Invent. math. 113, 511–529 (1993)

Inventiones mathematicae

A Schur-Horn-Kostant convexity theorem for the diffeomorphism group of the annulus

A.M. Bloch[1], **H. Flaschka**[2], **and T. Ratiu**[3]

[1] Department of Mathematics, The Ohio State University, Columbus, OH 43210, USA
[2] Department of Mathematics, The University of Arizona, Tucson AZ 85721, USA
[3] Department of Mathematics, University of California, Santa Cruz, CA 95064, USA

Oblatum 25-XI-1992 & 17-III-1993

Summary. The group of area preserving diffeomorphisms of the annulus acts on its Lie algebra, the globally Hamiltonian vectorfields on the annulus. We consider a certain Hilbert space completion of this group (thinking of it as a group of unitary operators induced by the diffeomorphisms), and prove that the projection of an adjoint orbit onto a "Cartan" subalgebra isomorphic to $L^2([0, 1])$ is an infinite-dimensional, weakly compact, convex set, whose extreme points coincide with the orbit, through a certain function, of the "permutation" semigroup of measure preserving transformations of $[0, 1]$.

1 Introduction

Diffeomorphism groups are huge, infinite-dimensional Lie groups, but in some respects they can be strikingly similar to finite-dimensional semisimple groups. Certain analogies remain, temporarily, at the formal level – for instance, the "continuous root systems" introduced by Saveliev and Vershik [SV]. Others, such as the decomposition of unitary representations under the action of a permutation group [VGG], are completely rigorous. Our paper presents a new rigorous point of coincidence between semisimple and diffeomorphism groups: an infinite-dimensional version of the $SU(n)$ Schur-Horn-Kostant convexity theorem. We prove such a result for a completion of the group $\mathrm{SDiff}(\mathscr{A})$ of area (but not necessarily orientation) preserving diffeomorphisms of the annulus

$$\mathscr{A} \overset{\text{def}}{=} \{0 \leqq z \leqq 1\} \times \{\exp(2\pi i\theta) | 0 \leqq \theta < 1\};$$

this group appears to be a particularly natural infinite-dimensional analog of $SU(n)$.

---

* Dedicated to Professor Takeshi Kotake on the occasion of his 60th birthday

Nowadays, the Schur-Horn theorem is considered to be an easy corollary of the Atiyah-Guillemin-Sternberg theorem about the convexity of the image of the momentum map. Our tools are more primitive. Thinking of SDiff($\mathscr{A}$) as a version of "$SU(\infty)$", we establish a generalization of Schur's theorem and a generalization of Horn's theorem, and place them both in a setting reminiscent of Kostant's Lie-theoretic version of those classical results. As just mentioned (and we explain this in more detail later in the introduction), our result requires the completion of SDiff($\mathscr{A}$) to the group of invertible measure preserving transformations, and even beyond – to the semigroup of not necessarily invertible measure preserving maps on $\mathscr{A}$.

A second interpretation of our main result involves the dispersionless Toda lattice, but this is described later. We begin with a review of the Schur, Horn, and Kostant theorems.

Notation. Let $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$.
a) $\sum_n x$ is the orbit of x under the symmetric group on n letters, i.e. the collection of points $(x_{s(1)}, \ldots, x_{s(n)})$, where s ranges over all $n!$ permutations.
b) For $C \subset \mathbf{R}^n$, $\hat{C}$ denotes the convex hull of C.

Schur's theorem [Sch] *Let A be a hermitean matrix with eigenvalues λ_j (arranged in non-increasing order[1]). Let $\lambda = (\lambda_1, \ldots, \lambda_n)$, and let $A^0 = (A_{11}, \ldots, A_{nn})$ be the diagonal of A. Then*

$$A^0 \in \widehat{\textstyle\sum_n \lambda}\,.$$

Horn's theorem [H] *Let $\lambda \in \mathbf{R}^n$, with components arranged in non-increasing order. If $A^0 \in \widehat{\sum_n \lambda}$, there is a hermitean matrix A with eigenvalues λ whose diagonal is A^0.*

The translation into Lie theory goes as follows. Let $G = SU(n)$, and let T be the subgroup of diagonal matrices. Their Lie algebras are denoted by $\mathscr{G}$ and $\mathscr{T}$, respectively; we identify the algebra $\mathscr{G}$ of skew-hermitean matrices with the set of hermitean ones via multiplication by $\sqrt{-1}$. Let

$$W \overset{\text{def}}{=} N(T)/T$$

(where $N(T)$ is the normalizer of T) be the Weyl group, which in this case is just the symmetric group. Finally, for $\lambda \in \mathscr{T}$, let $\mathcal{O}_\lambda$ denote the orbit $\{g\lambda g^{-1} | g \in G\}$ of hermitean matrices with fixed eigenvalues λ_j.

Kostant's theorem [K] *Let π be the projection from $\mathscr{G}$ to $\mathscr{T}$ that sends a matrix to its diagonal. Then $\pi(\mathcal{O}_\lambda)$ is the convex hull of its extreme points, which coincide with the Weyl group orbit $W \cdot \lambda$.*

This result applies to all compact Lie groups G, and provides a far-reaching generalization of the Schur-Horn theorems. We will state our infinite-dimensional Schur-Horn theorem in Kostant's language. It should be noted that even for $\mathscr{G} = su(n)$, Kostant's version is a nontrivial addition to Schur-Horn: it asserts that all matrices with given eigenvalues λ_j lie on a single orbit of the group $SU(n)$. This

[1] The order plays no role in the finite-dimensional setting, but it is crucial in the infinite-dimensional generalization. It is therefore best to introduce it at this early stage

may be undergraduate linear algebra, but the corresponding result in infinite dimensions is not so obvious any more.

The group G is replaced by $\mathrm{SDiff}(\mathscr{A})$, defined above. The Lie algebra $\mathscr{G}$ of $\mathrm{SDiff}(\mathscr{A})$ is the algebra of divergence-free vector fields tangent to the boundary of $\mathscr{A}$. These vector fields are Hamiltonian with respect to the area from $\omega = dz \wedge d\theta$ and their Hamiltonian functions $x(z, \theta)$ satisfy $\partial x(z_0, \theta)/\partial\theta = 0$ for $z_0 = 0, 1$.

Convention. In the subsequent development, we shall identify the Lie algebra $\mathscr{G}$ with the Poisson algebra $\mathscr{P}$ of functions obeying those boundary conditions. The two algebras are in fact not the same: $\mathscr{P}$ is a trivial central extension of $\mathscr{G}$, or equivalently, $\mathscr{G} \cong \mathscr{P}/\{\text{constant functions}\}$. It is easier to work with $\mathscr{P}$, and by referring to it as the Lie algebra of $\mathrm{SDiff}(\mathscr{A})$ we intend to reinforce analogies with the finite-dimensional convexity theorem. At the very end of the paper, we will point out that a convexity theorem for $\mathscr{G}$ follows immediately, by taking quotients, from our convexity theorem for $\mathscr{P}$.

The maximal torus T_0 of $\mathrm{SDiff}(\mathscr{A})$ is the subgroup of *pure twist maps,*

$$(z, \theta) \mapsto (z, \theta + \phi(z)) ; \tag{1.1}$$

see §2 for a justification of this choice for T_0. (The subscript zero is used to distinguish this torus from a certain completion introduced below.) The "diagonal" subalgebra $\mathscr{T}_0$ is, in accordance with our convention, identified with the Hamiltonians that depend only on z. "Projection onto the diagonal" means:

$$\pi : f(z, \theta) \mapsto \int_0^1 f(z, \theta)\, d\theta . \tag{1.2}$$

Morally, we prove the following result.

Desideratum. *Let $\lambda \in \mathscr{T}_0$ be a non-increasing function of z, and let $\mathscr{O}_\lambda$ be the orbit of* $\mathrm{SDiff}(\mathscr{A})$ *through λ. The image $\pi(\mathscr{O}_\lambda)$ is the convex hull of its extreme points, which coincide with the orbit of the Weyl group $N(T_0)/T_0$.*

In fact, this very attractive statement is wrong because, as we show in §2, the Weyl group $N(T_0)/T_0$ has only two elements. Only after we pass to a completion of $\mathrm{SDiff}(\mathscr{A})$, in order to produce a nontrivial Weyl group, will we be able to prove a reasonable convexity theorem.

