OPTIMAL TRANSPORT AND THE GEOMETRY OF $L^1(\mathbb{R}^d)$

IVAR EKELAND AND WALTER SCHACHERMAYER

Abstract. A classical theorem due to R. Phelps states that if C is a weakly compact set in a Banach space E, the strongly exposing functionals form a dense subset of the dual space E'. In this paper, we look at the concrete situation where $C \subset L^1(\mathbb{R}^d)$ is the closed convex hull of the set of random variables $Y \in L^1(\mathbb{R}^d)$ having a given law ν. Using the theory of optimal transport, we show that every random variable $X \in L^\infty(\mathbb{R}^d)$, the law of which is absolutely continuous with respect to the Lebesgue measure, strongly exposes the set C. Of course these random variables are dense in $L^\infty(\mathbb{R}^d)$.

1. Introduction

Throughout this paper we deal with a fixed probability space $(\Omega, \mathcal{F}, P)$. It will be assumed that $(\Omega, \mathcal{F}, P)$ has no atoms. The space of d-dimensional random vectors will be denoted by $L^0\left(\Omega, \mathcal{F}, P; \mathbb{R}^d\right)$, and the space of p-integrable ones by $L^p\left(\Omega, \mathcal{F}, P; \mathbb{R}^d\right)$, shortened to L^0 and L^p if there is no ambiguity. The *law* μ_X of a random vector X is the probability on $\mathbb{R}^d$ defined by:

$$\forall f \in C^b(\mathbb{R}^d), \ \int_\Omega f\left(X\left(\omega\right)\right) dP = \int_{\mathbb{R}^d} f\left(x\right) d\mu_X$$

where $C^b(\mathbb{R}^d)$ is the space of continuous and bounded functions on $\mathbb{R}^d$. The last term is, as usual, denoted by $\mathbb{E}_{\mu_X}[f]$. Clearly, $X \in L^p\left(\mathbb{R}^d\right)$ iff $E_{\mu_X}[|x|^p] < \infty$.

Our aim is to prove the following result:

Theorem 1. *Let $X \in L^1\left(\mathbb{R}^d\right)$ be given, and let $C \subset L^1(\mathbb{R}^d)$ be the closed convex hull of all random variable Y such that $\mu_X = \mu_Y$. Take any $Z \in L^\infty\left(\mathbb{R}^d\right)$ the law of which is absolutely continuous with respect to Lebesgue measure. Then there exists a unique $\overline{X} \in C$ where Z attains its maximum on C. The law of $\overline{X}$ is μ_X, and for every sequence $X_n \in C$ such that*

$$\left\langle Z, X_n \right\rangle \to \left\langle Z, \overline{X} \right\rangle$$

we have $\|X_n - \overline{X}\|_1 \to 0$.

This will be proved as Theorem 18 at the end of this paper. In addition, Theorem 19 will provide a converse.

2. Preliminaries

2.1. **Law-invariant subsets and functions.** We shall write $X_1 \sim X_2$ to mean that X_1 and X_2 have the same law. This is an equivalence relation on the space of random vectors. A set $C \subset L^0$ will be called *law-invariant* if:

$$[X_1 \in C \text{ and } X_1 \sim X_2] \implies X_2 \in C,$$

and a function $\varphi : L^0 \to R$ is law-invariant if $\varphi(X_1) = \varphi(X_2)$ whenever $X_1 \sim X_2$. Given $\mu \in \mathcal{P}(\mathbb{R}^d)$, we shall denote by $M(\mu)$ the equivalence class consisting of all X with law μ:

$$M(\mu) := \{X \mid \mu_X = \mu\}$$

The set $M(\mu)$ is not convex in general.

Lemma 2. *If μ has finite p-moment, $\int |x|^p d\mu < \infty$, for $| \leqslant p \leqslant \infty$, the set $C(\mu)$ is closed in the L^p-norm.*

Proof. If $X_n \in C(\mu)$ and $\|X_n - X\|_p \to 0$, then we can extract a subsequence which converges almost everywhere. If $f \in \mathcal{C}^b(\mathbb{R}^d)$, applying Lebesgue's dominated convergence theorem, we have $\int f(x)dP = \lim_n \int f(X_n)dP$ for every $f \in \mathcal{C}^b(\mathbb{R}^d)$. But the right-hand side is equal to $\int f(x)d\mu$ for every n. □

