A LITTLE TASTE OF SYMPLECTIC GEOMETRY: THE SCHUR-HORN THEOREM

TIMOTHY E. GOLDBERG

ABSTRACT. This is a handout for a talk given at Bard College on Tuesday, 1 May 2007 by the author. It gives careful versions of some of the basic definitions from symplectic geometry, describes the Atiyah/Guillemin-Sternberg and Schur-Horn theorems, and gives an elementary proof of the latter theorem for the case $n = 2$.

CONTENTS

Section 2 is somewhat technical, but an attempt has been made to make the rest of this document accessible to a wider audience of math students.

1. STATEMENT OF THE SCHUR-HORN THEOREM

Definition 1.1. Let V be a finite-dimensional real vector space. A subset $C \subset V$ is **convex** if the line segment between any two points in C is itself contained in C. For any subset $X \subset V$, the **convex hull** of X is the minimal convex set C in V that contains X. In this case we say that C is the convex set generated by X. A subset $P \subset V$ is a **convex polytope** if it is the convex hull of a finite set of points. $\triangle$

Let $\mathcal{H}(n)$ be the set of $n \times n$ Hermitian matrices, i.e. $n \times n$ complex matrices A such that $A^* = A$, where $A^* := \bar{A}^T$ is the adjoint of A. Recall from linear algebra that every Hermitian matrix:

- is diagonalizable;
- has all real eigenvalues;
- has all real diagonal entries.

Let $\lambda = (\lambda_1, \dots, \lambda_n)$ be a list of n real numbers, and let $\mathcal{O}_\lambda$ be the subset of $\mathcal{H}(n)$ consisting of Hermitian matrices whose eigenvalues are given by λ. Such a subset is sometimes called an **isospectral set**.

Theorem 1.2 (Schur-Horn). *Let $f\colon \mathcal{O}_\lambda \to \mathbb{R}^n$ be defined by*

$$f(A) = \begin{pmatrix} a_{11} \\ \vdots \\ a_{nn} \end{pmatrix},$$

where $a_{11}, \dots, a_{nn}$ are the diagonal entries of A. Then $f(\mathcal{O}_\lambda)$ is the convex polytope generated by the vectors in $\mathbb{R}^n$ whose entries are precisely $\lambda_1, \dots, \lambda_n$, written is some order.

Notice that the points that generate the convex polytope all satisfy the equation

$$x_1 + \ldots + x_n = \sum_{i=1}^{n} \lambda_i,$$

so the polytope itself lies in the hyperplane of $\mathbb{R}^n$ corresponding to this equation.

Example 1.3. Let $\lambda = (3, 2, 1)$. Then if we take every Hermitian matrix with these eigenvalues, then pluck off their diagonal entries and map them in $\mathbb{R}^3$, we will fill out the hexagon whose vertices are the six points

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

(See Figure 1.) This hexagon lies in the plane $x + y + z = 6$.

$\diamondsuit$

This theorem is attributed to I. Schur and A. Horn. In 1970, B. Kostant showed that this is a special case of a more general theorem about compact Lie groups. In 1982, M. Atiyah and, independently, V. Guillemin and S. Sternberg showed that Kostant's result is a special case of a still more general theorem from symplectic geometry involving Hamiltonian manifolds.

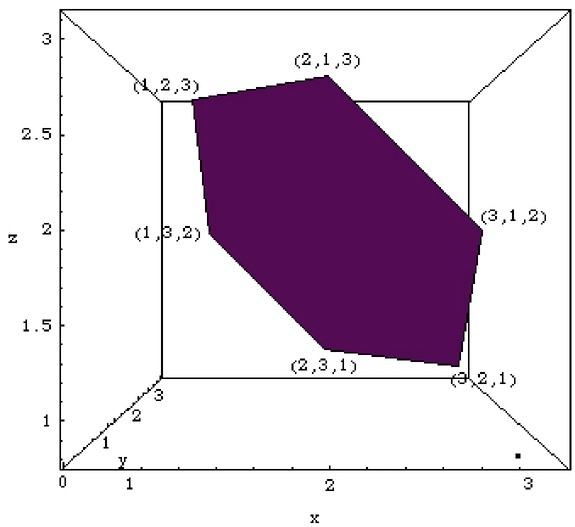

FIGURE 1. The hexagon from Example 1.3

2. CAREFUL DEFINITIONS

Let M be a smooth manifold, and let $\mathrm{Diff}(M)$ denote the set of diffeomorphisms $M \to M$. This forms a group under function composition.

