constraints*, J. Structural Mech., 11 (1983), pp. 561–577.

[32] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[33] A. SEEGER, *Convex analysis of spectrally-defined matrix functions*, SIAM J. Optim., to appear.

[34] A. SHAPIRO, *Extremal problems on the set of nonnegative definite matrices*, Linear Algebra Appl., 67 (1985), pp. 7–18.

[35] C. M. THEOBALD, *An inequality for the trace of the product of two symmetric matrices*, Math. Proc. Cambridge Philosophical Soc., 77 (1975), pp. 265–266.

[36] N.-K. TSING, M. K. H. FAN, AND E. I. VERRIEST, *On analyticity of functions involving eigenvalues*, Linear Algebra Appl., 207 (1994), pp. 159–180.

[37] G. A. WATSON, *An algorithm for optimal l_2 scaling of matrices*, IMA J. Numer. Anal., 11 (1991), pp. 481–492.

[38] ———, *Characterization of the subdifferential of some matrix norms*, Linear Algebra Appl., 170 (1992), pp. 33–45.

[39] ———, *On matrix approximation problems with Ky Fan k norms*, Numer. Algorithms, 5 (1993), pp. 263–272.

[40] K. ZIĘTAK, *On the characterization of the extremal points of the unit sphere of matrices*, Linear Algebra Appl., 106 (1988), pp. 57–75.

[41] ———, *Subdifferentials, faces and dual matrices*, Linear Algebra Appl., 185 (1993), pp. 125–141.