We next describe this completion of $\mathrm{SDiff}(\mathscr{A})$. The adjoint representation of $G = \mathrm{SDiff}(\mathscr{A})$ on its Lie algebra $\mathscr{G}$ is then the map[2] $P_g : F \mapsto F \circ g$ for $g \in \mathrm{SDiff}(\mathscr{A})$ and $F \in \mathscr{G}$. The operators P_g extend to unitary operators on $L^2(\mathscr{A})$. The strong operator topology induces a topology on $\mathrm{SDiff}(\mathscr{A})$. In this topology, $\mathrm{SDiff}(\mathscr{A})$ is dense in the group $\mathrm{SMeas}(\mathscr{A})$ of invertible measure preserving transformations of the annulus, and in fact in the semigroup $\overline{\mathrm{SMeas}}(\mathscr{A})$ of not necessarily invertible measure preserving transformations. The latter corresponds to the strong closure of the subgroup $\{P_g \mid g \in \mathrm{SDiff}(\mathscr{A})\}$ of the group of unitary operators on $L^2(\mathscr{A})$. The proof requires no work on our part: two of the approximation results come directly from the literature [Al, Br], and the third was communicated to us by Jürgen Moser.

[2] This is the *right* action of the diffeomorphism group, which – see [EM] – is differentiable; the left action $F \mapsto F \circ g^{-1}$ is not

It is known [Ha1] that $\mathrm{SMeas}(\mathscr{A})$ is a complete topological group. The subgroup T of measurable pure twist maps (maps of the form (1.1) with measurable ϕ) is shown to be maximal abelian; it plays the role of maximal torus of $\mathrm{SMeas}(\mathscr{A})$. We prove that the "Weyl group" $W = N(T)/T$ can be identified with the group of maps

$$(z, \theta) \mapsto (a(z), j(z)\theta)\,,$$

where $a(z)$ is invertible measure preserving on the interval $0 \leqq z \leqq 1$, and $j(z) = \pm 1$ almost everywhere. The strong closure of W, in which $a(z)$ is merely measure preserving, is denoted by $\bar{W}$.

Finally, we replace the Lie algebra $\mathscr{G}$ by its completion $L^2(\mathscr{A})$. The closure in $L^2(\mathscr{A})$ of the Lie algebra $\mathscr{T}_0$ of the torus is the subspace $\mathscr{T}$ of functions that are independent of θ. It is convenient to identify it with $L^2([0, 1])$. The Weyl semigroup $\bar{W}$ acts on $\mathscr{T}$ only through the first component $a(z)$; it is, so to say, the permutation group of the interval $0 \leqq z \leqq 1$. The projection (1.2), which is an orthogonal projection in Hillbert space, is the same for the measurable case as for the smooth case.

Convexity theorem. *Let $\lambda \in L^2(\mathscr{A})$ be a bounded, nonincreasing, right continuous function of z alone, and let $\mathcal{O}_\lambda$ be the orbit of $\overline{\mathrm{SMeas}(\mathscr{A})}$ through λ. Then $\pi(\mathcal{O}_\lambda) \subset L^2([0, 1])$ is a weakly compact, convex set. Its set of extreme points is the orbit $\bar{W} \cdot \lambda$ of the Weyl semigroup through λ.*

Remark. There are convexity theorems about the action of finite-rank tori contained in infinite-dimensional groups (see [AP] and [KP], for instance). In those situations, the Cartan subalgebra is finite-dimensional, and the momentum polytope is therefore an unbounded convex subset of a finite-dimensional Euclidean space. Our Cartan subalgebra, however, is an infinite-dimensional function space, and the analog of the momentum polytope is an infinite-dimensional convex set. In addition, $\overline{\mathrm{SMeas}(\mathscr{A})}$ is neither a group nor (as far as we know) a manifold, so the finite-dimensional Lie-theoretic or symplectic proofs are unlikely to generalize.

Working with orbits of the semigroup of not necessarily invertible measure-preserving transformations requires a little care. For example, if $x = \lambda \circ g$ for $g \in \overline{\mathrm{SMeas}(\mathscr{A})}$ and λ is as in the statement of the theorem, there need not be any $h \in \overline{\mathrm{SMeas}(\mathscr{A})}$ for which $\lambda = x \circ h$ (see Example 5.6). For this reason, we speak of "the orbit through λ", rather than just of "an orbit". Orbits through the bounded, nonincreasing functions λ of z alone have a simple characterization:

Orbit theorem. *Let $x \in L^2(\mathscr{A}) \cap L^\infty(\mathscr{A})$. There is a unique (necessarily bounded) nonincreasing, right continuous function λ of z alone such that $x \in \mathcal{O}_\lambda$. The orbit $\mathcal{O}_\lambda$ consists of all functions for which*

$$\int_{\mathscr{A}} y^p = \int_0^1 \lambda(z)^p\, dz \left(= \int_{\mathscr{A}} x^p \right), \quad p = 1, 2, \ldots . \tag{1.3}$$

(We ask that $x \in L^\infty$ because all moments of x must exist for our setting to make sense.) If one thinks of the set of nonincreasing functions of z alone as the positive Weyl chamber in the Cartan subalgebra $\mathscr{T}$, the first part of this theorem asserts that every maximal orbit intersects the positive Weyl chamber in a unique point. The function λ is the continuous analog of a diagonal matrix; condition (1.3) can be interpreted as saying that the traces of powers of the diagonal matrix λ equal the traces of powers of the "matrix" x. To explain how a function $x(z, \theta)$ is analogous to a matrix, and to provide additional perspective and motivation, we now outline a different approach to the convexity theorem. It was this line of thought that led us to our result.

Define the following difference operators on the space of doubly infinite square summable sequences:

$$L = a_n e^{\partial/\partial n} + b_n + e^{-\partial/\partial n} a_n ,$$

$$B = a_n e^{\partial/\partial n} - e^{-\partial/\partial n} a_n ,$$

where $e^{\partial/\partial n}$ and $e^{-\partial/\partial n}$ denote forward and backward shift, $(e^{\partial/\partial n} y)_n = y_{n+1}$, $(e^{-\partial/\partial n} y)_n = y_{n-1}$. Suppose the coefficients a_n, b_n depend on a variable t. The Lax equation

$$\frac{dL}{dt} = [B, L]$$

is equivalent to the system

$$\dot{a}_n = a_n(b_{n+1} - b_n), \quad \dot{b}_n = 2(a_n^2 - a_{n-1}^2) . \tag{1.4}$$

These are the *Toda lattice* equations[3] [T]. They are (for suitable boundary conditions on a_n, b_n) a completely integrable Hamiltonian system. The Poisson commuting constants of motion are given by $\operatorname{Tr} L^p, p = 1, 2, \ldots$ (the traces are only defined after appropriate regularization).

Let us now take the continuum limit of the Toda lattice equations. The shift $e^{\partial/\partial n}$ is replaced by $e^{\varepsilon\partial/\partial z}$, and ε tends to zero. The scaling is adjusted so that the continuous variable z takes values in $[0, 1]$. The functions $a_n(t), b_n(t)$ are replaced by functions $v(z, t), u(z, t)$ of two variables, and the Eqs. (1.4) become

$$\frac{\partial v}{\partial t} = v \frac{\partial u}{\partial z}, \quad \frac{\partial u}{\partial t} = 2 \frac{\partial v^2}{\partial z} . \tag{1.5}$$

This is a quasilinear hyperbolic system, called the *dispersionless Toda equations* [Bo, BB, BBKR, Sa1, Sa2]. The naive continuum limits I_p of the constants of motion $\operatorname{Tr} L^p$ of (1.4) give constants of motion for (1.5). For example,

$$\operatorname{Tr} L^2 = \sum a_n^2 + a_{n-1}^2 + b_n^2$$

becomes

$$I_2 = \int_0^1 2v(z)^2 + u(z)^2 \, dz .$$

Brockett and Bloch [BB] show that

$$I_p = \int_0^1 \int_0^1 (u(z) + v(z) e^{2\pi i \theta} + v(z) e^{-2\pi i \theta})^p \, dz \, d\theta .$$

---

[3] Historical note: in his lecture at an April 1991 workshop at the Fields Institute in Waterloo, Ontario, David Watkins explained that those equations had been derived in the context of QR and related algorithms, and then solved, by the numerical analyst Heinz Rutishauser in the 1950's; see [WE]

We think of $(u(z) + v(z)e^{2\pi i\theta} + v(z)e^{-2\pi i\theta})$ as the continuous analog of a tridiagonal matrix. The exponentials $\exp(\pm 2\pi i\theta)$ label the first super- and sub-diagonals. The variable z parametrizes the diagonal direction.