Definition 2.1. An **action** of a Lie group G on M is a group homomorphism $A\colon G \to \mathrm{Diff}(M)$. This action is called **smooth** if the associated map $G \times M \to M, (g, p) \mapsto A(g)(p)$ is smooth. We use the notation $A(g)(p) = g \cdot p$. △

Definition 2.2. A **symplectic form** on M is a differential two-form ω which is closed and nondegenerate, i.e. the exterior derivative of ω vanishes ($d\omega = 0$), and for each $p \in M$, the bilinear form ω_p on the tangent space T_pM is nondegenerate. (This can be thought of as a collection of nondegenerate, skew-symmetric, bilinear forms on each tangent space of M, subject to certain conditions.) The pair (M, ω) is called a **symplectic manifold**.

Given a smooth function $f\colon M \to \mathbb{R}$, the **symplectic gradient** of f is the unique vector field $\nabla_\omega f$ such that, for all $p \in M$ and $\vec{v_p} \in T_pM$ we have

$$D_{\vec{v_p}} f = \omega_p\left(\vec{v_p}, \nabla_\omega f(p)\right),$$

where $D_{\vec{v_p}}$ is the directional derivative at p in the direction $\vec{v_p}$. △

A smooth action of G on M induces a map $\mathfrak{g} \to \mathrm{Vec}(M)$, $\xi \mapsto \xi_M$, where $\mathfrak{g} = T_eG$ is the **Lie algebra** of G and $\mathrm{Vec}(M)$ is the set of smooth tangent vector fields on M. Intuitively, this is because the identity e of G corresponds to the identity map on

M, so an infinitesimal displacement from e in G corresponds to an infinitesimal displacement from each element of M. An infinitesimal displacement is essentially just a tangent vector. The vector field ξ_M is often called the **fundamental vector field** corresponding to ξ.

Definition 2.3. A **moment map** for the action of G on (M, ω) is a function $\Phi\colon M \to \mathfrak{g}^*$ such that Φ is G-equivariant and, for all $\xi \in \mathfrak{g}$,

$$\nabla_\omega(\Phi^\xi) = \xi_M,$$

where $\Phi^\xi\colon M \to \mathbb{R}$ is defined by $\Phi^\xi(p) = \Phi(p)(\xi)$ for all $p \in M$. By G-*equivariant*, we mean that

$$\Phi(g \cdot p) = g \cdot \Phi(p)$$

for all $g \in G$ and $p \in P$, where $g \cdot \Phi(p)$ is an instance of the coadjoint action of G on the dual of its Lie algebra, $\mathfrak{g}^*$. If a moment map for the action of G on M exists, then the action is called **Hamiltonian**, and M is called a **Hamiltonian** G**-manifold**. △

Example 2.4. Every Lie group G acts on its Lie algebra $\mathfrak{g}$, and this action induces an action on the dual of the Lie algebra, $\mathfrak{g}^*$. The coadjoint orbit of G through the element $\lambda \in \mathfrak{g}^*$ is the subset $\mathcal{O}_\lambda \subset \mathfrak{g}^*$ defined by

$$\mathcal{O}_\lambda = \{g \cdot \lambda \mid g \in G\}.$$

It is known that every **coadjoint orbit** can be given the structure of a symplectic manifold, canonically, and of course G acts on it by the coadjoint action. This action of G on $\mathcal{O}_\lambda$ is actually Hamiltonian, with a moment map given by the inclusion $\mathcal{O}_\lambda \hookrightarrow \mathfrak{g}^*$. ♢

Definition 2.5. A **torus** is a compact, connected, abelian Lie group. Equivalently (although not obviously nor trivially), a torus is a Lie group T which is isomorphic to a product $S_1 \times \ldots \times S^1$ of the circle group with itself some number of times. Here, the isomorphism is required to be a smooth group isomorphism. △

Theorem 2.6 (Atiyah/Guillemin-Sternberg)**.** *Let T be a torus, and let M be a compact and connected Hamiltonian T-manifold with moment map $\Phi\colon M \to \mathfrak{t}^*$, where $\mathfrak{t}$ is the Lie algebra of T. Then $\Phi(M)$ is a convex polytope generated by the image under Φ of the fixed point set M^T of T:*

$$M^T = \{p \in M \mid t \cdot p = p \textit{ for all } t \in T\}.$$

This theorem was considerably strengthened to apply to nonabelian Lie groups by F. Kirwan, and these convexity theorems have since been generalized to many, many, many difference contexts. (For instance, the author has been working to apply this to real Lagrangian subsets of invariant subvarieties of Hamiltonian Kähler manifolds, whatever all of that means.)