We shall say that $\sigma : \Omega \to \Omega$ is a *measure-preserving transformation* if it is a bijection, σ and σ^{-1} are measurable, and $P(\sigma^{-1}(A)) = P(A) = P(\sigma(A))$ for all $A \in \mathcal{A}$. The set Σ of all measure-preserving transformations is a group which operates on random vectors and preserves the law:

$$\forall \sigma \in \Sigma,\ \forall X \in L^0,\ \ X \sim X \circ \sigma.$$

The converse is not true, that is, equivalence classes do not coincide with orbits for the group action. However, it comes close. By Lemma A.4 from [2], we have:

Proposition 3. *Let C be a norm-closed subset of $L^p(\Omega, \mathcal{A}, P)$, $1 \leqslant p \leqslant \infty$. Then C is law-invariant if and only if it is transformation-invariant. As a consequence:*

$$\forall X \in M(\mu), \quad M(\mu) = \overline{\{X \circ \sigma \mid \sigma \in \Sigma\}}$$

the closure being taken L^p-norm.

2.2. **Choquet ordering of probability laws.** Denote by $\mathcal{P}(\mathbb{R}^d)$ the space of probability laws on $\mathbb{R}^d$, and endow it with the weak-* topology induced by $C^0(\mathbb{R}^d)$, the space of continuous functions on $\mathbb{R}^d$ which go to zero at infinity. It is known that there is complete metric on $\mathcal{P}(\mathbb{R}^d)$ which is compatible with this topology:

$$[\mu_n \to \mu \quad \text{weak-*}] \Longleftrightarrow \left[\forall f \in C^0(R^d),\ \int f_n d\mu \to \int f d\mu\right]$$

Denote by $\mathcal{P}_1(\mathbb{R}^d)$ the set of probability laws on $\mathbb{R}^d$ which have finite first moment:

$$\mu \in \mathcal{P}_1(\mathbb{R}^d) \Longleftrightarrow \int_{\mathbb{R}^d} |x|\, d\mu < \infty$$

Note that $\mathcal{P}_1(\mathbb{R}^d)$ is convex, but not closed in $\mathcal{P}(\mathbb{R}^d)$. If $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, every linear function $f(x)$ is μ-integrable. The point:

$$x := \int_{\mathbb{R}^d} y d\mu(y)$$

will be called the *barycenter* of the probability μ.

Since every convex function on $\mathbb{R}^d$ is bounded below by an affine function, we find that $\mathbb{E}_\mu[f]$ is well-defined (possibly $+\infty$) for every convex function. So the following definition makes sense:

Definition 4. For ν and μ in $\mathcal{P}_1\left(\mathbb{R}^d\right)$, we shall say that $\nu \preccurlyeq \mu$ if, for every convex function $f : \mathbb{R}^d \to \mathbb{R}$, we have:

$$\int_{\mathbb{R}^d} f(x)\, d\nu \leqslant \int_{\mathbb{R}^d} f(x)\, d\mu$$

For technical reasons, in order to avoid infinities, we shall introduce an equivalent definition. Denote by $\mathcal{C}$ the set of convex functions $f : \mathbb{R}^d \to \mathbb{R}$ which have linear growth:

$$\exists M, m : \forall x, \quad f(x) \leqslant m + M\|x\|$$

If $f \in \mathcal{C}$ then $\int f(x) d\mu < \infty$.

Lemma 5. *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then $g = \sup_n f_n$, for some increasing sequence $f_n \in \mathcal{C}$.*

Proof. Define the family f_n as follows:

$$\begin{aligned} f_n(x) &= \sup\left[\{\langle y, x\rangle - a \mid (y, a) \in A_n\}, 0\right] \\ A_n &= \{(y, a) \mid \|y\| \leqslant n \ \text{ and } \ \langle y, x\rangle - a \leqslant f(x) \quad \forall x\} \end{aligned}$$

□

Two results follow immediately:

Lemma 6. *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then the linear functional $\mu \to \int g(x) d\mu$ is l.s.c on $\mathcal{P}_1(\mathbb{R}^d)$*

Proof. We can find an affine functional $\ell(x)$ such that $\ell \leqslant f_n \leqslant g$ for all n. Since $\mu \in \mathcal{P}_1$, the function ℓ is integrable, and we can apply the motone convergence theorem:

$$\int g(x) d\mu = \sup_n \int f_n d\mu$$

Since each map $\mu \to \int f_n d\mu$ is weak-* continuous, the map $g \to \int g(x) d\mu$ is weak-* l.s.c. □

Lemma 7. *For $\nu, \mu \in P_1(\mathbb{R}^d)$, $\nu \preceq \mu$ holds iff:*

$$\int f(x) d\nu \leqslant \int f(x) d\mu$$

for every $f \in \mathcal{C}$.