It may now seem reasonable to consider a series

$$x(z, \theta) = \sum_{-\infty}^{\infty} x_n(z)e^{2\pi n i\theta}$$

the analog of a general matrix, and to think of the condition $x_{-n} = \overline{x_n}$ as defining a "continuous" hermitean matrix. In the L^2 sense, such a series determines a real-valued function of (z, θ); diagonal matrices are functions of z alone. Fixing the eigenvalues of such a matrix $x(z, \theta)$ would be equivalent to prescribing the moments I_p. The second part of the orbit theorem justifies this interpretation; the first half says only that a hermitean matrix can be diagonalized. (For the tridiagonal case, $x(z, \theta) = u(z) + 2v(z)\cos 2\pi\theta$, this is shown in [BB]; our proof in §4 appeals to Hausdorff's solution of the moment problem.)

The analogy between the Toda lattice and the dispersionless Toda system can be pushed quite far. For instance, if we now use L to denote the continuous tridiagonal matrix $u(z) + 2v(z)\cos 2\pi\theta$, then (1.5) can be written as a *Brockett double bracket equation* [Bro]

$$\frac{dL}{dt} = \{L, \{L, z\}\},$$

where $\{\cdot, \cdot\}$ is the Poisson bracket for the symplectic form $dz \wedge d\theta$ on $\mathscr{A}$ (we learned this from L. Faybusovich in June 1990). The interplay between gradient flows, sorting, convexity, and the double bracket equation established for the *finite* Toda lattice equation in [BBR, Bro] carries over to the dispersionless Toda system, at least at a formal level – since (1.5) is quasilinear hyperbolic, solutions will in general develop shocks, and the analytic meaning of various formal properties is still murky. We hope to return to those problems in another paper.

We have tried to provide motivation for the convexity theorem; something should be said about the proof. It is modeled on the pre-Kostant, pre-symplectic proofs of the Schur-Horn theorems. The basic tools are the theory of majorization and the notion of a doubly stochastic matrix, as generalized to function space by Ryff [R1, R2, R3] and Brown [Br]. The arguments are not difficult once the work of the authors just cited is taken into account; we are simply putting an established theory to novel use. All the finite-dimensional facts can be found in the encyclopaedic monograph by Marshall and Olkin [MO]; we give a very brief summary.

(1.6) Definition [HLP] For $x \in \mathbf{R}^n$, let x^* denote the vector obtained by rearranging the components of x in nonincreasing order. We say that y *majorizes* x, written $x \prec y$, if

$$x_1^* + \ldots + x_k^* \leqq y_k^* + \ldots y_k^*, \quad \text{for } 1 \leqq k \leqq n-1,$$

$$\sum_{j=1}^{n} x_j^* = \sum_{j=1}^{n} y_j^*.$$

(1.7) Definition. An $n \times n$ real matrix P is called *doubly stochastic* if $P_{ij} \geqq 0$, and the sum of each row and column is 1.

(1.8) Theorem. a) [Mo] *P is doubly stochastic if, and only if, $Pe = e$ and $e'P = e'$, where e is the column vector all of whose entries are 1, and e' is its transpose.*
b) [MO] *P is doubly stochastic if, and only if, $Px \prec x$ for all $x \in \mathbf{R}^n$.*
c) [HLP] *$x \prec y$ if, and only if, there is a doubly stochastic P such that $x = Py$.*
d) [Bi] *The set of doubly stochastic matrices is the convex hull of its extreme points, which are precisely the permutation matrices.*

(1.9) Corollary. $\{x | x \prec y\} = \widehat{\sum_n y}$.

The Schur theorem now follows easily. Diagonalize the hermitean matrix A, $A = Q\lambda Q^*$, Q unitary. Then $A_{ii} = \sum_j |Q_{ij}|^2 \lambda_j$. If Q is unitary, the matrix $P_{ij} \stackrel{\text{def}}{=} |Q_{ij}|^2$ must be doubly stochastic. Corollary 1.9 then gives the conclusion. Horn proved the converse by a rather intricate argument, deducing that when $x \prec \lambda$, there must be a doubly stochastic P of the form $P_{ij} = |Q_{ij}|^2$, with Q unitary, satisfying $x = P\lambda$; $Q\lambda Q^*$ is then the desired hermitean A having eigenvalues λ_j and diagonal x.

We will use infinite-dimensional versions of parts of Theorem 1.8. Our proof of the Schur theorem is closely related to the argument just sketched. Our proof of the Horn part, if pushed further than necessary for our purposes, could also be seen as analog of the finite-dimensional argument.

(1.10) Definition [HLP, R1] Let $f \in L^1([0, 1])$. Set $m(y) = |\{z | f(z) > y\}|$ (absolute value denotes Lebesgue measure on [0, 1]) and, for $0 \leqq z < 1$, set

$$f^*(z) = \sup\{y | m(y) > z\} .$$

The nonincreasing, right continuous function f^* is called the *nonincreasing rearrangement* of f.

(1.11) Definition [HLP,R1] Let $f, g \in L^1([0, 1])$. We say that f *majorizes* g (written $g \prec f$) if

$$\int_0^s g^*(z)\,dz \leqq \int_0^s f^*(z)\,dz, \quad 0 \leqq s < 1 ,$$

$$\int_0^1 g^*(z)\,dz = \int_0^1 f^*(z)\,dz .$$

(1.12) Definition [R1] A linear operator P on $L^1([0, 1])$ is called *doubly stochastic* if $Pf \prec f$ for all $f \in L^1([0, 1])$.

(1.13) Theorem [R2, R3] *In $L^1([0, 1])$: $g \prec f$ if, and only if, there is a doubly stochastic P such that $g = Pf$. The set $\Omega(f) \stackrel{\text{def}}{=} \{g | g \prec f\}$ is weakly compact and convex. Its set of extreme points is $\{f^* \circ \phi | \phi$ is a measure preserving transformation of $[0, 1]\}$.*

Here is an outline of our proof. The orbit theorem comes first: given $x(z, \theta)$, there is a unique (nonincreasing, right-continuous) $\lambda(z)$ defined on [0, 1], with the same moments. This fact is trivial, granted Hausdorff's solution of the moment problem. Now let $\lambda_2(z, \theta)$ be the function of two variables defined by $\lambda_2(z, \theta) = \lambda(z)$ (so λ_2 is constant on vertical lines in the square $\{0 \leqq z \leqq 1, 0 \leqq \theta \leqq 1\}$). We then verify from [Ha2] that a measure preserving transformation of $\mathscr{A}$ takes λ_2 to x. This, as we will see, immediately implies that the diagonal part $\pi(x)$ of x is obtained from

λ by the action of a doubly stochastic operator – this gives Schur's theorem. For the Horn part, we assume that $X \prec \lambda$, which, according to (1.13), means that $X = P\lambda$ for a doubly stochastic P. From this P, we build an $x(z, \theta)$ with the same moments as λ, for which $\pi(x) = X$.

In §2 we state the basic facts about SDiff($\mathscr{A}$) and SMeas($\mathscr{A}$). The rest of the paper is broken up into small pieces, for the reader's convenience. The approximation theorems are collected in §3, the orbit theorem is proved in §4, Schur's theorem in §5, and Horn's theorem in §6.