3. Symplectic interpretation of the Schur-Horn theorem

Let $\mathbf{U}(n)$ denote the $n \times n$ unitary matrices, i.e. the set of $n \times n$ complex matrices a such that $a^* = a^{-1}$. This is a compact connected Lie group of dimension n^2. Its Lie algebra is denoted $\mathfrak{u}(n)$, and is the set of skew-Hermitian matrices, ($X^* = -X$). The adjoint action of $\mathbf{U}(n)$ on its Lie algebra $\mathfrak{u}(n)$ is simply conjugation:

$$g \cdot X = gXg^{-1}$$

for all $g \in \mathbf{U}(n)$ and $X \in \mathfrak{u}(n)$, where the multiplication on the right hand side is matrix multiplication.

The set $\mathcal{H}(n)$ of Hermitian matrices can be identified with the dual space $\mathfrak{u}(n)^*$ via the linear isomorphism $L\colon \mathcal{H}(n) \to \mathfrak{u}(n)^*$ defined by $L(A)(X) = \mathrm{Tr}(iAX)$, for all $A \in \mathcal{H}(n)$ and $X \in \mathfrak{u}(n)$. This isomorphism is actually $\mathbf{U}(n)$-equivariant, so the coadjoint action of $\mathbf{U}(n)$ on $\mathfrak{u}(n)^*$ induces an action of $\mathbf{U}(n)$ on $\mathcal{H}(n)$, which just turns out to be conjugation.

Let $\lambda = (\lambda_1, \ldots, \lambda_n)$ be a list of n real numbers, and let A_λ denote the diagonal matrix with entries given by λ. Notice that A_λ is a Hermitian matrix. We have the following very powerful fact from linear algebra.

Fact. *Conjugation by unitary matrices is a transitive action on each isospectral set of Hermitian matricies.*

If you unpack this fact, you obtain the following.

- For all $g \in \mathbf{U}(n)$ and $A \in \mathcal{O}_\lambda$, we have $gXg^{-1} \in \mathcal{O}_\lambda$.
- For all $A \in \mathcal{O}_\lambda$, there exists $g \in \mathbf{U}(n)$ such that $gAg^{-1} = A_\lambda$.
- For all $A, B \in \mathcal{O}_\lambda$, there exists $g \in \mathbf{U}(n)$ such that $gAg^{-1} = B$.

Therefore the coadjoint orbit $\mathcal{O}_{A_\lambda}$ of $\mathbf{U}(n)$ through A_λ is equal to the isospectral set $\mathcal{O}_\lambda$ of Hermitian matrices with eigenvalues given by λ. Hence $\mathcal{O}_\lambda$ is a Hamiltonian $\mathbf{U}(n)$-manifold.

There is a very nice torus T sitting inside $\mathbf{U}(n)$ as a subgroup. Let T be the set of diagonal complex matrices $\begin{pmatrix} z_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & z_n \end{pmatrix}$ with $\|z_1\| = \ldots = \|z_n\| = 1$. The Lie algebra $\mathfrak{t}$ of T is the set of diagonal matrices with pure imaginary entries, and its dual $\mathfrak{t}^*$ can be identified with the space of diagonal matrices with real entries, via the linear isomorphism $L\colon \mathcal{H}(n) \to \mathfrak{u}(n)^*$ defined above. The inclusion map $\mathfrak{t} \hookrightarrow \mathfrak{u}(n)$ induces a map $\Phi\colon \mathcal{H}(n) \cong \mathfrak{u}(n)^* \to \mathfrak{t}^*$ which projects a matrix A to the diagonal matrix with the same diagonal entries as A. The isospectral set $\mathcal{O}_\lambda$ is a

Hamiltonian T-manifold, and the restriction of Φ to this set is a moment map for the action. If we each diagonal matrix with its diagonal (thought of as a vector in $\mathbb{R}^n$), then this moment map is exactly the function $f\colon \mathcal{O}_\lambda \to \mathbb{R}^n$ from the Schur-Horn theorem.

The Atiyah/Guillemin-Sternberg theorem tells us that $f(\mathcal{O}_\lambda)$ is a convex polytope. Because T is a group of diagonal matrices and T acts on $\mathcal{O}_\lambda$ by conjugation, it can be shown that a matrix $A \in \mathcal{H}(n)$ is fixed under conjugation by every element of T if and only if A is diagonal. Because the eigenvalues of a diagonal matrix are simply its diagonal entries, the only diagonal matrices in $\mathcal{O}_\lambda$ are those with diagonal entries $\lambda_1, \ldots, \lambda_n$, in some order. Therefore these are exactly the elements of the fixed point set $\mathcal{O}_\lambda^T$, so the diagonals of these matrices generate the convex polytope $f(\mathcal{O}_\lambda)$. This is exactly the statement of the Schur-Horn theorem.