Proof. For any g convex, we have, by the preceding lemma $g = \sup_m f_m$, for some sequence $f_m \in \mathcal{C}$. The inequality holds for each f_m, and we conclude by Fatou's lemma. □

This is an (incomplete) order relation on the set of probability measures with finite first moment. It is known in potential theory as the *Choquet ordering* (see [5], chapter XI.2). Note that if f is linear, both f and $-f$ are convex, so that, if $\nu \preccurlyeq \mu$, then:

$$\int_{\mathbb{R}^d} f(x)\, d\nu = \int_{\mathbb{R}^d} f(x)\, d\mu$$

In particular, if $\nu \lesssim \mu$ then ν and μ have the same barycenter.

Informally speaking, $\nu \preccurlyeq \mu$ means that they have the same barycenter, but μ is more spread out than ν. In potential theory, this is traditionally expressed by saying that "μ est une balayée de ν", that is, "μ is swept away from ν". The following elementary properties illustrates this basic intuition:

(1) (*certainty equivalence*) If $x_0 = E_\mu[x]$ (x_0 is the barycenter of μ) and δ_{x_0} is the Dirac mass carried at x_0, then $\delta_{x_0} \leqslant \mu$
(2) *(diversification)* If $X_1 \sim X_2$ have law μ, and $Y = \frac{1}{2}(X_1 + X_2)$ has law ν, then $\nu \leqslant \mu$. Indeed, if f is convex:

$$\int_{\mathbb{R}^d} f(x)\,d\nu = \int_\Omega f(Y)\,dP \leqslant \frac{1}{2}\int_\Omega f(X_1)\,dP + \frac{1}{2}\int_\Omega f(X_2)\,dP$$
$$= \left(\frac{1}{2} + \frac{1}{2}\right)\int_{\mathbb{R}^d} f(x)\,d\mu = \int_{\mathbb{R}^d} f(x)\,d\mu$$

Lemma 8. *Let $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ and let $I[\mu]$ be the Choquet order interval of μ in $\mathcal{P}_1(\mathbb{R}^d)$*

$$I[\mu] = \{\nu \in \mathcal{P}_1(\mathbb{R}^d) : \nu \lll \mu\}.$$

Then $I[\mu]$ is a compact subset of $\mathcal{P}(\mathbb{R}^d)$ with respect to the weak-star topology induced by $C^0(\mathbb{R}^d)$.

Proof. As the weak-star topology on $\mathcal{P}_1(\mathbb{R}^d)$ is metrisable it will suffice to show that every sequence $(\nu_n)_{n=1}^\infty$ in $I[\mu]$ has a cluster point.

The relation $\nu_n \lll \mu$ implies in particular that the first moment of ν_n are bounded by the first moment of μ. This in turn implies that Prokhorov's condition is satisfied, i.e. for $\varepsilon > 0$ there is a compact $K \subseteq \mathbb{R}^d$ such that $\nu_n(K) \geqslant 1 - \varepsilon$, for all $n \in \mathbb{N}$.

By Prokhorov's theorem we may find a subsequence, still denoted by $(\nu_n)_{n=1}^\infty$, and a probability measure $\nu \in \mathcal{P}(\mathbb{R}^d)$ which is the weak-star limit. To show that $\nu \in I[\mu]$, let $f : \mathbb{R}^d \to \mathbb{R}$ be convex. By Lemma 6, we have:

$$\langle f, \nu\rangle \leqslant \limsup_{n\to\infty} \langle f, \nu_n\rangle \leqslant \langle f, \mu\rangle.$$

□

The relationship with weak convergence in L^1 is given by the next result. To motivate it, consider a sequence of i.i.d. random variables X_n such that $P[X = -1] = 1/2 = P[X = 1]$. Then $X_n \to 0$ weakly, and the law of the limit is δ_0, but all the X_n have the law $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. Clearly $\delta_0 \lll \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$.

