Positivity 13 (2009), 559–574

1385-1292/030559-16, *published online* October 4, 2008
DOI 10.1007/s11117-008-2172-4

Positivity

Muirhead-Rado inequality for compact groups

Leonard J. Schulman

Abstract. Muirhead's majorization inequality was extended by Rado to the case of arbitrary permutation groups. We further generalize this inequality to compact groups and their linear representations over the reals. We characterize saturation of the inequality, and describe the saturation condition in detail for the case of actions on Hermitian operators.

Mathematics Subject Classification (2000). 52A40, 26D05, 60E15.

Keywords. Inequality, majorization, symmetric function.

1. Introduction

The purpose of this paper is to give a natural extension of a classic majorization inequality due in its initial form to Muirhead, and in more general form to Rado; and to describe the cases in which the inequality is satisfied with equality. We begin by recalling some definitions.

1. Definition: For vectors $a, b \in \mathbb{R}^n$, say that $b \succeq a$ ("b majorizes a") if there exists a probability distribution t on the symmetric group S_n such that $\sum_{\pi \in S_n} t(\pi)\pi b = a$. (Here $(\pi b)_i = b_{\pi(i)}$.)

Equivalently: $b \succeq a$ if $\sum_i a_i = \sum_i b_i$, and if, after sorting the entries of each vector into nonincreasing order, the sum of the entries in every prefix of b is no less than the corresponding sum in a.

A *monomial symmetric function* on n variables is specified by an "exponent vector" $a = (a_1, \dots, a_n)$. We allow the $\{a_i\}$ to be real numbers. (For the extensive algebraic theory of the case of nonnegative integer $\{a_i\}$, see [15].) The nonnegative reals are denoted $\mathbb{R}_+$, and the positive reals $\mathbb{R}_{++}$. The monomial symmetric function $\mu_a(x)$ is defined for vectors $x \in \mathbb{R}^n_{++}$ by

$$\mu_a(x) = \frac{1}{n!} \sum_{\pi \in S_n} \prod_{i=1}^{n} x_{\pi(i)}^{a_i}.$$

Supported in part by the National Science Foundation and the Center for the Mathematics of Information.

Muirhead's inequality

2. Theorem: [Muirhead] *$b \succeq a$ if and only if $\mu_b(x) \geq \mu_a(x)$ for all $x \in \mathbb{R}^n_{++}$.*

Muirhead demonstrated his inequality for the case of integer vectors [9]. Its full statement is given in the classic monograph of Hardy, Littlewood and Polya [5] (§2.18-2.20, pp. 44-49). Among its implications are the power-mean inequality and Newton's inequality for elementary symmetric functions.

Rado's inequality

Rado gave an elegant extension of Muirhead's inequality [10]. To describe it we need to define G-majorization. Let G be a finite group acting on $\{1, \ldots, n\}$.

3. Definition: *Let $a, b \in \mathbb{R}^n$. Say that b G-majorizes a, written $b \succeq_G a$, if there exists a probability distribution t on G such that $a = \sum t(g)(b \circ g)$, or in other words for all $i \in \{1, \ldots, n\}$, $a(i) = \sum_g t(g) b(g(i))$.*

Next we need the "(G,a)-mean" defined by a vector $a \in \mathbb{R}^n$:

$$\mu_{G,a}(x) = \frac{1}{|G|} \sum_{g \in G} \prod_{i=1}^{n} x_{g(i)}^{a_i}. \tag{1}$$

Rado's inequality is:

4. Theorem: [Rado] *$b \succeq_G a$ if, and only if, $\mu_{G,b}(x) \geq \mu_{G,a}(x)$ for all $x \in \mathbb{R}^n_{++}$.*

It is more natural to write Rado's theorem and proof in terms of the logarithms $w_i = \log x_i$, and a pairing among vectors in $\mathbb{R}^n$:

5. Definition: *For $w, a \in \mathbb{R}^n$ let $\langle w, a \rangle_G = \frac{1}{|G|} \sum_g \exp \sum_i a_{g(i)} w_i$.*

Comments: (i) This definition is symmetric: $\langle w, a \rangle_G = \langle a, w \rangle_G$. (ii) $\log \langle a, w \rangle_{S_n}$ $> a^\dagger w$, with equality if and only if one of w or a is constant.

In terms of this definition Rado's inequality is that $b \succeq_G a$ if and only if $\langle w, b \rangle_G \geq \langle w, a \rangle_G$ for all $w \in \mathbb{R}^n$.

Literature. The volume by Marshall and Olkin gives a wealth of material about majorization and majorization inequalities [8]. An extension of the Muirhead inequality to function spaces can be found in [6,12]; it may be of interest to pursue a common generalization of that and the present work. For an extension of the Muirhead-Rado inequality in a direction different from ours, see [2]. For recent majorization-related results on symmetric functions see [4].

Outline. In Section 2 we extend the Muirhead-Rado inequality from finite permutation groups to unitary actions of compact groups (we should add that there is little difficulty in this); and perhaps more interestingly, characterize saturation of the inequality. In Section 3 we consider the case of group actions on Hermitian operators and give a representation-theoretic characterization of the saturation condition for this case. But first, a couple of examples.

Examples

6. Example: Let $n \geq 4$ and let G be the cyclic group C_n. The action of G on $\{1, \ldots, n\}$ by addition modulo n yields the following inequality for $x \in \mathbb{R}^n_{++}$:

$$\sum_{i=1}^{n} x_i^3 x_{i+1}^6 \geq \sum_{i=1}^{n} x_i^1 x_{i+1}^2 x_{i+2}^2 x_{i+3}^4.$$

7. Example: We show how to obtain Hölder's inequality from Rado's. Write $\mu_t(x)$ for $\mu_{t,0,\ldots,0}(x)$ (with $n-1$ 0's). Hölder's inequality states that for $x, y \in \mathbb{R}^n_+$ and $0 < \alpha < 1$, and with z representing the vector $(x_1y_1, \ldots, x_ny_n)$,

$$(\mu_{\frac{1}{\alpha}}(x))^{\alpha}(\mu_{\frac{1}{1-\alpha}}(y))^{1-\alpha} \geq \mu_1(z). \tag{2}$$