4. Proof of the Schur-Horn theorem for $n = 2$

An arbitrary 2×2 Hermitian matrix is of the form $A = \begin{pmatrix} a & c+di \\ c-di & b \end{pmatrix}$. The eigenvalues of A are the roots of its characteristic polynomial:

$$\det\left[\begin{pmatrix} a & c+di \\ c-di & b \end{pmatrix} - x\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] = \det\begin{pmatrix} a-x & c+di \\ c-di & b-x \end{pmatrix}$$
$$= (a-x)(b-x) - (c+di)(c-di)$$
$$= x^2 - (a+b)x + (ab - c^2 - d^2).$$

Since the leading coefficient of this quadratic is 1, the sum of its roots is $(a+b)$ and the product of its roots is $(ab - c^2 - d^2)$.

Let $\lambda = (\lambda_1, \lambda_2)$ be a pair of real numbers. If $A \in \mathcal{O}_\lambda$ then

$$\lambda_1 + \lambda_2 = a + b \quad \text{and} \quad \lambda_1\lambda_2 = ab - c^2 - d^2.$$

Hence $a + b = \lambda_1 + \lambda_2$ and $ab \geq \lambda_1\lambda_2$. Define $\alpha, \beta\colon \mathbb{R}^2 \to \mathbb{R}$ by $\alpha(x,y) = x + y$ and $\beta(x,y) = xy$ for $(x,y) \in \mathbb{R}^2$. Then our equations can be expressed as

$$\alpha(a,b) = \alpha(\lambda_1,\lambda_2) \quad \text{and} \quad \beta(a,b) \geq \beta(\lambda_1,\lambda_2).$$

Note that (λ_1, λ_2) is contained in the following intersection of solution sets

$$\left\{(x,y) \in \mathbb{R}^2 \mid \alpha(x,y) = \lambda_1 + \lambda_2\right\} \cap \left\{(x,y) \in \mathbb{R}^2 \mid \beta(x,y) = \lambda_1\lambda_2\right\},$$

and (a,b) is contained in the intersection

$$\left\{(x,y) \in \mathbb{R}^2 \mid \alpha(x,y) = \lambda_1 + \lambda_2\right\} \cap \left\{(x,y) \in \mathbb{R}^2 \mid \beta(x,y) \geq \lambda_1\lambda_2\right\}.$$

By plotting these curves, we see that the latter intersection is exactly the line segment between the points (λ_1, λ_2) and (λ_2, λ_1). (Figure 2 shows this plot under the assumption $0 \leq \lambda_1 \leq \lambda_2$. For each other case there is an entirely analogous picture.) Of course, this line segment is exactly the convex hull of these two points. This proves the Schur-Horn theorem for $n = 2$.

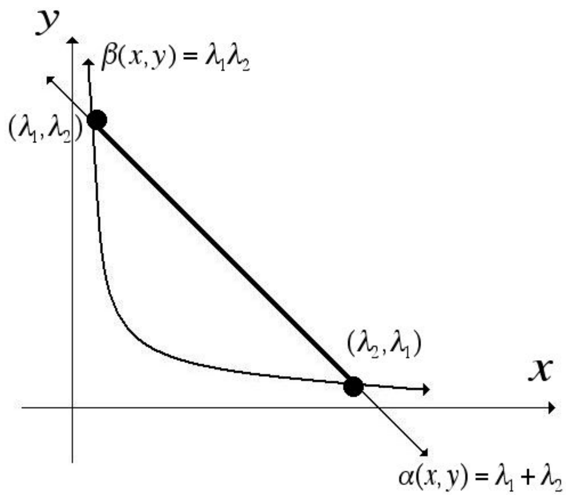

FIGURE 2. The dark line segment must contain the point (a, b).

Remark 4.1. Note that if $\lambda_1 = \lambda_2$, then the line segment is actually just a point. So in this case, both diagonal entries of every matrix in $\mathcal{O}_\lambda$ are $\lambda_1 = \lambda_2$. $\diamondsuit$

Timothy E. Goldberg
graduate student
Department of Mathematics
Cornell University
Ithaca, New York
goldberg@math.cornell.edu