Proposition 9. *Suppose X_n is a sequence in $L^1(\Omega, \mathcal{A}, P; \mathbb{R}^d)$, converging weakly to Y. Denote by μ_n the law of X_n and by ν the law of Y. Suppose μ_n converges weak-* to some $\bar{\mu} \in \mathcal{P}_1(\mathbb{R}^d)$. Then $\nu \leqslant \bar{\mu}$, with equality if and only if $\|X_n - Y\|_1 \to 0$*

Proof. First note that $\mu \succeq \delta_{E[y]}$. Indeed, for any convex function f with linear growth, we have, by Jensen's inequality:

$$\int f(x)d\mu_n = \int_\Omega f(X_n)dP \geqslant f(E[X_n])$$

and the left hand side converges to $\int f(x)d\mu$ while the right-hand side converges to $f(E[y])$.

Now consider a finite σ-algebra $G \subset \mathcal{F}$. Denote by $\mathcal{A}$ the collection of atoms of $\mathcal{G}$. We have:

$$\int f(x)d\mu_n = \int E[f(X_n)|G]dP$$

and by the same method we show that:

$$\mu \succeq \sum_{A\in G} P[A]\delta_{E[Y|A]}$$

Now let $(\mathcal{G}_k), k \in \mathbb{N}$, be a sequence of finite sub-sigma-algebras of $\mathcal{F}$ such that Y is measurable w.r.t. $\sigma(\cup_k \mathcal{G}_k)$. Denoting by ν_k the law of $E[Y \mid \mathcal{G}_k]$, we have by the above argument:

$$\bar{\mu} \geq \nu_k \text{ for all } k$$

and hence $\bar{\mu} \geq \nu$ by taking the limit when $k \to \infty$.

Turning to the final assertion, it follows from Lebesgue's dominated convergence theorem that, since X_n converges to Y in the L^1 norm, the law μ_n of X_n converges to the law ν of Y weak-* in $\mathcal{P}_1(\mathbb{R}^d)$.

Conversely suppose that $(X_n)_{n=1}^{\infty}$ converges to Y weakly in $L^1(\mathbb{R}^d)$ and $\bar{\mu} \preccurlyeq \nu$. For every $A \in \mathcal{F}$ and every convex function f with $\limsup_{x\to\infty} \frac{|f(x)|}{(x)} < \infty$ we then have

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)\mathbb{1}_A] = \mathbb{E}[f(Y)\mathbb{1}_A]. \tag{2.1}$$

Indeed the inequality $\geqslant$ follows from Jensen as above. The reverse inequality follows from the fact that

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \lim_{n\to\infty} \langle f, \mu_n \rangle = \langle f, \bar{\mu} \rangle = \langle f, \nu \rangle = \mathbb{E}[f(Y)].$$

In conjunction with

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)\mathbb{1}_{\Omega\setminus A}] \geqslant \mathbb{E}[f(Y)\mathbb{1}_{\Omega\setminus A}]$$

this yields the inequality $\leqslant$ in (2.1).

Now suppose that $(X_n)_{n=1}^{\infty}$ fails to converge to Y in the norm of $L^1(\mathbb{R}^d)$, i.e., there is $1 > \alpha > 0$ such that

$$\mathbb{P}[|X_n - Y| \geqslant \alpha] \geqslant \alpha,$$

for infinitely many n. Approximating Y by step functions we may find a set $A \in \mathcal{F}$, $P[A] > 0$, and a point $y_0 \in A$ such that $|Y - y_0| < \frac{\alpha^2}{5}$ on A and

$$\mathbb{P}[A \cap |X_n - <_0| \geqslant \tfrac{\alpha}{2}] \geqslant \tfrac{\alpha}{2}\mathbb{P}[A].$$

