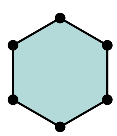
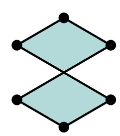
Dynamical Systems Linear Programming and the Monge Kantorovich Problem

Smale95

Anthony Bloch with Tudor Ratiu

(earlier work with Brockett, Flaschka, more recently Karp)

- **Monge problem in finite-dimensions**
- **Dynamics and Linear Programing**
- **Schur-Horn**
- **Gradient**
- **Infinite-Dimensions: diffeomorphism group of the annulus**
- **Other things: Positivity**

$\mathsf{Perm}(3,2,1) =$  , $\mu(\mathcal{J}^{\geq 0}_{(3,2,1)}) =$  .

Smale Connections:

- **The University of Michigan! Smale, me, Tudor, Indika:**
- **Topology and Mechanics**
- **The Mathematics of Time**
- **Dynamical Systems**
- **Linear Programming**
- **Economics.**
- **Biological Systems and networks (I. Rajapakse, Fred Leve)**

Consider a geometric approach to analyzing a particular setting of the Monge–Kantorovich problem and its connection to Schur–Horn theory. We begin by analyzing the finite-dimensional setting (as in Brezis 2018).

Then we consider the Monge–Kantorovich problem on the diffeomorphism group of the annulus. This is a particular interesting case of the general infinite-dimensional problem, as presented in, for example, Evans and Gangbo and McCann.

We relate these problems, respectively, to the classical Schur–Horn theorem, and its infinite-dimensional generalization to the diffeomorphism group of the annulus as proved in Bloch, Flashcka, Ratiu 93. We also consider the dual problems. In addition, we relate these problems to the gradient flows discussed in work with Brockett, Ratiu, Flashcka, Karp

1. The finite dimensional case

Brezis introduces the following finite-dimensional setting.

Consider two sets X, Y **consisting of** m **points** $\{P_i\}$ **and** $\{N_i\}$, $1 \le i \le m$, **i.e.,**

$$X = \{P_1, P_2 \ldots, P_m\}, \quad Y = \{N_1, N_2 \ldots, N_m\}. \tag{1.1}$$

Let $c : X \times Y \to \mathbb{R}$ **be a smooth "cost" function. Brezis introduces three problems denoted M, K, and D (for dual):**

$$\mathbf{M} := \mathbf{min}_{\sigma \in S_m} \sum_{i=1}^{m} c(P_i, N_{\sigma(i)}), \tag{1.2}$$

where S_m **is the permutation group of** $\{1, \ldots, m\}$**, and**

$$\mathbf{K} := \mathbf{min}_A \left\{ \sum_{i,j=1}^{m} a_{ij} c(P_i, N_j) \,\middle|\, A = (a_{ij}) \text{ \textbf{is doubly stochastic}} \right\}. \tag{1.3}$$

For our purposes, we formulate the dual problem as:

(1.4)
$$\mathbf{D} := \mathbf{sup}_{\varphi: X \to \mathbb{R},\, \psi: Y \to \mathbb{R}} \left\{ \sum_{i=1}^{m} (\varphi(P_i) - \psi(N_i)) \,\middle|\, \varphi(x) - \psi(y) \le c(x, y), \forall x \in X, y \in Y \right\}.$$

Brezis proves that M = K = D. We shall consider these equalities from the point of view of majorization and dynamics. It is clear, for example, that M $\geq$ K since a permutation matrix (every row and column has exactly one entry equal to 1 and all other entries equal to 0) is a special case of a doubly stochastic matrix (all entries are ≥ 0 and the sum of all entries in each row and each column is 1). We shall also show this is true in our infinite-dimensional setting. We also show that K $\geq$ D in both cases. Brezis also shows D $\geq$ M in the finite case completing the equalities.

1.1. **The adjoint orbit and dynamical setting. We can arrive also at this finite setting by considering the problem of minimizing Trace**(LN)**, where** L **belongs to the isospectral set of skew Hermitian matrices defined by the purely imaginary diagonal matrix** Λ **and** N **is a constant skew Hermitian matrix, as in Bloch, Brockett, Ratiu [1992].**

More precisely we consider

$$\mathbf{min}_\Theta \|L - N\|^2 := \mathbf{min}_\Theta \langle L - N, L - N\rangle := \mathbf{min}_\Theta \left\langle \Theta^T \Lambda \Theta - N, \Theta^T \Lambda \Theta - N \right\rangle \tag{1.5}$$

where Θ **belongs to the group of unitary matrices. This is equivalent to minimizing Trace**(LN) **and is the Monge problem M in this setting.**

We arrive at the solution by following the gradient dynamics

$$\dot{L} = [L, [L, N]] \tag{1.6}$$

which is the gradient flow of Trace(LN) **on an adjoint orbit of the unitary group** $\mathrm{U}(n)$ **with respect to the normal metric.**