2 SDiff and SMeas

We first collect information about the groups SDiff($\mathscr{A}$) and SMeas($\mathscr{A}$). This section and the next one are logically independent of §§5–6, where the convexity theorems are proved. Those results are straight functional analysis, and SDiff($\mathscr{A}$) or SMeas($\mathscr{A}$) do not help with the proofs. We feel, however, that our results appear natural (and interesting) only in the group-theoretic setting.

We begin with the smooth case; more details can be found in [BR]. SDiff($\mathscr{A}$) is the group of C^∞ area preserving diffeomorphisms of the annulus

$$\mathscr{A} \stackrel{\text{def}}{=} \{0 \leqq z \leqq 1\} \times \{\exp(2\pi i\theta) | 0 \leqq \theta < 1\}$$

(more generally, one could consider Sobolev maps in H^s for some $s > 2$). Its Lie algebra $\mathscr{G}$, according to the convention set down in Sect. 1, is identified with the Poisson algebra of functions x satisfying

$$\frac{\partial x}{\partial \theta}(z_0, \theta) \equiv 0, \quad z_0 = 0, 1 .$$

The Hamiltonian vector field X_x will then be tangent to the boundary.

Let $\mathscr{T}_0$ denote the subalgebra of θ-independent functions. A function $f \in \mathscr{T}_0$ generates the vector field

$$\dot{z} = \frac{\partial f}{\partial \theta} = 0, \quad \dot{\theta} = -\frac{\partial f}{\partial z},$$

which integrates to give the symplectic twist map

$$(z, \theta) \mapsto \left(z, \theta - t\frac{\partial f(z)}{\partial z}\right) . \tag{2.1}$$

The subgroup T_0 of pure twist maps (2.1) behaves like a maximal torus of SDiff($\mathscr{A}$). The following supporting evidence is presented in [BR]:

Theorem. a) *T_0 is a smooth, H^s-closed, path connected submanifold of* SDiff($\mathscr{A}$); b) *it is a maximal abelian subgroup of* SDiff($\mathscr{A}$), *and its formal Lie algebra is* $\mathscr{T}_0$; c) *it is totally geodesic and flat with respect to the hydrodynamic metric*

$$\langle X, Y \rangle = \int_{\mathscr{A}} X \cdot Y \, d(\text{area}) .$$

In these respects, T_0 mimics a maximal torus of a compact Lie group. Other properties do not survive transition to infinite dimensions. For example, it is no longer true that every element of SDiff$(\mathscr{A})$ can be conjugated to an element of T_0, since most symplectic maps of the annulus do not have a continuum of invariant circles. It is not known whether all maximal abelian subgroups of SDiff$(\mathscr{A})$ are conjugate (it would be surprising if they were). As explained in the introduction, the failure of another finite-dimensional property influences the setting of the convexity theorem:

(2.2) Proposition. *The normalizer $N(T_0)$ of T_0 is the group of twist maps (possibly orientation reversing):*

$$(z,\theta) \mapsto (z, j\theta + \phi(z)), \quad j = +1 \text{ or } -1 .$$

The Weyl group $N(T_0)/T_0$ is the two-element group.

The proof is postponed; it will follow from a more general result for the measurable case.

We next turn to SMeas$(\mathscr{A})$, the group of invertible measure preserving transformations of the annulus. Each $g \in$ SMeas$(\mathscr{A})$ determines a unitary operator P_g on $L^2(\mathscr{A})$ by $P_g x = x \circ g$. The strong operator topology induces a topology on SMeas$(\mathscr{A})$. It is traditionally called the *weak topology*, because the strong and weak operator topologies coincide on unitary operators. Halmos [Ha1] described a basis for this topology, and Alpern [A1] showed it to be the topology determined by the metric

$$\rho(g_1, g_2) \stackrel{\text{def}}{=} \inf\{\mu | m(\{(z,\theta) || g_1(z,\theta) - g_2(z,\theta)| \geqq \mu\}) < \mu\} \tag{2.3}$$

(m denotes planar Lebesgue measure). Evidently, $\rho(g_n, g) \to 0$ if, and only if, the g_n converge to g in measure. It is easy to see that this happens if, and only if, $P_{g_n} \to P_g$ in the strong operator topology.

Let T be the subgroup of pure twist maps, $(z,\theta) \mapsto (z, \theta + \phi(z))$, with measurable ϕ. This subgroup will play the role of maximal torus.

(2.4) Proposition. *The normalizer $N(T)$ of T is the group of maps*

$$(z,\theta) \mapsto (a(z), j(z)\theta + k(z)) ,$$

with a an invertible measure preserving transformation of $[0, 1]$, *j measurable and equal to ± 1 almost everywhere, and k measurable. The "Weyl group" $N(T)/T$ may be identified with the group of maps of the form*

$$(z,\theta) \mapsto (a(z), j(z)\theta) .$$

Proof. We need to find all $g \in$ SMeas$(\mathscr{A})$ with the following property: for every $h \in T$, there is an $\bar{h} \in T$ such that

$$g \circ h = \bar{h} \circ g . \tag{2.5}$$

Write the maps $g, h, \bar{h}$ in components as

$$(a(z,\theta), b(z,\theta)), \quad (z, \theta + \phi(z)), \quad (z, \theta + \bar{\phi}(z)) ,$$

respectively. Condition (2.5) translates to

$$a(z, \theta + \phi(z)) = a(z,\theta) , \tag{2.6.a}$$

$$b(z, \theta + \phi(z)) = b(z,\theta) + \bar{\phi}(a(z,\theta)) . \tag{2.6.b}$$

From (2.6.a) it follows immediately that a is a function of z alone. Rewrite (2.6.b) as

$$b(z, \theta + \phi(z)) - b(z, \theta) = \bar{\phi}(a(z)) .$$

The right side is independent of θ, and therefore $b(z, \theta)$ must be linear in $\theta, b(z, \theta) = j(z)\theta + k(z)$. Equation (2.6.b) then reads

$$j(z)\phi(z) = \bar{\phi}(a(z)) ,$$

which has a solution $\bar{\phi}$ for every measurable ϕ only if a is one-to-one.

So far, we have seen that g has the form

$$(a(z), j(z)\theta + k(z)) ,$$

with the function a being one-to-one. Since the measure of all strips

$$\{z \in E\} \times \{\exp 2\pi i\theta \,|\, \theta \in [0, 1]\}$$

($E \subset [0, 1]$ measurable) must be preserved, a should be a measure preserving transformation of $[0, 1]$. Also, the measure of strips

$$\{z \in [0, 1]\} \times \{\exp 2\pi i\theta \,|\, \theta \in F\}$$

($F \subset [0, 1]$ measurable) is preserved, and so it follows that $j(z) = \pm 1$ almost everywhere.

Finally, if $g_i(z, \theta) = (a_i(z), j_i(z)\theta + k_i(z))$, $i = 1, 2$, one checks easily that $g_1 = g_2 \circ h$ for some $h \in T$ exactly when $a_1 = a_2$ and $j_1 = j_2$. The quotient $N(T)/T$ can therefore be identified with the set of maps described in the statement of the proposition. □

Proof of Proposition 2.2 In the preceding proof, the function a must be smooth, preserve length, and fix the boundaries $z = 0, 1$; thus $a(z) = z$. The function j must be smooth, and therefore identically equal to 1 or -1, so that the "Weyl group" is the two-element group $\{(z, \theta), (z, -\theta)\}$. □

3 Completion of SDiff

The disclaimer at the beginning of the preceding section remains in effect: the next result only places the convexity and orbit theorems in their natural context.

We denote by m the planar Lebesgue measure on the annulus $\mathscr{A}$. The operator on $L^2(\mathscr{A})$ induced by a measure-preserving transformation g of $\mathscr{A}$ is called P_g.