We then have

$$\mathbb{E}[|Y - y_0|\mathbb{1}_A] \leqslant \tfrac{\alpha^2}{5}\mathbb{P}[A]$$

while

$$\mathbb{E}[|X_n - y_0|\mathbb{1}_A] \leqslant \tfrac{\alpha^2}{4}\mathbb{P}[A],$$

a contradiction to (2.1). □

The Choquet ordering can be completely characterized in terms of Markov kernels

Definition 10. A Borel map $\alpha : \mathbb{R}^d \to \mathcal{P}_1(\mathbb{R}^d)$ is a *Markov kernel* if, for every $x \in X$, the barycenter of α_x is x:

$$\forall x \in X, \quad \int_{\mathbb{R}^d} y d\alpha_x = x$$

If α is a Markov kernel, and $\nu \in \mathcal{P}(\mathbb{R}^d)$, we define $\mu := \int_{\mathbb{R}^d} \alpha_x d\nu \in \mathcal{P}(\mathbb{R}^d)$ by:

$$\int_{\mathbb{R}^d} f(x)\, d\mu = \int_{\mathbb{R}^d} \alpha_x(f)\, d\nu$$

Theorem 11. *If ν and μ are in $\mathcal{P}_1(\mathbb{R}^d)$ we have $\nu \preccurlyeq \mu$ if and only if there exists a Markov kernel α_x such that $\mu = \int_{\mathbb{R}^d} \alpha_x d\nu$*

Proof. Suppose there exists such a Markov kernel. For any convex function f, since x is the barycenter of α_x, Jensen's inequality implies that $\alpha_x(f) \geqslant f(x)$. Integrating, we get:

$$\int_{\mathbb{R}^d} f(x)\, d\mu = \int_{\mathbb{R}^d} \alpha_x(f) d\nu \geqslant \int_{\mathbb{R}^d} f(x)\, d\nu$$

so $\nu \preccurlyeq \mu$. The converse is known as Strassen's theorem (see [7], [5]) □

2.3. **Optimal transport.** In the sequel, μ and ν will be given in $\mathcal{P}_1(\mathbb{R}^d)$, and μ will have bounded support. We are interested in the following problem: maximize

$$\int_{\mathbb{R}^d} \langle x, T(x) \rangle\, d\mu$$

among all Borel maps $T : \mathbb{R}^d \to \mathbb{R}^d$ which map μ on ν:

$$T\natural\mu = \nu \iff \int f(y) d\nu = \int f(T(x)) d\mu \ \forall f \in C^0(\mathbb{R})$$

In the sequel, this will be referred to as the *basic problem,* and denoted by $(\mathrm{BP}[\mu,\nu])$. If there is an optimal solution T, it has the property that if X is any r.v. with law μ, then, among all r.v. Y with law ν, the one such that the correlation $E_\mu[\langle X, Y\rangle]$ is maximal is $T(X)$.

There is also a *relaxed problem*, denoted $(\mathrm{RP}[\mu,\nu])$. It consists of maximizing:

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle\, d\lambda$$

among all probability measures λ on $\mathbb{R}^d \times \mathbb{R}^d$ which have μ and ν as marginals. Obviously, we have $\sup(BP) \leqslant \sup(RP)$, and the latter is finite because μ has bounded support and ν has finite first moment.

Finally, there is a *dual problem*, defined by $(\mathrm{DP}[\mu,\nu])$, which consists of minimizing

$$\int_{\mathbb{R}^d} \varphi(x) d\mu + \int_{\mathbb{R}^d} \psi(y) d\nu$$

over all pairs of functions $\varphi(x)$ and $\psi(y)$ such that $\varphi(x) + \psi(y) \geqslant \langle x, y\rangle$.

The following theorem summarizes results due to Kantorovitch [3], Kellerer [4] Rachev and Ruschendorf [6], and Brenier [1]. It was originally formulated for the case when μ and ν have finite second moment, and this is also what is found in [8]. Indeed, in this case, since $T\natural\mu = \nu$,we have:

$$\begin{aligned}\int \|x - T(x)\|^2 d\mu &= \int \|x\|^2 d\mu + \int \|T(x)\|^2 d\mu - 2\int \langle x, T(x)\rangle\, d\mu \\ &= \int \|x\|^2 d\mu + \int \|y\|^2 d\nu - 2\int \langle x, T(x)\rangle\, d\mu\end{aligned}$$

Since the two first terms on the right-hand side do not depend on T, the problem of maximising $\int \langle x, T(x)\rangle\, d\mu$ (bilinear cost) is equivalent to the problem of maximizing $\int \|x - T(x)\|^2 d\mu$ (quadratic cost), for which general techniques are available. In the case at hand, we will not assume that ν has finite second moment, so this approach is not available: the square distance is not defined, while the correlation maximisation still makes sense.