Proof. By continuity it suffices to show this for $x, y \in \mathbb{R}^n_{++}$ and α rational, $\alpha = c/(c+d)$ for positive integer c, d. In other words we are to show $(\mu_{\frac{c+d}{c}}(x))^c(\mu_{\frac{c+d}{d}}(y))^d \geq (\mu_1(z))^{c+d}$. Let T be the $2 \times n$ matrix

$$T = \begin{pmatrix} \log x_1 & \ldots & \log x_n \\ \log y_1 & \ldots & \log y_n \end{pmatrix}.$$

Let S_n act on the columns of T. Then

$$\begin{aligned}
\mu_{\frac{c+d}{c}}(x) &= \left\langle \begin{pmatrix} \frac{c+d}{c} & 0 \ldots 0 \\ 0 & 0 \ldots 0 \end{pmatrix}, T \right\rangle_{S_n} \\
\mu_{\frac{c+d}{d}}(y) &= \left\langle \begin{pmatrix} 0 & 0 \ldots 0 \\ \frac{c+d}{d} & 0 \ldots 0 \end{pmatrix}, T \right\rangle_{S_n} \\
\mu_1(z) &= \left\langle \begin{pmatrix} 1 & 0 \ldots 0 \\ 1 & 0 \ldots 0 \end{pmatrix}, T \right\rangle_{S_n}
\end{aligned}$$

If C is a multiset of $\{1, \ldots, n\}$ of size c, in other words a nonnegative integer function with $\sum_1^n C(i) = c$, let $p(C) = n^{-c}\binom{c}{C(1), \ldots, C(n)}$. If E is a multiset of size $c+d$ then it can be written as the sum of multisets C of size c, and D of size d, in several ways; $p(E) = \sum_{C+D=E} p(C)p(D)$.

For a multiset C, T_C will represent the $2 \times c$ matrix consisting of the columns of T indexed by C, with repetition (in any order). Then

$$\begin{aligned}
(\mu_{\frac{c+d}{c}}(x))^c &= \left(\left\langle \begin{pmatrix} \frac{c+d}{c} & 0 \ldots 0 \\ 0 & 0 \ldots 0 \end{pmatrix}, T \right\rangle_{S_n} \right)^c \\
&= \sum_C p(C) \left\langle \begin{pmatrix} \frac{c+d}{c} & \cdots & \frac{c+d}{c} \\ 0 & \cdots & 0 \end{pmatrix}, T_C \right\rangle_{S_c}.
\end{aligned}$$

Likewise for multisets D of size d and E of size $c+d$,

$$(\mu_{\frac{c+d)}{d}}(y))^d = \sum_D p(D) \left\langle \begin{pmatrix} 0 & \cdots & 0 \\ \frac{c+d}{d} & \cdots & \frac{c+d}{d} \end{pmatrix}, T_D \right\rangle_{S_d}$$

$$(\mu_1(z))^{c+d} = \sum_E p(E) \left\langle \begin{pmatrix} 1 & \ldots & 1 \\ 1 & \ldots & 1 \end{pmatrix}, T_E \right\rangle_{S_{c+d}}.$$

Finally, $(\mu_{\frac{c+d}{c}}(x))^c (\mu_{\frac{c+d}{d}}(y))^d - (\mu_1(z))^{c+d}$

$$= \sum_E p(E) \left[\left\langle \begin{pmatrix} \frac{c+d}{c} & \cdots & \frac{c+d}{c} & 0 & \cdots & 0 \\ 0 & \ldots & 0 & \frac{c+d}{d} & \cdots & \frac{c+d}{d} \end{pmatrix}, T_E \right\rangle_{S_{c+d}} - \left\langle \begin{pmatrix} 1 & \ldots & 1 \\ 1 & \ldots & 1 \end{pmatrix}, T_E \right\rangle_{S_{c+d}} \right]$$

$$\geq 0$$

because by Rado's inequality each term in the summation is nonnegative. □

2. Extension to compact groups

2.1. Preliminaries

Here we extend Rado's inequality in two ways. The more straightforward is that G, which in Rado's work is confined to permuting the axes in $\mathbb{R}^n$, will be permitted to be any finite linear group acting on $\mathbb{R}^n$. The second is that G will be allowed to be a compact, rather than finite, group, with a unitary action on $\mathbb{R}^n$. Both directions of the Muirhead-Rado theorem will survive these extensions. Later we will consider the special case that $\mathbb{R}^n = \mathbb{R}^{d^2}$ is the space of Hermitian operators on $\mathbb{C}^d$, on which G acts by conjugation.

The nondegenerate inner product on $\mathbb{R}^n$ between vectors a, w (with respect to which G has a unitary action) will be written $a^\dagger w$.

Let η be the left invariant Haar measure on G, normalized so that the measure of G is 1.

8. Definition: *For $a, b \in \mathbb{R}^n$, and G a compact group acting on $\mathbb{R}^n$, say that b G-majorizes a, written $b \succeq_G a$, if there is a probability measure t on G such that $a = \int t(g) gb \, d\eta(g)$.*

9. Definition: *For $a, w \in \mathbb{R}^n$, $\langle a, w \rangle_G = \int_G e^{a^\dagger g w} d\eta(g)$.*

10. Comment: Uniqueness of the left invariant measure η, up to a multiplicative constant, is implied by local compactness of G. For locally compact groups the left and right invariant measures may, however, be different. Observing that the right invariant measure is $\eta'(g) = \eta(g^{-1})$, we can write $\langle a, w \rangle_G = \int_G e^{a^\dagger g w} \, d\eta(g) = \int_G e^{a^\dagger g^{-1} w} \, d\eta'(g) = \int_G e^{w^\dagger g^{-1\dagger} a} \, d\eta'(g) = \int_G e^{w^\dagger g a} \, d\eta'(g)$.