(3.1) Proposition. *The strong closure of* $\{P_\phi \,|\, \phi \in \mathrm{SDiff}(\mathscr{A})\}$ *is* $\{P_g \,|\, g \in \overline{\mathrm{SMeas}(\mathscr{A})}\}$. *Equivalently, the completion of* $\mathrm{SDiff}(\mathscr{A})$ *in the topology given by the metric* (2.3) *is* $\overline{\mathrm{SMeas}(\mathscr{A})}$.

Proof. It is shown in [Br, Theorem 5] that $\{P_g \,|\, g \in \overline{\mathrm{SMeas}(\mathscr{A})}\}$ is the strong closure of $\{P_g \,|\, g \in \mathrm{SMeas}(\mathscr{A})\}$. Therefore, we must prove that every P_g, $g \in \mathrm{SMeas}(\mathscr{A})$, can be approximated by a P_ϕ, $\phi \in \mathrm{SDiff}(\mathscr{A})$. Put differently: an invertible measure preserving map of $\mathscr{A}$ can be approximated (in measure) by a measure preserving diffeomorphism. Alpern [A1] proves that an invertible measure preserving transformation of the *square* can be approximated, in his metric (2.3), by a measure preserving homeomorphism that fixes the boundary of the square (this clearly implies the same approximation for the annulus). It now remains to approximate such a homeomorphism of the square by a measure preserving diffeomorphism, which is a special case of a result due to Moser:

Theorem. *In the metric* (2.3), SDiff($\mathscr{A}$) *is dense in the group of boundary fixing, measure preserving homeomorphisms of the square (and hence there is a similar relation, in the strong operator topology on* $L^2(\mathscr{A})$*, between the corresponding groups of unitary operators).*

The elegant proof of this theorem may be found in the preprint [Mo].

4 Orbit theorem

In this section, we prove a basic theorem of linear algebra for our "continuum matrices": every hermitean matrix can be diagonalized, and all hermitean matrices with the same eigenvalues are conjugate under the unitary group. Planar Lebesgue measure on $\mathscr{A}$ is denoted by m, and linear Lebesgue measure on $0 \leqq z \leqq 1$ and $0 \leqq \theta \leqq 1$ is denoted by the absolute value symbol $|\cdot|$ (the meaning will be clear from context).

(4.1) Spectral theorem. *Let* $x \in L^2(\mathscr{A}) \cap L^\infty(\mathscr{A})$*, and set*

$$I_p = \int_{\mathscr{A}} x^p \, dm, \quad p \in \mathbf{Z}^+ .$$

There exists a unique, nonincreasing, right-continuous function λ *on* $[0, 1]$ *such that*

$$I_p = \int_0^1 \lambda^p(z) \, dz, \quad p \in \mathbf{Z}^+ .$$

Proof. We need Theorem 2.6.4 of [Akh]:

Hausdorff's theorem. *The moment problem*

$$I_p = \int_0^1 u^p \, d\sigma(u), \quad p = 0, 1, 2, \ldots ,$$

has a unique solution in the class of nondecreasing, right continuous functions u taking values in $[0, 1]$*, if and only if the inequalities*

$$\sum_{i=0}^{m} (-1)^i \binom{m}{i} I_{i+k} \geqq 0 \tag{4.2}$$

hold for all $m, k = 0, 1, 2, \ldots$.

Suppose first that $0 \leqq x(z, \theta) \leqq 1$. Then (4.2) is trivial, since it asks that

$$\int_0^1 \int_0^1 x(z, \theta)^k (1 - x(z, \theta))^m \, d\theta \, dz \geqq 0 .$$

Hence the function σ in Hausdorff's theorem exists and is unique.

Now define

$$r_p(y) = \int_{\{u^p > y\}} d\sigma(u) ,$$

and introduce the "inverse function" $f_p^*(z) = \sup\{y \mid r_p(y) > z\}$. These f_p^* are nonincreasing and right continuous, and one checks that $r_p(y) = r_1(y^{1/p})$ and $f_p^*(z) = f_1^*(z)^p$. It is clear that

$$\int_0^1 f_p^*(z)\,dz = \int_0^1 u^p\,d\sigma(u)\,,$$

so that

$$\int_0^1 f_1^*(z)^p\,dz = \int_0^1 u^p\,d\sigma(u)\,.$$

The function $\lambda(z) = f_1^*(z)$ has the required property. It is unique, because from λ we can recover σ, and that function is completely determined by the I_p according to Hausdorff's theorem.

If $x(z,\theta)$ does not take almost all of its values between 0 and 1, let us assume that $-K \leqq x \leqq K$ almost everywhere ($\|x\|_\infty \leqq K$). Our criterion (4.2) for solvability of the moment problem now becomes the condition on the I_p implied by

$$\int_0^1\int_0^1 (x(z,\theta) + K)^j (K - x(z,\theta))^{m-j}\,d\theta\,dz \geqq 0\,.$$

This is, of course, still true, and the rest of the proof is modified in the obvious way. □

Remark. In Theorem 4.1, we need only require that $\mathscr{A}$ be a probability measure space.

(4.3) Diagonalization theorem. *Let $x \in L^2(\mathscr{A}) \cap L^\infty(\mathscr{A})$, and let λ be as in Theorem 4.1. Define $\lambda_2(z,\theta) = \lambda(z)$ (the subscript 2 indicates the extension of λ to two dimensions). There exists a measure preserving map ψ: $\mathscr{A} - \mathscr{Z}_1 \to \mathscr{A} - \mathscr{Z}_2$, with $\mathscr{Z}_i$ of measure zero, such that $x = \lambda_2 \circ \psi$.*

Proof. It is known [Ha2, §41] that a nonatomic, separable, totally finite normalized measure algebra is isomorphic to the measure algebra of the unit interval [0, 1]. We use Halmos' proof strategy for this result to get a particularly convenient isomorphism. Our terminology follows Halmos'.

Think of $\mathscr{A}$ as the unit square, with horizontal axis z and vertical axis θ. For each $N = 1, 2, 3, \ldots$, divide $\mathscr{A}$ into vertical strips of width $1/2^N$, and divide each strip into squares of side $1/2^N$. This gives a partition P_N of $\mathscr{A}$. The P_N are decreasing and dense in the metric space of measurable subsets of $\mathscr{A}$ (the metric being given by the measure of the symmetric difference of two sets). Define a map T from the squares in $\{P_N\}$ to subintervals of [0, 1] by mapping each vertical strip of P_N, in sequence, to intervals of length $1/2^N$; the squares in each strip are mapped to smaller subintervals in the obvious way. As explained by Halmos (who uses a more general sequence of partitions in more general measure spaces), the map T extends to an isometry $\bar{T}$ of measure algebras. A theorem of von Neumann [Ro, p. 329] says that $\bar{T}$ is implemented by a point mapping p: $\mathscr{A} \to [0, 1]$ which is one-to-one off two sets of measure zero (in domain and range).

We are now ready to define the desired measure preserving map on $\mathscr{A}$. Clearly, $x \circ p^{-1} \stackrel{\text{def}}{=} g$ is equimeasurable with x. According to [R2], there is then a measure preserving map ϕ of [0, 1] such that $g = \lambda \circ \phi$. By design, $\lambda = \lambda_2 \circ p^{-1}$. Therefore $x = \lambda_2 \circ p^{-1} \circ \phi \circ p$. The map $\psi = p^{-1} \circ \phi \circ p$ is the required measure preserving transformation of $\mathscr{A}$. □

5 Schur's theorem

In this section and the next, we shall need two facts about doubly stochastic operators; it is convenient to state them here.

(5.1) Proposition [R1, Br] *If P is doubly stochastic on $L^1([0,1])$ (see (1.12)), then:* (i) $Pf \geqq 0$ *whenever* $f \geqq 0$; (ii) $\|P\|_1 \leqq 1$; (iii) $\|P\|_\infty \leqq 1$ *on* $L^1([0,1]) \cap L^\infty([0,1])$ *(and so, by interpolation,* $\|P\|_p \leqq 1$ *for* $1 \leqq p \leqq \infty$).