We now recall Brenier's theorem [1] in the present setting. In order to obtaine a transport of Monge type rather than Kantorovich type, we assume that μ is absolutely continuous w.r.t Lebesgue measure. We note in passing that it would be sufficient to assume that only that μ does not "give mass to small sets", i.e. that sets of Haussdorff dimension $d-1$ have μ-measure 0 ([9]

Theorem 12. *Suppose μ has compact support and is absolutely continuous w.r.t. Lebesgue measure. Suppose also ν has finite first moment. Then the basic problem ($BP[\mu,\nu]$) has a solution T, which is unique up to negligible subsets, and there is a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that $T(x) = \nabla\varphi(x)$ a.e..*

The relaxed problem ($RP[\mu,\nu]$) has $\lambda = \int \delta_{T(x)} d\mu(x)$ as a unique solution.

Denoting by ψ the Fenchel transform of φ, all solutions to the dual problem ($DP[\mu,\nu]$) are of the form $(\varphi + a, \psi - a)$ for some constant a, up to μ-, resp ν-., a.s. equivalence. The values of the minimum in problem (DP) and of the maximum in problems (BP) and (RP) are equal:

$$\max(BP[\mu,\nu]) = \max(RP[\mu,\nu]) = \min(DP[\mu,\nu]) \tag{2.2}$$

Let us denote by $\mathbf{mc}[\mu,\nu]$ this common value. We shall call it the *maximal correlation* between μ and ν. It follows from the theorem that for any $T', \lambda', \varphi', \psi'$ satisfying the admissibility conditions, we have:

$$\int_{\mathbb{R}^d} \langle x, T'(x) \rangle \, d\mu \leqslant \mathbf{mc}[\mu,\nu]$$

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle \, d\lambda' \leqslant \mathbf{mc}[\mu,\nu]$$

$$\int_{\mathbb{R}^d} \varphi'(x) d\mu + \int \psi'_{\mathbb{R}^d}(y) d\nu \geqslant \mathbf{mc}[\mu,\nu]$$

As an interesting consequence, we have:

Proposition 13. *Let μ,ν_1, ν_2 be probability measures on $\mathbb{R}^d$ such that μ is absolutely continuous w.r.t the Lebesgue measure and has bounded support, while ν_1 and ν_2 have finite first moment. Suppose $\nu_1 \preccurlyeq \nu_2$ and $\nu_1 \neq \nu_2$. Then $\mathbf{mc}[\mu,\nu_1] < \mathbf{mc}[\mu,\nu_2]$.*

Proof. By Theorem 11, there is a Markov kernel α such that:

$$\nu_2 = \int_{\mathbb{R}^d} \alpha_x d\nu_1 \tag{2.3}$$

Let T_1be the optimal solution of (BP$[\mu,\nu_1]$). Consider the probability measure λ on $\mathbb{R}^d \times \mathbb{R}^d$ defined by:

$$\int f(x,y)\, d\lambda(x,y) = \int d\mu(x) \int f(x,y)\, d\alpha_{T_1(x)}(y) \tag{2.4}$$

Since $\alpha_{T_1(x)}$ is a probability measure, the first marginal of λ is μ. Let us compute the second marginal. We have, for any $f \in C^0(\mathbb{R}^d)$,

$$\begin{aligned}
\int_{\mathbb{R}^d \times \mathbb{R}^d} f(y) d\lambda(x,y) &= \int_{\mathbb{R}^d} \alpha_{T_1(x)}(f) d\mu(x) \\
&= \int_{\mathbb{R}^d} \alpha_x(f) d\nu_1(x) \\
&= \nu_2(f)
\end{aligned}$$

where the second equality comes from the fact that T_1 maps μ on ν_1 and the second from equation (2.3). So the second marginal of λ is ν_2, and λ is admissible in problem (RP$[\mu,\nu_2]$). A similar computation gives:

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d}\langle x,y\rangle\,d\lambda(x,y) &= \int_{\mathbb{R}^d}\left\langle x,\int_{\mathbb{R}^d}d\alpha_{T_1(x)}(y)\right\rangle d\mu\left(x\right)\\ &= \int_{\mathbb{R}^d}\langle x,T_1(x)\rangle\,d\mu\left(x\right)=\mathbf{mc}[\mu,\nu_1]\end{aligned}$$