Therefore if $\eta = \eta'$, the pairing is symmetric: $\langle a, w\rangle_G = \langle w, a\rangle_G$. Groups for which the left and right invariant measures are equal are called unimodular; every compact group is unimodular [11]. Since we are concerned here only with compact groups, our pairing is indeed symmetric. (In integrals we'll now write dg for $d\eta(g)$.)

2.2. Inequality

11. Theorem: *Let G be a compact group acting unitarily on $\mathbb{R}^n$ and let $a, b \in \mathbb{R}^n$. Then $b \succeq_G a$ if, and only if, $\langle b, w\rangle_G \geq \langle a, w\rangle_G$ for all $w \in \mathbb{R}^n$.*

Proof. There are two directions; in both we can follow the outline of the proof of the finite case.

Direction I: Assume $b \succeq_G a$.

In this direction the proof is unchanged from the finite case (replacing only summations by integrals) and is included only for completeness.

Let t be a probability measure on G such that

$$a = \int_G t(h)hb\, dh. \tag{3}$$

Then:

$$\begin{aligned}
\langle w, a\rangle_G &= \int_G e^{w^\dagger ga}\, dg \\
&= \int_G e^{\int_G t(h)w^\dagger ghb\, dh}\, dg \\
&\leq \int_G \int_G t(h)e^{w^\dagger ghb}\, dh\, dg && \text{Arithmetic-geometric mean inequality (AGM)} \\
&= \int_G t(h) \int_G e^{w^\dagger kb}\, dk\, dh && \text{The integration is w.r.t. left-invariant measure} \\
&= \int_G e^{w^\dagger kb}\, dk \\
&= \langle w, b\rangle_G
\end{aligned}$$

Direction II: Assume $b \not\succeq_G a$.

In this direction, although our proof still follows the outline of the finite case, we have to handle a small complication: in the finite case a single "outlying" element of the group already has positive Haar measure, but this is not so for infinite groups. So, first we use the hyperplane separation lemma to obtain positive "separated" measure in the space the group is acting upon; then with an averaging argument we "lift" this bound to the group, and show that in a compact group there is a positive amount of "separated" Haar measure.

Let S be the unit sphere in $\mathbb{R}^n$. Let $f(v) = \sup_{h\in G} v^\dagger hb$. By the hyperplane separation lemma, there exist $v \in S$ and $\delta_0 > 0$ such that $v^\dagger a > f(v) + \delta_0$.

For $u \in S$, let

$$N_{u,\delta} = \{w \in S : \cos\delta < u^\dagger w\}.$$

For v and δ_0 as above, let

$$M_{v,\delta_0} = \{w \in S : w^\dagger a > f(v) + \delta_0\}.$$

There is a $\delta_1 > 0$ such that $N_{v,\delta_1} \subseteq M_{v,\delta_0}$. Let $A_{\delta_1} = \{h \in G : hv \in N_{v,\delta_1}\}$. Let ν denote the uniform measure on S, with the normalization $\nu(S) = 1$; let $\hat{\nu}(\delta) = \nu(N_{v,\delta})$ and note that $\hat{\nu}(0) = 0$ and that $\hat{\nu}$ is strictly increasing. The proof now relies on the following bound.

12. Lemma: $\eta(A_{\delta_1}) \geq \hat{\nu}(\delta_1/2) > 0$.

Proof. For a vector $u \in S$ and $r > 0$ let $t(u,r) = \eta(\{h \in G : hv \in N_{u,r}\})$. The average of $t(u,r)$ over u chosen uniformly in S is $\hat{\nu}(r)$. We argue that there is a vector $u \in N_{v,\delta_1/2}$ for which $t(u,\delta_1/2) \geq \hat{\nu}(\delta_1/2)$. Otherwise, by G-invariance of t, no u for which $t(u,\delta_1/2) \geq \hat{\nu}(\delta_1/2)$ would have any point of the form hv (for $h \in G$) within its neighborhood $N_{u,\delta_1/2}$, contradicting the positivity of $\hat{\nu}(\delta_1/2)$.

Now, $N_{u,\delta_1/2} \subseteq N_{v,\delta_1}$, hence $\eta(A_{\delta_1}) \geq \hat{\nu}(\delta_1/2)$. □

To prove the theorem write, for any $c > 0$:

$$\begin{aligned}
\langle a, vc\rangle_G &= \int_G e^{a^\dagger gvc}\, dg \\
&\geq \int_{A_{\delta_1}} e^{a^\dagger gvc}\, dg \\
&\geq \int_{A_{\delta_1}} e^{(f(v)+\delta_0)c}\, dg \\
&= e^{(f(v)+\delta_0)c}\eta(A_{\delta_1}) \\
&\geq e^{c\delta_0}\eta(A_{\delta_1})\langle b, vc\rangle_G
\end{aligned}$$

The theorem follows by choosing $w = vc$ for c sufficiently large that $e^{c\delta_0}\eta(A_{\delta_1}) > 1$. □

Remark: Deviation bounds for G-symmetric random walks. Define the "(G,b) random walk" in $\mathbb{R}^n$ to be a walk which takes independent steps each selected from the set $\{gb\}_{g\in G}$ according to the Haar measure on g.

The Chernoff bound on the displacement of this walk from the origin in the direction of a vector $w \in \mathbb{R}^n$, is governed by the quantity $E(e^{w^\dagger gb}) = \langle w, b\rangle_G$. Specifically the bound is (for any real β)

$$P(\frac{1}{m}w^\dagger \sum_1^m g_i b \geq t) \leq \left(\frac{\langle \beta w, b\rangle_G}{e^{\beta t}}\right)^m \tag{4}$$

(this can be converted to an explicit bound after optimizing over β). Theorem 11 is therefore equivalent to stating that $b \succeq_G a$ if and only if the Chernoff bound (4) for the (G,a) walk is at least as strict, in every direction w, as the corresponding deviation bound for the (G,b) walk.

2.3. Saturation of the inequality

For $b \succeq_G a$ we wish to characterize

$$W(b,a) = \{w : \langle a, w\rangle_G = \langle b, w\rangle_G\},$$

the set of w for which equality is attained in Theorem 11.