(5.2) Proposition [R1, p. 1382] *Let P be a bounded operator on $L^\infty([0,1])$ satisfying* (i) $0 \leqq P\chi_E \leqq 1$ *and* (ii) $\int_0^1 P\chi_E = |E|$ *for every characteristic function of a measurable set $E \subset [0,1]$. Then P has a unique extension $\bar{P}$ to a doubly stochastic operator on $L^1([0,1])$. Conversely, every doubly stochastic operator satisfies* (i) *and* (ii).

We also want to use the L^2 version of Ryff's theorem 1.13. Ryff's setting was L^1, which is the most difficult case. To preserve the analogy with the finite-dimensional situation, we prefer to have the "projection onto the diagonal part", (1.2), be an orthogonal projection in Hilbert space.

(5.3) Theorem [R2, R3] *In $L^2([0,1]) \cap L^\infty([0,1])$: $g \prec f$ if, and only if, there is a doubly stochastic P such that $g = Pf$. The set $\Omega(f) \overset{\text{def}}{=} \{g \mid g \prec f\}$ is weakly compact and convex. Its set of extreme points is $\{f^* \circ \phi \mid \phi$ is a measure preserving transformation of $[0,1]\}$.*

Proof. The equivalence of $g \prec f$ and $g = Pf$ follows from Ryff's Theorem 1.13, since $L^2 \cap L^\infty \subset L^1$. Convexity of $\Omega(f)$ remains true for the same reason. The weak L^2 topology (which is Hausdorff) is stronger than the topology determined by the linear functionals $(\cdot, h)$, $h \in L^\infty([0,1])$. In the latter topology, $\Omega(f)$ is compact, and therefore the (necessarily continuous) inclusion map to $\Omega(f) \subset (L^2, \text{weak})$ is an isomorphism. Thus, $\Omega(f)$ is weakly compact in L^2. Finally, we should check that extreme points in L^1 (as in (1.13)) remain extreme points in L^2, and that no new extreme points are introduced. This is trivial, since all of $\Omega(f)$ consists of essentially bounded, hence L^1, functions, to which Theorem 1.13 applies. □

With these preparations out of the way, we can proceed to the first half of our convexity theorem.

(5.4) Schur's theorem. *Let $x \in L^2(\mathscr{A}) \cap L^\infty(\mathscr{A})$, let $\pi(x)$ be the zeroth Fourier coefficient of x,*

$$\pi(x)(z) = \int_0^1 x(z, \theta)\, d\theta ,$$

and let λ be as in Theorem 4.1. Then $\pi(x)$ belongs to the closed convex hull of the orbit of the Weyl semigroup $\bar{W}$ through λ.

Proof. By (2.4), the action of W on λ (which is a function of z alone) is just the action of the group of invertible measure preserving transformations of $[0,1]$ on an element of $L^2([0,1])$. By [Br], the action of $\bar{W}$ on λ is then the action of the semigroup of (not necessarily invertible) measure preserving transformations of $[0,1]$. According to (5.3), $\pi(x)$ will be in the closed convex hull of this orbit precisely when $\pi(x) \prec \lambda$. We now prove that this is the case.

According to Theorem 4.3, there is a measure preserving map

$$\psi(z, \theta) = (\tau(z, \theta), \eta(z, \theta))$$

of $\mathscr{A}$ to itself, such that $x = \lambda_2 \circ \psi$ (where, as earlier, we define $\lambda_2(z, \theta) = \lambda(z)$). It is clear that

$$(Pf)(z) \stackrel{\text{def}}{=} \int_0^1 f(\tau(z, \theta))\, d\theta$$

defines a doubly stochastic operator on $L^1([0, 1])$: condition (i) in (5.2) holds trivially, and condition (ii) is true because

$$\int_0^1 (P\chi_E)(z)\, dz = \int_0^1 \int_0^1 \chi_E \circ \psi\, d\theta\, dz = |E| ,$$

since ψ preserves the measure of the strip $E \times [0, 1]$.

But then

$$(P\lambda)(z) = \int_0^1 \lambda(\tau(z, \theta))\, d\theta = \int_0^1 x(z, \theta)\, d\theta = \pi(x)(z) , \tag{5.5}$$

which shows (by 1.13) that $\pi(x) \prec \lambda$. □

We remark that the proof follows the finite-dimensional one quite closely. The measure preserving map ψ of $\mathscr{A}$ is analogous to a unitary matrix, the relation $x = \lambda_2 \circ \psi$ is conjugation of a diagonal matrix by a unitary one, and the unitary matrix is used to define a doubly stochastic matrix.

(5.6) *Example.* Let $x(z, \theta) = z - z^2$, so that $\pi(x)(z) = X(z) = z - z^2$, with the nonincreasing rearrangement λ of X being $\lambda(z) = (1 - z^2)/4$. As promised by Schur's Theorem 5.4, $X = P\lambda$, where P is the doubly stochastic operator defined by $Pf = f \circ \phi$ with

$$\phi(z) = 2|z - \tfrac{1}{2}| .$$

This ϕ is measure preserving from [0, 1] to itself, i.e., it belongs to our Weyl semi-group, but it is not invertible. Ryff [R1] shows that there is no doubly stochastic Q for which $\lambda = QX$. In other words, X is on the $\bar{W}$-orbit through λ, but λ is not on the $\bar{W}$-orbit through X.

6 Horn's theorem

(6.1) Horn's theorem. *Let λ be a bounded, nonincreasing function on* [0, 1], *and let X lie in the closed convex hull of the Weyl semigroup orbit through λ,*

$$\bar{W} \cdot \lambda = \{\lambda \circ \phi \mid \phi \text{ is a measure preserving transformation of } [0, 1]\}.$$

Then there exists an $x \in L^2([0, 1]) \cap L^\infty([0, 1])$ such that

(i) $X(z) = \pi(x)(z) = \int_0^1 x(z, \theta)\, d\theta$,

(ii) $\int_0^1 \int_0^1 x(z, \theta)^p d\theta\, dz = \int_0^1 \lambda(z)^p\, dz, \quad p \in \mathbf{Z}^+$.

Remark. By Ryff's Theorem 1.13, the hypothesis on X is equivalent to $X \prec \lambda$, which in turn means that $X = P\lambda$ for a doubly stochastic operator P. This is the property we shall use.

An example might help to motivate our proof of this theorem.

(6.2) *Example.* Let $a_1 > a_2 > a_3 > a_4$, and define

$$\lambda(z) = a_i \text{ for } \frac{i-1}{4} \leqq z < \frac{i}{4}, \quad i = 1, \ldots, 4,$$

and

$$X(z) = \begin{cases} (a_1 + a_2)/2, & \text{if } 0 \leqq z < 1/2, \\ (a_3 + a_4)/2, & \text{if } 1/2 \leqq z < 1. \end{cases}$$

The operator P defined by

$$(Pf)(z) = \begin{cases} 2 \int_0^{1/2} f(t)\,dt, & \text{if } 0 \leqq z < 1/2, \\ 2 \int_{1/2}^{1} f(t)\,dt, & \text{if } 1/2 \leqq z \leqq 1 \end{cases} \tag{6.3}$$

is doubly stochastic, and maps λ to X: $X = P\lambda$.

Equation (5.5) in the proof of Schur's theorem indicates a connection between λ, P, and x. It suggests that we should find a function $\tau(z, \theta)$ so that $\lambda(\tau(z, \theta))$ gives one possible of the Horn problem for the prescribed λ and X.