Since λ has marginals μ and ν_2, it is admissible in the relaxed problem(RP$[\mu,\nu_2]$), so that the left-hand side is at most $\mathbf{mc}[\mu,\nu_2]$ while the right-hand side is equal to $\mathbf{mc}[\mu,\nu_1]$. It follows that $\mathbf{mc}[\mu,\nu_1]\leqslant\mathbf{mc}[\mu,\nu_2]$. If there is equality, then λ is an optimal solution to (RP$[\mu,\nu_2]$). By the uniqueness part of Theorem 12, we must have $\lambda=\int\delta_{T_1(x)}d\mu(x)$. Comparing with equation (2.4), we find $\alpha_y=\delta_y$, holding true ν_1-almost surely. Writing this in equation (2.3) we get $\nu_1=\nu_2$. □

2.4. **Strongly exposed points.** Let E be a Banach space, and $C\subset E$ a closed subset. For $v\in E'$, consider the optimization problem:

$$\sup_{u'\in C}\left\langle v,u'\right\rangle \tag{2.5}$$

Definition 14. We say that $v\in E'$ *exposes* $u\in C$ if u solves problem (2.5) and is the unique solution. We shall say that $v\in E'$ it strongly exposes $u\in C$ if it exposes u and all maximizing sequences in problem (2.5) converge to u:

$$\left\{u_n\in C,\quad \lim_n\langle v,u_n\rangle=\langle v,u\rangle\right\}\Longrightarrow\lim_n\|u-u_n\|=0$$

We shall say that $u\in C$ is an *exposed point* (resp. *strongly exposed*) if it is exposed (resp. strongly exposed) by some continuous linear functional v. It is a classical result of Phelps that every weakly compact convex subset C of E is the closed convex hull of its strongly exposed points.

3. Some geometric properties of law-invariant subsets of L^1

Given $\nu\in\mathcal{P}_1\left(\mathbb{R}^d\right)$, we define two subsets $M(\nu)$ and $C(\nu)$ of $L^1$by:

$$\begin{aligned}M\left(\nu\right)&=\left\{X\in L^1\mid\mu_X=\nu\right\}\\ C\left(\nu\right)&=\left\{X\mid\mu_X\preccurlyeq\nu\right\}\end{aligned}$$

$M(\nu)$ is closed in L^1but not convex. To investigate the relation between $M\left(\nu\right)$ and $C\left(\nu\right)$, we shall need the following result:

Proposition 15. *Let $Y\in L^1(\mathbb{R}^d)$ with law ν, and let $\mu\in\mathcal{P}^1(\mathbb{R}^d)$ such that $\mu\gtrsim\nu$. Then there is a sequence $(X_n)_{n=1}^\infty$ in $M(\mu)$ such that $(X_n)_{n=1}^\infty$ converges weakly to Y in $L^1(\mathbb{R}^d)$.*

In addition, there is a sequence $(Y_n)_{n=1}^\infty$ in the convex hull of $(M(\mu))$ which converges strongly to Y in $L^1(\mathbb{R}^d)$.

We start by recalling a well-known result from ergodic theory.

Lemma 16. *Let $\Omega=\{-1,1\}^{\mathbb{Z}}$ equipped with the Borel sigma-algebra $\mathcal{F}$ and Haar-measure $\mathbb{P}$, and let T_n the n'th shift, i.e.*

$$T_n\left((\eta_k)_{k\in\mathbb{Z}}\right)=(\eta_{k-n})_{k\in\mathbb{Z}}$$

Let $Z \in L^1(\{-1,1\}^{\mathbb{Z}}, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$*. Then* $(Z_n)_{n=1}^\infty := (Z \circ T_n)_{n=1}^\infty$ *converges weakly to the constant function* $\mathbb{E}[Z]$.

Proof. Suppose that Z depends only on finitely many coordinates and let $A \in \mathcal{F}$ also depend only on finitely many coordinates of $\{-1,1\}^{\mathbb{Z}}$. Then, for n large enough, $Z_n := Z_0 T_n$ is independent of A so that

$$\mathbb{E}[Z_n|A] = \mathbb{E}[Z_n] = \mathbb{E}[Z].$$

The general case follows from approximation. □

Proof. (of the Proposition): Assume w.l.o.g. that $L^1(\Omega)$ is separable. Recall that, $(\Omega, \mathcal{F}, \mathbb{P})$ has no atoms. Suppose first that Y takes only finitely many values, i.e.

$$Y = \sum_{j=1}^N y_j \mathbb{1}_{A_j}$$

where $(y_j)_{j=1}^N \in \mathbb{R}^d$ and $(A_1, \ldots, A_N)$ forms a partition of Ω into sets in $\mathcal{F}$ with strictly positive measure.