Theorem 1.1. *Equation* (1.6) *is the gradient flow of* $\mathrm{Trace}(LN)$ *with respect to the normal metric on an adjoint orbit of* $\mathrm{U}(n)$. *For* N *diagonal with distinct diagonal entries and* L *having initial condition with distinct eigenvalues, there are* $n!$ *equilibria corresponding to the* $n!$ *diagonal matrices with rearranged eigenvalues. The stable equilibrium is the one having the same ordering as the entries of* N*, after dividing both by* i*.*

1.2. **Convexity in finite dimensions. We now consider the Brezis equalities M = K = D from the Lie theoretic point of view presented above. We begin by recalling the following:**

Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$**; then** $S_n x$ **denotes the orbit of** x **under the symmetric group on** n **letters, i.e., the collection of all points** $(x_{s(1)}, \ldots, x_{s(n)})$**, where** s **ranges over all** $n!$ **permutations. For** $C \subset \mathbb{R}^n$**,** $\widehat{C}$ **denotes the convex hull of** C**, i.e., the smallest convex set containing** C**.**

Theorem 1.2. Schur's Theorem. *(1923) Let A be a Hermitian matrix with eigenvalues λ_j, arranged in non-increasing order. Let $\lambda = (\lambda_1, \ldots, \lambda_n)$ and $A^0 = (A_{11}, \ldots, A_{nn})$ be the diagonal of A. Then*

$$A^0 \in \widehat{S_n \lambda}.$$

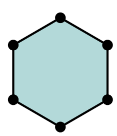
Theorem 1.3. Horn's Theorem. *(Horn 54) Let $\lambda \in \mathbb{R}^n$, with components arranged in non-increasing order. If $A^0 \in \widehat{S_n \lambda}$, there is a Hermitian matrix A with eigenvalues λ whose diagonal is A^0.*

Schur-Horn theorem.

Let $d_1, \ldots, d_n$ **and** $\lambda_1, \ldots, \lambda_n$ **be two sequences of real numbers arranged in nondecreasing order.**
Then there exists a Hermitian matrix with diagonal entries $d_1, \ldots, d_n$ **and eigenvalues** $\lambda_1, \ldots, \lambda_n$ **if and only if**

$$\begin{gathered} d_1 \le \lambda_1 \\ d_1 + d_2 \le \lambda_1 + \lambda_2 \\ \vdots \\ d_1 + d_2 \cdots + d_n = \lambda_1 + \lambda_2 \ldots \lambda_n \,. \end{gathered}$$

Equivalent: If d_i, λ_i **are as above there exists a Hermitian matrix with these as diagonal entries and eigenvalues respectively if and only if the vector** $(d_1, \ldots, d_n)$ **is in the permutohedron generated by** $(\lambda_1, \ldots, \lambda_n)$**.**

$\mathsf{Perm}(3, 2, 1) =$ 

Definition 1.4. For $x \in \mathbb{R}^n$, 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

$$x_1^* + \cdots + x_k^* \leq y_1^* + \dots y_k^*, \quad \text{for} \quad 1 \leq k \leq n-1, \quad \text{and} \quad \sum_{j=1}^{n} x_j^* = \sum_{j=1}^{n} y_j^*.$$

Definition 1.5. An $n \times n$ real matrix P is called *doubly stochastic* if $P_{ij} \geq 0$, and the sum of all entries in each row and each column is 1.

Theorem 1.6. (a) *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) *P is doubly stochastic if and only if $Px \prec x$ for all $x \in \mathbb{R}^n$.*
(c) *$x \prec y$ if and only if there is a doubly stochastic P such that $x = Py$.*
(d) *The set of doubly stochastic matrices is the convex hull of its extreme points, which are precisely the permutation matrices.*

Remark 1.7. $\{x \mid x \prec y\} = \widehat{S_n y}$.

The Schur theorem now follows easily. Diagonalize the Hermitian matrix A, $A = Q\lambda Q^*$, Q unitary, λ real diagonal. Then $A_{ii} = \sum_j |Q_{ij}|^2 \lambda_j$. If Q is unitary, the matrix $P_{ij} = |Q_{ij}|^2$ must be doubly stochastic. Theorem 1.6(d) 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 matrix P of the form $P_{ij} = |Q_{ij}|^2$, with Q unitary, satisfying $x = P\lambda$; $Q\lambda Q^*$ is then the desired Hermitian A having eigenvalues λ_j and diagonal x.

We can see clearly from these considerations that M $\geq$ K.

We note that problem K can be also be formulated in slightly more general form as follows.

We allow the cost matrix to be non-square, $n \times m$, $n \neq m$. In much of our analysis below we will require the square case which we will stipulate.