Let $b \succeq_G a$. Fix t as in Equation 3, and pick an $h_0 \in G$ such that for every open neighborhood N of h_0, $t(N) > 0$. There is such an h_0 because otherwise each h has an open neighborhood $N(h)$ for which $t(N(h)) = 0$; these neighborhoods have a finite subcover J, so $1 = t(G) \leq \sum_{h\in J} t(N(h)) = 0$. (For this argument we don't need G to be compact; the Lindelöf condition guaranteeing that J is countable would suffice.)

Let $\Delta = h_0 b - a = \int_G t(h)(h_0 b - hb)\, dh$.

13. Theorem: *Let $b \succeq_G a$. Then $\langle a, w\rangle_G = \langle b, w\rangle_G$ if and only if $w^\dagger g\Delta = 0$ for all $g \in G$. In other words, $W(b,a) = \{w : w^\dagger g\Delta = 0$ for all $g \in G\}$.*

Rado didn't fully address saturation of the inequality in his work, but he examined the regular representation of an abelian group, and showed that in that case the above characterization holds.

Proof. First assume that $w^\dagger g\Delta = 0$ for all $g \in G$. Rewrite the beginning of the proof in Theorem 11:

$$\begin{aligned}
\langle a, w\rangle_G &= \int_G e^{w^\dagger g a}\, dg \\
&= \int_G e^{w^\dagger g(h_0 b - \Delta)}\, dg \\
&= \int_G e^{w^\dagger g h_0 b}\, dg \\
&= \int_G e^{w^\dagger g b}\, dg \\
&= \langle b, w\rangle_G.
\end{aligned}$$

For the reverse implication, observe that the only inequality in Theorem 11 is in the use of the AGM:

$$\begin{aligned}
\langle a, w\rangle_G = \ldots &= \int_G e^{\int_G t(h) w^\dagger g h b\, dh}\, dg \\
&\leq \int_G \int_G t(h) e^{w^\dagger g h b}\, dh\, dg
\end{aligned}$$

If the inequality is tight then the AGM must be tight for all g except a set of measure 0; this in turn means that for such g, $w^\dagger g h b$ is a constant function on sets of positive probability in t, or to be more precise, for any measurable sets $A, B \subseteq G$,

$$t(A) \int_B t(h) w^\dagger g h b\, dh = t(B) \int_A t(h) w^\dagger g h b\, dh. \tag{5}$$

Each side of Equation 5 is continuous in g, so it in fact holds for all g. Allowing A to range over a nested sequence of open neighborhoods of h_0, and using continuity of $w^\dagger ghb$ in h as well as the fact that $t(A) > 0$ for all A, we find that for all g and all measurable B,

$$\int_B t(h)w^\dagger ghb\, dh = t(B)w^\dagger gh_0 b$$

$$w^\dagger g \int_B t(h)(hb - h_0 b)\, dh = 0$$

Letting $B = G$,

$$w^\dagger g \int_G t(h)(hb - h_0 b)\, dh = 0$$

$$w^\dagger g\Delta = 0.$$

□

3. Conjugation actions on Hermitian operators

3.1. Preliminaries

In this section we specialize to the interesting case that $n = d^2$ and $\mathbb{R}^n$ is the real vector space $\mathcal{H}_d$ of $d \times d$ Hermitian matrices, acted upon by conjugation.

We focus on this case because (besides that it supplies interesting examples, see Section 3.4), we can improve the characterization of the saturation condition (Theorem 13) to a more useful form (Theorem 20), in which the characterization does not depend upon checking a universal quantification over elements of the group.

There is a nondegenerate inner product on $\mathcal{H}_d$ defined by Tr $(a^\dagger w)$, or what is the same, Tr (aw), for $a, w \in \mathcal{H}_d$. (The restriction of the standard inner product Tr $(a^\dagger w)$ from End$(\mathbb{C}^d)$.) The inner product gives rise to a norm in the usual way. If G is a compact linear group acting on $\mathbb{C}^d$, then conjugation by elements of G is a unitary action on $\mathcal{H}_d$ (because Tr $(g^{-1}a^\dagger g g^{-1} b g) =$ Tr $(a^\dagger b)$). Majorization specializes to:

14. Definition: *For $a, b \in \mathcal{H}_d$, and G a compact group acting on $\mathbb{C}^d$, say that b G-majorizes a, written $b \succeq_G a$, if there is a probability measure t on G such that $a = \int t(g)g^{-1}bg\, d\eta(g)$.*

15. Definition: *For $a, w \in \mathcal{H}_d$, $\langle a, w\rangle_G = \int_G e^{Tr\ (g^{-1}agw)} d\eta(g)$.*

Applying Theorem 11 gives

16. Corollary: *Let G be a compact group acting on $\mathbb{C}^d$ and let $a, b \in \mathcal{H}_d$. Then $b \succeq_G a$ if, and only if, $\langle b, w\rangle_G \geq \langle a, w\rangle_G$ for all $w \in \mathcal{H}_d$.* □

A natural example of this is:

17. Corollary: *Let compact G act irreducibly on $\mathbb{C}^d$. Let $b \in \mathcal{H}_d$. Then $b \succeq_G (Tr\ b/d)I$, and $\langle b, w\rangle_G \geq \langle (Tr\ b/d)I, w\rangle_G$ for all $w \in \mathcal{H}_d$.*

Proof. $\int_G g^{-1}bg\ dg$ is a G-invariant inner product on $\mathbb{C}^d$. Since G is irreducible, Schur's lemma [13] indicates that $\int_G g^{-1}bg\ dg = (\mathrm{Tr}\ b/d)I$. □

18. Example: *(Hölder inequality in the Hermitian case.) Let G be a compact group acting on $\mathbb{C}^d$ and let $a, b, w \in \mathcal{H}_d$, $0 < \alpha < 1$. Then $(\langle a, \frac{w}{\alpha}\rangle_G)^\alpha (\langle b, \frac{w}{1-\alpha}\rangle_G)^{(1-\alpha)} \geq \langle a + b, w\rangle_G$.*

Useful references for the representation theory employed in this section include [3], [14], [7] and [1].