Let

$$\sigma(z, y) = P\chi_{[0,y]}(z), \tag{6.4}$$

which works out to

$$\sigma(z, y) = \begin{cases} 2y, & \text{if } 0 \leqq z < 1/2, 0 \leqq y \leqq 1/2, \\ 1, & \text{if } 0 \leqq z < 1/2, 1/2 \leqq y \leqq 1; \\ 0, & \text{if } 1/2 \leqq z \leqq 1, 0 \leqq y < 1/2, \\ 2(y - \frac{1}{2}), & \text{if } 1/2 \leqq z \leqq 1, 1/2 \leqq y \leqq 1. \end{cases}$$

If we now define

$$\tau(z, \theta) = \inf\{y \mid \sigma(z, y) > \theta\} \tag{6.5}$$

for $0 \leqq \theta < 1$, we find

$$\tau(z, \theta) = \begin{cases} \theta/2, & \text{if } 0 \leqq z < 1/2, \\ (1 + \theta)/2, & \text{if } 1/2 \leqq z \leqq 1. \end{cases}$$

The function

$$x(z, \theta) = \lambda(\tau(z, \theta)) \tag{6.6}$$

takes on the values a_1, a_2, a_4, a_3 on the four squares $0 \leqq z, \theta < 1/2$; $0 \leqq z < 1/2$, $1/2 \leqq \theta \leqq 1$; etc., clockwise around $\mathscr{A}$. It is easy to check that x and λ have the same moments, and that $\pi(x) = X$. Thus, X is one solution of Horn's problem.

We will see that (6.4), (6.5), and (6.6) work in general.

Remark. The operator defined by

$$(P_1 f)(z) = \begin{cases} \frac{1}{2}f(\frac{z}{2}) + \frac{1}{2}f(\frac{z}{2}+\frac{1}{4}), & \text{if } 0 \leqq z < 1/2\,, \\ \frac{1}{2}f(\frac{z}{2}+\frac{1}{4}) + \frac{1}{2}f(\frac{z}{2}+\frac{1}{2}), & \text{if } 1/2 \leqq z \leqq 1\,, \end{cases}$$

is doubly stochastic, and also maps λ in the preceding example to the same X. So the doubly stochastic operator is not unique. It so happens that P and P_1 produce the same x by the prescription in Example 6.2, but one can easily see that x is not unique either.

Proof of theorem. By assumption, we have a doubly stochastic P for which $P\lambda = X$. We want to set $\sigma(z, y) = P\chi_{[0,y]}(z)$ and to construct a suitable x via (6.5) and (6.6). There is a slight technical problem: the function $y \mapsto P\chi_{[0,y]}$ is defined from $[0, 1]$ to $L^1([0, 1])$, while the desired $\sigma(z, y)$ is a function from $\mathscr{A}$ to $\mathbf{R}$. The identification of $\mathscr{C}([0, 1], L^1([0, 1]))$ with a subset of $L^1([0, 1] \times [0, 1])$ is a standard exercise in function spaces, but we fill in the details just to establish the monotonicity properties we shall need.

Let $F(y) = P\chi_{[0,y]}$; F is a continuous function from $[0, 1]$ to $L^1([0, 1])$. When $s < r$, (5.1.i) implies that

$$F(r)(z) - F(s)(z) = (P\chi_{(s,r]})(z) \in [0, 1] \tag{6.7}$$

for almost every z, and (5.1.ii) shows that

$$\| F(r) - F(s) \|_1 = \| P\chi_{(s,r]}) \|_1 \leqq (r - s)\,. \tag{6.8}$$

When r and s are dyadic rationals, (6.7) will hold for z not in a set E_{rs} of measure zero. Let E be the countable union of the E_{rs}.

Now set $I_{j,N} = [(j-1)/2^N, j/2^N)$, and

$$\sigma_N(z, y) = \sum_{j=1}^{2^N} \chi_{I_{j,N}}(y) F((j-1)/2^N)(z)\,.$$

By (6.7) and the definition of E, σ_N is an increasing step function in y, with values in $[0, 1]$, for every $z \notin E$. Furthermore, when $N < M$ then

$$\sigma_N(z, y) \leqq \sigma_M(z, y), \quad z \notin E, y \in [0, 1]\,.$$

Hence $\lim_{N\to\infty} \sigma_N(z, y) \overset{\text{def}}{=} \sigma(z, y)$ exists for $z \notin E$, $y \in [0, 1]$. The limit σ is measurable, since each σ_N is, and σ is nondecreasing in y for each $z \notin E$.

For each y, $\sigma_N(\cdot, y)$ tends to $F(y)$ in $L^1([0, 1])$; this is immediate when (6.8) is applied to

$$\sigma_N(z, y) - F(y)(z) = \sum_{j=1}^{2^N} \chi_{I_{j,N}}(y)(F((j-1)/2^N)(z) - F(y)(z))\,.$$

In particular,

$$\sigma(\cdot, y) = P\chi_{[0,y]}, y \in [0, 1]\,. \tag{6.9}$$

Next, introduce τ as in (6.5), but only for $z \notin E$:

$$\tau(z, \theta) = \inf\{y \,|\, \sigma(z, y) > \theta\}\,.$$

We show that τ is measurable. Consider the set

$$\{(z, \theta) \,|\, \tau(z, \theta) \leqq \alpha\} \subset \mathscr{A}\,. \tag{6.10}$$

It is easy to see that (for $z \notin E$)

$$\tau(z, \theta) \leqq \alpha \text{ if and only if } \sigma(z, \alpha) \geqq \theta \tag{6.11}$$

(the "if" part uses the fact that σ is nondecreasing in its second argument). Hence the set in (6.10) is the same as

$$\{(z, \theta) | 0 \leqq \theta \leqq \sigma(z, \alpha)\} . \tag{6.12}$$

By (6.9), $\sigma(\cdot, \alpha) = P\chi_{[0,\alpha]} \overset{\text{def}}{=} h$. The set (6.12) is just the collection of points under the graph of h, and is measurable with respect to planar Lebesgue measure because h is measurable on [0, 1] [S, Theorem III.10.3]). Hence the set (6.10) is measurable.

Now we define $x(z, \theta) = \lambda(\tau(z, \theta))$. We will verify that (i) and (ii) in the theorem are satisfied.

To prove (i), let $f(z) = \sum_1^n c_j \chi_{A_j}(z)$ be a nonincreasing step function with $A_j = [a_j, a_{j+1})$. We have

$$\int_0^1 \chi_{A_j}(\tau(z, \theta))\, d\theta = |\{\theta | \tau(z, \theta) \in A_j\}| . \tag{6.13}$$

But if $a_j \leqq \tau(z, \theta) < a_{j+1}$, then by (6.11), $\sigma(z, a_j) \leqq \theta < \sigma(z, a_{j+1})$. Hence the integral in (6.13) is $\sigma(z, a_{j+1}) - \sigma(z, a_j)$, and

$$\int_0^1 f(\tau(z, \theta))\, d\theta = \sum_{j=1}^n c_j(\sigma(z, a_{j+1}) - \sigma(z, a_j)) = \sum_{j=1}^n c_j(P\chi_{A_j})(z) = (Pf)(z) \tag{6.14}$$

(for a.e. z). Now let f_k be a sequence of nonincreasing step functions that converges monotonically from above to λ (such a sequence exists, because λ is upper semicontinuous). Clearly, $f_k \to \lambda$ in L^1 also, so $Pf_k \to P\lambda$ in L^1. By (5.1.i), $\{Pf_k\}$ is also monotone, so it must converge almost everywhere to $P\lambda$. Furthermore, for each $z \notin E$, $f_k(\tau(z, \theta))$ converges monotonically to $\lambda(\tau(z, \theta))$ for all θ. Thus, replacing f by f_k in (6.14) and passing to the limit, we obtain (i).

To prove (ii), we first integrate (6.14) over z. According to (5.2.ii),

$$\int_0^1 (P\chi_A)(z)\, dz = |A|$$

for all (measurable) $A \subset [0, 1]$. Therefore,

$$\int_0^1 \int_0^1 f(\tau(z, \theta))\, d\theta\, dz = \sum_{j=1}^n c_j |A_j| = \int_0^1 f(z)\, dz .$$

In this relation, replace f by f_k^p, and proceed as before. □

Remark. We believe, but have not tried very hard to prove (because this refinement is not needed for our argument), that our function $\tau(z, \theta)$ is the first component of a measure preserving map $\psi(z, \theta) = (\tau(z, \theta), \eta(z, \theta))$ of $\mathscr{A}$. If this were true (it is for Example 6.2), we could think of the doubly stochastic operator P as infinite-dimensional analog of Horn's [H] "orthostochastic" matrix, as used in his proof of the finite dimensional theorem.