Suppose we are given nonnegative numbers c_{ij}, μ_i^+, μ_j^-, $i = 1, \ldots, n$, $j = 1, \ldots, m$, satisfying

$$\sum_{i=1}^{n} \mu_i^+ = \sum_{j=1}^{m} \mu_j^- \, . \tag{1.7}$$

The goal is to minimize over $\mu_{ij} \geq 0$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \mu_{ij} c_{ij} \tag{1.8}$$

subject to the constraints

$$\sum_{j=1}^{m} \mu_{ij} = \mu_i^+ \, , \qquad \sum_{i=1}^{n} \mu_{ij} = \mu_J^- \, . \tag{1.9}$$

For $m = n$ and normalizing so that the sums are unity we recover K. More generally, we can consider $m \neq n$.

Can do the dual problem similarly.

2. The SDiff and SMeas Setting

Many of the ideas above generalize to an appropriate infinite-dimensional setting based on our work on the infinite-dimensional Schur–Horn theorem in Bloch, Flaschka, Ratiu, Inv. Math.

Consider the following (initially smooth) setting; $\mathrm{SDiff}(\mathcal{A})$ **is the group of** C^∞ **area preserving diffeomorphisms of the annulus**

$$\mathcal{A} = \{0 \leq z \leq 1\} \times \{\exp(2\pi i\theta) \mid 0 \leq \theta < 1\}$$

(more generally, one could consider Sobolev maps in H^s **for some** $s > 2$**). Its Lie algebra** $\mathfrak{g}$ **is identified with the Poisson algebra of functions** x **satisfying**

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

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

We next consider $\mathrm{SMeas}(\mathcal{A})$**, the group of invertible measure preserving transformations of the annulus. Each** $g \in \mathrm{SMeas}(\mathcal{A})$ **determines a unitary operator** P_g **on** $L^2(\mathcal{A})$ **by** $P_g x = x \circ g$**. The strong operator topology induces a topology on** $\mathrm{SMeas}(\mathcal{A})$**. It is traditionally called the** *weak topology***, because the strong and weak operator topologies coincide on unitary operators.**

We now want to define majorization and doubly stochastic operators in this setting. We use the following.

Definition 2.1. (Ryff) Let $f \in L^1([0,1])$. Set $m(y) = \big|\{z \mid f(z) > y\}\big|$ (absolute value denotes Lebesgue measure on $[0,1]$) and, for $0 \le z < 1$, set

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

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

Definition 2.2. Definition . Let $f, g \in L^1([0,1])$. We say that f *majorizes* g (written $g \prec f$) if

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

Definition 2.3. A linear operator P on $L^1([0,1])$ is called *doubly stochastic* if $Pf \prec f$ for all $f \in L^1([0,1])$.

Theorem 2.4. *(Ryff) In $L^1([0,1])$: $g \prec f$ if and only if there is a doubly stochastic P such that $g = Pf$. The set $\Omega(f) = \{g \mid g \prec f\}$ is weakly compact and convex. Its set of extreme points is* $\{f^* \circ \phi \mid \phi \text{ is a measure preserving transformation of } [0,1]\}$.

2.1. **Some spectral results. The following results are proved in Bloch, Flashcka, Ratiu: so we shall omit the proofs.**

Theorem 2.5. *Spectral Theorem. Let* $x \in L^2(\mathcal{A}) \cap L^\infty(\mathcal{A})$*, and set*

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

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

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

Theorem 2.6. *Diagonalization Theorem. Let* $x \in L^2(\mathcal{A}) \cap L^\infty(\mathcal{A})$ *and let* λ *be as in the Spectral Theorem. Define* $\lambda_2(z, \theta) = \lambda(z)$ *There exists a measure preserving map* $\psi : \mathcal{A} \setminus \mathcal{Z}_1 \to \mathcal{A} \setminus \mathcal{Z}_2$*, with* $\mathcal{Z}_i$ *of measure zero, such that* $x = 2 \circ \psi$*.*

Theorem 2.7 (Ryff)**.** *In* $L^2([0,1]) \cap L^\infty([0,1])$*:* $g \prec f$ *if and only if there is a doubly stochastic operator* P *such that* $g = Pf$*. The set* $\Omega(f) = \{g \mid g \prec f\}$ *is weakly compact and convex. Its set of extreme points is* $\{f^* \circ \phi \mid \phi$ is a measure preserving transformation of $[0,1]\}$.

We now need an analogue of the permutation group in our setting in order to formulate the Schur and Horn theorems. This corresponds to the Weyl group of the unitary group, generalized to our setting.