3.2. Restatement of Theorem 13 for Hermitian operators

$W(b, a)$ is a subspace of the real vector space $\mathcal{H}_d$, and is described by

$$W(b, a) = \{w :\ w \perp g^{-1}\Delta g \text{ for all } g \in G\}.$$

(Naturally $w \perp g^{-1}\Delta g$ means that $\mathrm{Tr}\ (wg^{-1}\Delta g) = 0$.) Here Δ is obtained as in Section 2.3.

Now, instead of considering G as a subgroup of $GL(d, \mathbb{C})$ (as we have to this point), we view G as an abstract group and let (ρ, V) be the representation of G that we have implicitly discussed until now; here $V \simeq \mathbb{C}^d$, and ρ is a homomorphism $G \to GL(V)$, or rather, without loss of generality, a homomorphism of G into the unitary group $U(V)$. Consider the vector space of $d \times d$ complex matrices $V^\dagger \otimes V$. We are interested in the representation $(\rho^* \otimes \rho, V^\dagger \otimes V)$ in which ρ acts on $V^\dagger \otimes V$ by conjugation (here ρ^* is the complex conjugate representation of ρ; recall that ρ is unitary). In other words for a $d \times d$ complex matrix m,

$$((\rho^* \otimes \rho)(g))(m) = \rho^{-1}(g)m\rho(g).$$

The order $\succeq_G$ must be interpreted with respect to this group action: $b \succeq_G a$ if there is a probability measure t on G such that $a = \int t(g)\rho(g^{-1})b\rho(g)\ d\eta(g)$.

We can now restate Theorem 13 as:

$$W(b, a) = \{w : w \perp \rho(g^{-1})\Delta\rho(g)\ \forall g \in G\}. \tag{6}$$

The representation $(\rho^* \otimes \rho, \mathcal{H}_d)$ is a *real* subrepresentation of $(\rho^* \otimes \rho, V^\dagger \otimes V)$, meaning that $\mathcal{H}_d$ is a real subspace of the complex space $V^\dagger \otimes V$, and is preserved under the group action. We refer to $(\rho^* \otimes \rho, \mathcal{H}_d)$ and its subrepresentations as Hermitian representations of G.

3.3. Effective form of the saturation condition for Hermitian operators

The main thing we need to understand, in order to arrive at the effective form of Theorem 13 (i.e., Theorem 20), is the decomposition of the Hermitian representation $(\rho^* \otimes \rho, \mathcal{H}_d)$ into real irreps. Fortunately, this can be approached through the plethysm of $(\rho^* \otimes \rho, V^\dagger \otimes V)$.

Corresponding to a complex vector space U there is a real vector space $U_{\mathbb{R}}$ with $\dim_{\mathbb{R}}(U_{\mathbb{R}}) = 2\dim_{\mathbb{C}}(U)$, obtained by converting each complex coordinate $a + bi$ to the pair (a, b); this also converts a unitary representation (α, U) to an orthogonal representation $(\alpha_{\mathbb{R}}, U_{\mathbb{R}})$.

We need to distinguish among the three types of complex irreps (α, U). Let (α^*, U) be the complex conjugate of α, the irrep of G obtained by entry-wise conjugation: $\alpha^*_{i,j}(g) = (\alpha_{i,j}(g))^*$. Every complex irrep is of one of the following types; moreover, the process below in which a real irrep $\widetilde{\alpha}$ is constructed from a complex irrep α, accounts for all real irreps. (See for example [14] III.5A.4.)

1. α is *of real type,* also called *integer,* if there is a basis in which $\alpha = \alpha^*$ (in other words all matrix entries are real).

   In this case $\alpha_{\mathbb{R}}$ is a direct sum of two copies of a real irrep $\widetilde{\alpha}$ of dimension $\dim_{\mathbb{R}}(\widetilde{\alpha}) = \dim_{\mathbb{C}}(\alpha)$, and character $\chi_{\widetilde{\alpha}} = \chi_\alpha$. (In a basis in which $\alpha = \alpha^*$, the real subspaces of U of fixed phase, such as the purely real and purely imaginary subspaces, are invariant for α.)
2. α is *of quaternionic type,* also called *pseudo real* or *half integer,* if it is not real, yet is equivalent to α^*.

   In this case $\widetilde{\alpha} = \alpha_{\mathbb{R}}$ is a real irrep with dimension $\dim_{\mathbb{R}}(\widetilde{\alpha}) = 2\dim_{\mathbb{C}}(\alpha)$, and character $\chi_{\widetilde{\alpha}} = 2\chi_\alpha$.
3. α is *of complex type* if it is inequivalent to α^*.

   In this case we can construct a real irrep $\widetilde{\alpha}$ with dimension $\dim_{\mathbb{R}}(\widetilde{\alpha}) = 2\dim_{\mathbb{C}}(\alpha)$ and character $\chi_{\widetilde{\alpha}} = \chi_\alpha + \chi^*_\alpha$ by letting (U^*, U) be the "diagonal" real vector space $(U^*, U) = \{(u^*, u) : \ u \in U\}$, and setting $(\widetilde{\alpha}, \widetilde{U}) = ((\alpha^*, \alpha), (U^*, U))$.

   (To see that this is irreducible, note that its real dimension is the same as that of α, so an invariant subspace would yield an invariant subspace of $\alpha_{\mathbb{R}}$. Since $\alpha_{\mathbb{R}}$ is orthogonal, this would mean there are $u, v \in U_{\mathbb{R}}$ such that $u \perp \alpha_g(v)\ \forall g$. However, the inner product in U is preserved in $U_{\mathbb{R}}$, so this would imply the same statement for that pair u, v, taken as elements of U; and this would contradict the irreducibility of U.)

Let $(\rho^* \otimes \rho, V^\dagger \otimes V)$ decompose into equivalence classes of complex irreps τ_i (of dimensions d_i) as follows: If τ_i occurs with multiplicity m_i, then space $\overline{W}_i$ is the complex space of dimension $m_i d_i$ containing all irreps equivalent to τ_i. We'll use (τ_i, W_i) or $(\rho^* \otimes \rho, W_i)$ to refer to any irrep within $\overline{W}_i$. Since $(\rho^* \otimes \rho, V^\dagger \otimes V)$ is a unitary representation, $\overline{W}_i \perp \overline{W}_j$ for $i \neq j$.