Finally, we return to a point postponed from the very first section. The Lie algebra $\mathscr{G}$ of SDiff($\mathscr{A}$), as was noted there, is the algebra of divergence-free vector fields tangent to the boundary of the annulus $\mathscr{A}$. We have consistently identified $\mathscr{G}$ with the Poisson algebra $\mathscr{P}$ of Hamiltonian functions, which is actually a (trivial) central extension of $\mathcal{G}$. It is easy enough to use our convexity theorem to obtain the corresponding result for (a Hilbert space completion of) the correct Lie algebra $\mathcal{G}$.

One begins by noting that $\mathcal{G} \cong \mathcal{P}/\{\text{constant functions}\}$. Because the constant functions are invariant under the action of SDiff($\mathcal{A}$) and the diagonal projection π, (1.2), these descend to the quotient. As pre-Hilbert space, $\mathcal{G}$ is naturally isomorphic to the orthogonal complement in $\mathcal{P}$ of the constant functions, but this complement is not a Lie algebra under Poisson bracket. Upon completion, $\mathcal{G}$ becomes the quotient of $L^2(\mathcal{A})$ by the constant functions, and of course it is isomorphic, as Hilbert space, to the othogonal complement of the constants. An $\overline{\text{SMeas}(\mathcal{A})}$-orbit in the completion of $\mathcal{G}$ may thus be identified with an orbit through a function whose integral over $\mathcal{A}$ is zero. Consequently, the convexity theorem for the completion of $\mathcal{G}$ follows from application of our Schur-Horn theorem to orbits through functions whose first moment $I_1 = \int_A x\, dm$ vanishes.

Acknowledgements. In the early stages of our work, when we were trying to find our way through measure theory and area-preserving maps, we were helped by S. Alpern, R. Brockett, Ph. Boyland, Ya. Eliashberg, N. Falkner, W. Faris, C. Golé, D. Pickrell, and T. Ward; we are grateful for their advice. We thank Lenny Friedlander for curing an almost fatal case of confusion towards the end, and we are particularly indebted to Jürgen Moser for telling us, and writing up, his approximation theorem (summarized in §3) that allows us to use group theory language. Bloch was supported by NSF grant DMS-90-02136, PYI DMS-91-57556, and AFOSR Contract F49620-93-1-0037, Flaschka by DMS-90-01897, and Ratiu by DMS-91-22708.

References

[Akh] Akhiezer, N.: The Classical Moment Problem. (Univ. Math. Monogr.) Edinburgh London: Oliver & Boyd 1965

[Al] Alpern, S.: Approximation to and by measure preserving homeomorphisms. J. Lond. Math. Soc. **18**, 305–315 (1978)

[AP] Atiyah, M.F., Pressley, A.N.: Convexity and loop groups. In: Artin, M., Tate, J. (eds.) Arithmetic and Geometry: papers dedicated to I.R. Shafarevich on the occasion of his sixtieth birthday, vol. 2. Boston Basel Stuttgart: Birkhaüser 1983

[BR] Bao, D., Ratiu, T.: A candidate maximal torus in infinite dimensions. Contemp. Math., vol. **132**, pp. 117–124. Providence, RI: Am. Math. Soc. 1992, details to appear in another paper

[Bi] Birkhoff, G.: Three observations on linear algebra. Rev. Univ. Nac. Tucumán, Ser. A **5**, 147–151 (1946) (we have only seen the Math. Reviews summary of this paper)

[BBKR] Bloch, A.M., Brockett, R.W., Kodama, Y., Ratiu, T.: Spectral equations for the long wave limit of the Toda lattice equations. In: Proceedings of the CRM Workshop on Hamiltonian Systems, Transformation Groups and Spectral Transform Methods, pp. 97–402. Harnad, J., Marsden, J.E. (eds.) Publ. CRM. Montréal 1990

[BBR] Bloch, A.M., Brockett, R.W., Ratiu, T.: A new formulation of the generalized Toda lattice equations and their fixed point analysis via the momentum map. Bull. **23**, 447–456 (1990)

[Bo] Bogojavlenskiĭ, O.I.: Lax representations with spectral parameter for some dynamical systems. Izv. Akad. Nauk. SSSR **52**, 243–266 (1988)

[BB] Brockett, R.W., Bloch, A.: Sorting with the dispersionless limit of the Toda lattice. In: Proceedings of the CRM Workshop on Hamiltonian Systems, Transformation Groups and Spectral Transform Methods, pp. 103–112. Harnad, J., Marsden, J.E. (eds.) Publ. CRM. Montréal 1990

[Bro] Brockett, R.W.: Dynamical systems that sort lists, solve linear programming problems and diagonalize symmetric matrices. Proc. 1988 IEEE Conference on Decision and Control, 1988. Linear Algebra Appl. **122–124**, pp. 761–777 (1989)

[Br] Brown, J.R.: Approximation theorems for Markov operators. Pac. J. Math. **16**, 13–23 (1966)

[EM] Ebin, D.G., Marsden, J.: Groups of diffeomorphisms and the motion of an incompressible fluid. Ann. Math. **92**, 102–163 (1970)

[Ha1] Halmos, P.: Approximation theorems for measure preserving transformations. Trans. Am. Math. Soc. **55**, 1–18 (1944)

[Ha2] Halmos, P.: Measure Theory. Princeton, NJ: D. Van Nostrand 1950

[HLP] Hardy, G.H., Littlewood, J.E., Pólya, G.: Inequalities. Cambridge: Cambridge University Press 1934

[H] Horn, A.: Doubly stochastic matrices and the diagonal of a rotation matrix. Am. J. Math. **76** 620–630 (1954)

[K] Kostant, B.: On convexity, the Weyl group and the Iwasawa decomposition. Ann. Sci. Éc. Norm. Supér. **6**, 413–455 (1973)

[KP] Kac, V.G., Peterson, D.H.: Unitary structure in representations of infinite-dimensional groups and a convexity theorem. Invent. Math. **76**, 1–14 (1984)

[MO] Marshall, A.W., Olkin, I.: Inequalities: Theory of Majorization and its Applications. New York: Academic Press 1979

[Mo] Moser, J.: Remark on the smooth approximation of volume-preserving homeomorphisms by symplectic diffeomorphisms. Zürich: Forschungsinstitut für Mathematik ETH (Preprint, October 1992)

[Ro] Royden, H.: Real Analysis. New York: Macmillan 1963

[R1] Ryff, J.V.: On the representation of doubly stochastic operators. Pac. J. Math. **13**, 1379–1386 (1963)

[R2] Ryff, J.V.: Orbits of L^1 functions under doubly stochastic transformations. Trans. Am. Math. Soc. **117**, 92–100 (1965)

[R3] Ryff, J.V.: Extreme points of some convex subsets of $L^1(0, 1)$. Proc. Am. Math. Soc. **18**, 1026–1034 (1967)

[S] Saks, S.: Theory of the Integral, 2nd ed. (Dover Publ.) New York: Dover 1964

[Sa1] Savieliev, M.V.: Integro-differential nonlinear equations and continual Lie algebras. Commun. Math. Phys. **121**, 283–290 (1989)

[Sa2] Saveliev, M.V.: On the integrability problem of the continuous long wave approximation of the Toda lattice. Lyon: Éc. Norm. Supé. Preprint (1992)

[SV] Saveliev, M.V., Vershik, A.M.: New examples of continuum graded Lie algebras. Phys. Lett. **A143**, 121–128 (1990)

[Sch] Schur, I.: Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie. Sitzungsber. Berl. Math. Ges. **22**, 9–20 (1923)

[T] Toda, M.: Theory of Nonlinear Lattices. Berlin Heidelberg New York: Springer 1981

[VGG] Vershik, A.M., Gel'fand, I.M., Graev, M.I.: Representation of the group of diffeomorphisms. Usp. Mat. Nauk **30** (no. 6), 1–50 (1975)

[WE] Watson, D.S., Elsner, L.: On Rutishauser's approach to selfsimilar flows. SIAM J. Matrix. Anal. Appl. **11**, 301–311 (1990)