The analogue of the Weyl group to our setting is group W **of invertible measure preserving transformations of** $[0,1]$**. The action of the the Weyl group** W **on** λ **(which is a function of** z **alone) is just right composition of an element of** $L^2([0,1])$ **by an invertible measure preserving transformations of** $[0,1]$**. The Weyl semigroup** $\overline{W}$ **is the closure of** W **in in the strong operator topology and consists of not necessarily invertible measure preserving transformations of** $[0,1]$ **([?, Theorem 5]). The action of** $\overline{W}$ **on** λ **is again right composition.**

Theorem 2.8. *Schur's Theorem. Let* $x \in L^2(\mathcal{A}) \cap L^\infty(\mathcal{A})$*, 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 the spectral theorem. Then* $\pi(x)$ *belongs to the closed convex hull of the orbit of the Weyl semigroup* $\overline{W}$ *through* λ *which happens precisely when* $\pi(x) \prec \lambda$*.*

Theorem 2.9. *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*,*

$$\overline{W} \cdot \lambda = \{\lambda \circ \phi \mid \phi \text{ is a measure preserving transformation of } [0,1]\}. \tag{2.1}$$

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

$$(i) \quad X(z) = \pi(x)(z) = \int_0^1 x(z,\mathfrak{t}^{\mathfrak{u}})\, d\mathfrak{t}^{\mathfrak{u}},$$

$$(ii) \quad \int_0^1 \int_0^1 x(z,\mathfrak{t}^{\mathfrak{u}})^p\, d\mathfrak{t}^{\mathfrak{u}} dz = \int_0^1 z)^p\, dz, \; p \in \mathbb{Z}^+.$$

2.2. **Problems M and K. We formulate these two problems in our setting and then prove that M implies K. In both problems, and later in the dual problem D, the integrand $c(x, y)$ of the cost function is a continuous real-valued function on $\mathbb{R}^n \times \mathbb{R}^n$.**

<u>Problem K</u>. The analogue of K in our infinite dimensional setting is the following. Consider the class of probability measures μ on $\mathbb{R}^n \times \mathbb{R}^n$ with $\mathrm{proj_x}\mu = \mu^+ \, , \mathrm{proj_y}\mu = \mu^-$. **We wish to find μ which minimizes**

$$J[\mu] = \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) d\mu(x, y) \, . \tag{2.2}$$

<u>Problem M</u>. The analogue of M in our infinite dimensional setting is the following. Given are two nonnegative Radon measures μ^+, μ^- on $\mathbb{R}^n$ satisfying $\mu^+(\mathbb{R}^n) = \mu^-(\mathbb{R}^n)$. Consider the class of measurable 1-1 mappings $s : \mathbb{R}^n \to \mathbb{R}^n$ **which rearrange μ^+ to μ^-,** $s_\#(\mu^+) = \mu^-$ **(**$s_\#$ **denotes push forward of measures), i.e.,**

$$\int_X h(s(x) d\mu^+(x) = \int_Y h(s(x) d\mu^-(y) \, , \tag{2.3}$$

for any continuous function h, where X is the support of μ^+ and Y is the support of μ^-. We want to find s which minimizes

$$I[s] = \int_{\mathbb{R}^n} c(x, s(x)) d\mu^+(x) \, . \tag{2.4}$$

The cost function is often chosen to be quadratic:

$$c(x, y) = \tfrac{1}{2} \|x - y\|^2 \, .$$

This corresponds to the Wasserstein distance W_2.

Definition 2.10. Let $f^+, f^- : \mathbb{R}^n \to \mathbb{R}$ be smooth functions with compact supports. The Wasserstein distance W_2 between them is given by

$$d(f^+, f^-)^2 = \inf \left\{ \frac{1}{2} \int_{\mathrm{supp}\, f^+} \int_{\mathrm{supp}\, f^-} \|x - y\|^2 d\mu(x, y) \right\} \tag{2.5}$$

where the infimum is over all nonnegative Radon measures μ with projections $\mu^+ = f^+ dx$, $\mu^- = f^- dy$, i.e., μ^+, μ^- are assumed to be given by the smooth densities f^+, f^-.

In our finite dimensional setting we can think of $s(\Lambda) = \Theta^T \Lambda \Theta$ and $c = \|N - L\| = \|N - \Theta^T \Lambda \Theta\|$.

We can show that M $\geq$ K. Thus, in the context of our infinite dimensional setting, we consider the cost function

$$-\langle x(z, \theta), z \rangle = -\int_0^1 \int_0^1 x(z, \theta) z dz d\theta \, . \tag{2.6}$$

Dynamics:

$$\dot{x} = \{x, \{x, z\}\}$$

for the z, θ Poisson bracket.

Related: Dynamical approach to total positivity with Steve Karp in CMP etc.

Totally positive matrices: all minors positive. Origins: Gantmacher-Krein, Schoenberg in the 1930's.

Use similar dynamical techniques and gradient flows to prove topological results for totally positive Grassmannians, flag varieties and amplituhedra

Applications: Vibrations in mechanical systems, diminishing sign changes in a sequence, statistics, Markov processes, representation theory, high energy physics.