Observe that for any i, the irrep $W_i^\dagger$ is the complex conjugate of the irrep W_i (in other words $(\rho^* \otimes \rho, W_i^\dagger)$ is equivalent to τ_i^*). If W_i is of real or quaternionic type, then since it is equivalent to its complex conjugate, $\overline{W}_i = \overline{W}_i^\dagger$.

Now there are two cases. (1) If $\overline{W}_i = \overline{W}_i^\dagger$ then there is a nonzero Hermitian matrix h in $\overline{W}_i$. (Take any element of $\overline{W}_i$. If it is skew-Hermitian multiply it by i, otherwise add it to its conjugate transpose.) The images of h span (over $\mathbb{C}$) some copy of W_i; this space is carried to itself by complex conjugation. By the previous paragraph, this irrep and its complex conjugate are both equal to the action of $\rho^* \otimes \rho$ on this copy of W_i, and so they are identical. Hence W_i is of real type. It follows that $(\rho^* \otimes \rho, V^\dagger \otimes V)$ can contain no irreps of quaternionic type.

Specifically, for any W_i in $\overline{W}_i$, τ_i restricted to $W_i \cap \mathcal{H}_d$ is a real irrep, and $\overline{W}_i \cap \mathcal{H}_d$ contains this real irrep with multiplicity m_i. Note that $\dim_{\mathbb{R}}(\overline{W}_i \cap \mathcal{H}_d) = \dim_{\mathbb{C}}(\overline{W}_i)$.

(2) If $\overline{W}_i \neq \overline{W}_i^\dagger$ then $\overline{W}_i \perp \overline{W}_i^\dagger$. In this case $\overline{W}_i$ does not contain any Hermitian matrices (else the above logic would show that W_i is of real type, so that $\overline{W}_i = \overline{W}_i^\dagger$). Instead apply the generic construction described earlier, to produce the "diagonal" real irrep $\tau_i' = (\rho^* \otimes \rho, (\overline{W}_i^\dagger, \overline{W}_i))$. Note that $(\overline{W}_i^\dagger, \overline{W}_i) \subseteq \mathcal{H}_d$ and that $\dim_{\mathbb{R}}(\tau_i') = \dim_{\mathbb{C}}(\overline{W}_i^\dagger \oplus \overline{W}_i)$.

These two cases construct all Hermitian irreps of $\rho^* \otimes \rho$: count dimensions. Summarizing:

19. Lemma: *Let (ρ, V) be a unitary representation of a compact group G. For some k, ℓ, and $\{m_i\}$, there are inequivalent (complex) irreps $\alpha_1, \dots, \alpha_k$ of real type, and $\alpha_{k+1}, \dots, \alpha_{k+\ell}$ of complex type, such that $(\rho^* \otimes \rho, V^\dagger \otimes V)$ is the direct sum of $k + \ell$ invariant subspaces, of the equivalence classes of the respective α_i, with multiplicities m_i; and such that $(\rho^* \otimes \rho, \mathcal{H}_d)$ is the direct sum of $k+\ell$ invariant subspaces, of the respective equivalence classes $\widetilde{\alpha}_i$ (as constructed from α_i above), with multiplicities m_i.*

Theorem 13 (in the form given in Equation 6) and Lemma 19 together give a useful approach to determining $W(b, a)$ from a and b. Obtain an explicit majorization of a by b, or at least, a guarantee that $b \succeq_G a$, along with a suitable h_0; then compute Δ. Determine the list of Hermitian irreps (σ, V_σ); because of the lemma we can start by computing characters of $\rho^* \otimes \rho$. Finally, $W(b, a)$ is the span of the V_σ for which $\Delta \perp V_\sigma$.

In conclusion, the saturation condition for Hermitian representations has the following description:

20. Theorem: *Let (ρ, V) be a d-dimensional representation of a compact group G, let $a, b \in \mathcal{H}_d$, and let $b \succeq_G a$. Then $w \in W(b, a)$ if and only if for every real irrep $(\widetilde{\alpha}_i, W)$ of $\rho^* \otimes \rho$ described in Lemma 19, the projection of either w or Δ into W is zero.*

3.4. Examples

21. Example: As in Corollary 17, consider the special case of averaging uniformly over the group: $a = \int_G \rho(g^{-1}) b \rho(g)\, dg$. Here $W(b, a)$ is spanned by those Hermitian irreducibles σ of $\rho^* \otimes \rho$ on which $b|_\sigma = (\mathrm{Tr}\ b/d) I|_\sigma$.

22. Example: Example 6 is given in terms of real functions, rather than Hermitian matrices. It can be written in the latter notation with $G = C_n$, with b the $n \times n$ diagonal matrix with $(3, 6, 0, \dots, 0)$ on the diagonal, and with a the diagonal matrix with $(1, 2, 2, 4, 0, \dots, 0)$ on the diagonal. (If we only care to examine real functions, then w is also a diagonal matrix.) Here ρ is the regular representation. Now Δ is the diagonal matrix $(-1, -2, 1, 2, 0, \ldots, 0)$, and $\rho^* \otimes \rho$ decomposes into d copies of ρ; each irrep of C_n appears d times in $\rho^* \otimes \rho$. Each irrep is supported on the diagonal exactly once. If n is odd, the only complex irrep of real type is the trivial representation; if n is even the complex irreps of real type are the trivial and sign representations. The real irreps corresponding to these are one-dimensional, while those corresponding to the complex irreps of complex type are two-dimensional (coming in $(\cos, \sin)$ pairs). Examining Δ we see that the span of the cases of equality includes all real irreps supported off the diagonal, together with the identity, and, if n is even, the sign representation; but it includes none of the other real irreps supported on the diagonal. In terms of the original inequality for real functions, the only nontrivial case of equality (i.e., the only case other than a multiple of the identity) is $(x, x^{-1}, \ldots, x, x^{-1})$, for even n.

23. Example: The special unitary group SU_n in its "standard representation" (action on $\mathbb{C}^n$).

This is a particularly simple example. A theorem of Schur [8] states that $b \succeq_{SU_n} a$ if and only if there is eigenvalue majorization $(\lambda_i(b))_{i=1}^n \succeq_{S_n} (\lambda_i(a))_{i=1}^n$. To describe the cases of equality we note, first, that the identity matrix spans a one-dimensional, trivial, Hermitian irrep; and second, that by the spectral theorem, the orbit of any Hermitian matrix under conjugation by SU_n includes every Hermitian matrix with the same eigenvalues. It follows that, given any pair of nonzero traceless Hermitian matrices, one can be conjugated by some element of SU_n, so that the two have nonzero inner product. Thus the perp space in $\mathcal{H}_n$ to the trivial representation (i.e., trace 0 Hermitians) is itself irreducible. There are no nontrivial cases of saturation of the inequality.

24. Example: *Representations of SU_2.*

SU_2 has a single irrep ρ_d in every dimension $d \geq 1$, with character $\chi_d = \sum_{j=0}^{d-1} e^{i\alpha(d-1-2j)}$ on matrices in SU_2 having the pair of eigenvalues $e^{\pm i\alpha}$. The plethysm is

$$\rho_d^* \otimes \rho_d = \bigoplus_{j=0}^{d-1} \rho_{2j+1}$$

The odd-dimensional irreps are of real type. (In physics this decomposition arises in the addition of the angular momenta of two spin-d particles.) Each irreducible complex subspace (ρ_{2j+1}, W) of $V^\dagger \otimes V$ satisfies $W = W^\dagger$, so W is the direct sum, over $\mathbb{R}$, of its Hermitian and skew-Hermitian restrictions (each of which is invariant under $\rho_d^* \otimes \rho_d$). Thus $(\rho_d^* \otimes \rho_d, \mathcal{H}_d)$ is the direct sum of the real irreps $\widetilde{\rho_{2j+1}}$ for $0 \leq j \leq d-1$:

$$(\rho_d^* \otimes \rho_d, \mathcal{H}_d) = \bigoplus_{j=0}^{d-1} (\widetilde{\rho_{2j+1}}, \mathcal{H}_d)$$

Because of this rich structure, the cases of equality depend substantially on Δ. To illustrate, we spell out the first nontrivial instance, $d = 3$. The three-dimensional representation of the Lie algebra su_2 is generated by the angular momentum operators

$$J_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \qquad J_y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \qquad J_z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

These satisfy the commutation relations $[J_x, J_y] = iJ_z, [J_y, J_z] = iJ_x, [J_z, J_x] = iJ_y$. The space spanned by J_x, J_y, J_z is the three-dimensional real space of Hermitian matrices

$$xJ_x + yJ_y + zJ_z = \begin{pmatrix} 0 & -iz & iy \\ iz & 0 & -ix \\ -iy & ix & 0 \end{pmatrix}$$

This space is invariant under the action of $\rho^* \otimes \rho$. It is also irreducible (conjugating by e^{iJ_x} rotates the J_yJ_z plane, etc.); therefore it is the irrep $\widetilde{\rho}_3$. It's now easy to choose Δ and w to yield a nontrivial case of equality: let Δ, for example, be supported by $\widetilde{\rho}_3$, and w by $\widetilde{\rho}_5$. (One suitable choice for w is diagonal, with entries $(1, -1, 0)$.)

3.5. Converse to Theorem 20

We conclude this section by showing that every case of equality allowed by Theorem 20 can occur.

25. Theorem: *Let (ρ, V) be a unitary representation of a compact group G; let ρ contain k equivalence classes of irreps, with multiplicities $m_1, \dots, m_k$. Let $\{\widetilde{\alpha}_i\}_{i\in T}$ be a maximal set of mutually perpendicular Hermitian irreps of $(\rho^* \otimes \rho, \mathcal{H}_d)$, and let $T' \subseteq T$. Then there exist b and a such that the support of Δ in T equals T' if, and only if, $\{\widetilde{\alpha}_i\}_{i\in T'}$ does not contain any of the $\sum_1^k m_i^2$ copies of the trivial irrep that are in $(\rho^* \otimes \rho, \mathcal{H}_d)$. Equivalently, any set of Hermitian irreps that includes all of the $\sum_1^k m_i^2$ trivial irreps, can occur as the span of the cases of equality for some b and a.*

Proof. Enumeration of the trivial irreps of $(\rho^* \otimes \rho, V^\dagger \otimes V)$ is standard: they are a basis for the center of the group algebra generated by $\{\rho(g)\}_{g\in G}$. Let $\rho = \bigoplus \rho_j$ be a decomposition of ρ into irreps. Schur's lemma states that there is a copy of the trivial representation in $\rho_i^* \otimes \rho_j$ if and only if ρ_i is equivalent to ρ_j, which explains the contribution m_i^2. For our application it remains to note that $\widetilde{\alpha}$ (as constructed in Section 3.3) is a trivial irrep if and only if α is.

The projection of any b into a trivial irrep is unchanged by averaging over the group, hence Δ can have no support in a trivial irrep.

Take any set T' of mutually perpendicular nontrivial Hermitian irreps of $(\rho^* \otimes \rho, \mathcal{H}_d)$. We show how to choose b and a so that Δ is supported precisely on T'. For each $i \in T'$ pick a nonzero Hermitian b_i supported on $\widetilde{\alpha}_i$; set $b = \sum_{i\in T'} b_i$. For each b_i fix a $g_i \in G$ such that $\rho(g_i^{-1})b_i\rho(g_i) \neq b_i$. Take any positive coefficients $\{t_j\}_{j\in T'}$

such that $1 - \sum_{j \in T'} t_j > 0$. Set $a = (1 - \sum_{j \in T'} t_j)b + \sum_{j \in T'} t_j \rho(g_j^{-1}) b \rho(g_j)$, and note that $a = \sum_{i \in T'} a_i$, where a_i is the projection of a on $\widetilde{\alpha}_i$, and $a_i = (1 - \sum_{j \in T'} t_j)b_i + \sum_{j \in T'} t_j \rho(g_j^{-1}) b_i \rho(g_j)$. If $i \in T - T'$ then $b_i = a_i = 0$ so the projection of Δ on $\widetilde{\alpha}_i$ is 0. On the other hand if $i \in T'$ then the points $\{b_i, \{\rho(g_j^{-1}) b_i \rho(g_j)\}_{j \in T'}\}$ are all of equal, nonzero norm (in the trace norm, see Section 3.1), hence each lies at a vertex of their convex hull. Since b_i is unequal to at least one of the other points, a_i does not lie at a vertex of this convex hull. So $a_i \neq b_i$, and Δ has nonzero projection on $\widetilde{\alpha}_i$. □

4. Partition functions

We conclude with an application of this class of inequalities to statistical physics. Our description is brief and intended only to illustrate the connection. As is well known, a quantum mechanical system in a "pure" or perfectly known state is modeled by a vector in a Hilbert space. In many cases this vector space is effectively finite-dimensional (for example when considering transitions among a few energy levels of an atom); we restrict ourselves to this finite-dimensional case. Typically, a group of symmetries G acts naturally on this Hilbert space $\mathbb{C}^d$. For example, the state of a spin-2 particle is a unit vector in $\mathbb{C}^5$; physical rotation of the particle manifests in the Hilbert space through the irreducible representation of $SO(3)$ in $\mathbb{C}^5$.

A quantum mechanical system that is in a "mixed" or imperfectly known state cannot be modeled as a vector in the Hilbert space. Instead, it is described by a more general object: a density matrix with respect to the Hilbert space. This (continuing to assume the space is $\mathbb{C}^d$) is a $d \times d$ positive semi-definite Hermitian operator with trace 1. (A pure state $v \in \mathbb{C}^d$ is described by the outer product $v \otimes v^\dagger$.) The action of G on vectors in $\mathbb{C}^d$ is carried over to density matrices via the conjugation action.

The energy, or Hamiltonian, operator which determines the dynamics of the physical system, is a $d \times d$ Hermitian matrix H. The energy of a system with density matrix D is $\mathrm{Tr}\ (DH)$.

We turn to the statistical aspect. Suppose that a partial measurement of the system has been performed, so that its current state (or more precisely, the observer's knowledge of its state) is a density matrix D; subsequently, the system mixes under the dynamics of a group of symmetries G. Some time after the measurement, then, the system is in statistical equilibrium with respect to G. This does not mean that all memory of D is gone, but as much is gone as G allows: the probability density for the system to be in state $g^{-1}Dg$ is proportional to $e^{-\mathrm{Tr}\ (g^{-1}Dg\beta H)}$. Here β is inversely proportional to temperature.

The function which rescales these quantities into a probability distribution (the Gibbs distribution) is the *partition function,* which plays a central role in statistical physics:

$$Z = \int e^{-\mathrm{Tr}\ (g^{-1}Dg\beta H)} dg.$$

This is none other than $\langle D, -\beta H \rangle_G$. Theorem 11 shows how the partition function changes if one of the arguments (H or D) is kept fixed while the other is changed. For example, suppose that the experimenter may choose whether the system is subjected to Hamiltonians H_1 or H_2. Let Z_1 and Z_2 denote the respective partition functions. If $H_1 \succeq H_2$, then the experimenter may conclude that $Z_1 \geq Z_2$, no matter what D is. Or, suppose that the Hamiltonian of the system is fixed but that the experimenter wishes to make a prediction based on the results of the partial measurement. If D_1 and D_2 correspond to two possible measurement outcomes, and Z_1 and Z_2 are the respective partition functions, then if $D_1 \succeq D_2$, it follows that $Z_1 \geq Z_2$, for any Hamiltonian.

Acknowledgments

Thanks to an anonymous referee for comments that improved this paper.

References

[1] T. Bröcker, T. tom Dieck. *Representations of Compact Lie Groups*, Graduate Texts in Mathematics. vol. 98, Springer (1985).

[2] D.E. Daykin. *Generalisation of the Muirhead-Rado inequality.* Proc. Am. Math. Soc., 30 (1971) 84–86.

[3] W. Fulton, J. Harris. *Representation theory: a first course*, Graduate Texts in Mathematics. vol. 129, Springer, 3d edn, (1991).

[4] K. Guan. *Schur-convexity of the complete elementary symmetric function.* J. Inequal. Appl. 9 (2006).

[5] G. Hardy, J.E. Littlewood, G. Pólya. *Inequalities*, 2nd edn, Cambridge University Press (1952).

[6] G.H. Hardy, J.E. Littlewood, G.Pólya. *Some simple inequalities satisfied by convex functions.* Messenger Math. 58 (1929) 145–152.

[7] T. Inui, Y. Tanabe, Y. Onodera. *Group theory and its applications in physics*, vol. 78, Springer series in solid-state sciences. Springer (1996) (second, corrected print).

[8] A.W. Marshall, I. Olkin. *Inequalities: theory of majorization and its applications.* Academic Press, London (1979).

[9] R.F. Muirhead. *Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters.* Proc. Edinburgh Math. Soc., 21 (1903) 144–157.

[10] R. Rado. *An inequality.* J. Lond. Math. Soc., 27 (1952) 1–6.

[11] H.L. Royden. *Real analysis.* Macmillan, New York, 3rd edn (1988).

[12] J.V. Ryff. *On Muirhead's theorem.* Pacific J. Math., **21**(3) (1967) 567–576.

[13] J.-P. Serre. *Linear representations of finite groups.* Springer, Berlin (1977).

[14] B. Simon. *Representations of finite and compact groups*, vol. 10, Graduate Studies in Mathematics. American Mathematical Society (1996)

[15] R. Stanley. *Enumerative combinatorics.* Cambridge University Press, Cambridge (1999).

Leonard J. Schulman
Caltech
MC 256-80
Pasadena CA 91125
USA
e-mail: schulman@caltech.edu

Received 13 September 2007; accepted 15 August 2008