By an elementary version of Strassen's theorem we may find a Markov kernel $\alpha = (\alpha_{y_j})_{j=1}^N$ such that the barycenter of α_{y_j} is y_j and

$$\mu = \sum_{j=1}^N \mathbb{P}[A_j]\alpha_{y_j} \tag{3.1}$$

Each of the sets A_j, equipped with the conditional probability $P[A_j]^-1P|_{A_j}$ is Borel isomorphic to $\{-1,1\}^{\mathbb{Z}}$, equipped with the Haar measure. Hence for each $j = 1, \ldots, N$ we may find a random variable $Z_j : A_j \to \mathbb{R}^d$ under $P[A_j]^-1P|_{A_j}$ such that law $(Z_j) = \alpha_{y_j}$, so that its barycenter equals y_j, as well as a sequence $(T_{j,n})_{n=1}^\infty$ of measure-preserving transformations of A_j such that in the weak $L^1(\mathbb{R}^d)$ topology we have:

$$\lim_{n\to\infty} (Z_j \circ T_n)\mathbb{1}_{A_j} = y_j \mathbb{1}_{A_j}, \qquad j = 1, \ldots, N.$$

Letting

$$X_n = \sum_{j=1}^N (Z_j \circ T_n)\mathbb{1}_{A_j}$$

we obtain by (3.1) a sequence in $L^1(\mathbb{R}^d)$ such that the law of (X_n) is $= \mu$, converging weakly to $Y = \sum_{j=1}^N y_j \mathbb{1}_{A_j}$.

Now drop the assumption that Y is a simple function and fix a sequence $(\mathcal{G}_m)_{m=1}^\infty$ of finite sub-sigma-algebras of $\mathcal{F}$, generating $\mathcal{F}$. Note that for $Y_m = \mathbb{E}[Y|\mathcal{G}_m]$ and $\nu_m =$ law (Y_m) we have that $\nu_m \prec \nu$, by Jensen's inequality.

By the first part we may find, for each $m \geqslant 1$, a sequence $(X_{m,n})_{n=1}^\infty$ in $M(\mu)$ such that $(X_{m,n})_{n=1}^\infty$ converges weakly to Y_m . Noting that $(Y_m)_{m=1}^\infty$ converges to Y (in the norm of $L^1(\mathbb{R}^d)$ and therefore also weakly) we may find a sequence $(n_m)_{m=1}^\infty$ tending sufficiently fast to infinity, such that $(X_{m,n_m})_{m=1}^\infty$ converges weakly to Y.

As regards the final assertion it follows from Komlos' theorem that there is a sequence of convex combinations of the above $(X_{m,n_m})_{m=1}^\infty$ converging to Y in the norm of $L^1(\mathbb{R}^d)$. □

The relationship between $C(\nu)$ and $M(\mu)$ follows immediately:

Theorem 17. *The set $C(\nu)$ is convex, weakly compact, and equals the weak closure of $M(\nu)$:*

$$C(\nu) = \overline{M(\nu)}^w = \overline{co}\, M(\nu)$$

Proof. Obviously $\overline{M(\nu)}^w \subset C(\nu)$. Conversely, take any $X \in C(\nu)$. By Theorem 11, there is some Markov kernel α such that

$$\nu = \int_{\mathbb{R}^d} \alpha_x d\mu_X = \int_\Omega \alpha_{X(\omega)} dP$$

By the above proposition, there is a sequence X_n in $M(\nu)$ such that $X_n \to X$ weakly, so $X \in \overline{M(\nu)}^w$. This shows that $C(\nu) = \overline{M(\nu)}^w$.

By Proposition 9, $C(\nu)$ is convex. It remains to show that it is weakly compact. Since $C(\mu)$ is the weak closure or $M(\mu)$, it is enough to show that $M(\mu)$ is weakly relatively compact. To do that, we shall use the Dunford-Pettis criterion. We claim that $M(\nu)$ is equiintegrable. Indeed, for any $X \in M(\nu)$ and $m > 0$, we have:

$$\int_{|X|>m} |X| dP = \nu(|x| > m)$$

which goes to 0 when $m \to \infty$, independently of X. The result follows. □

We now investigate strongly exposing functionals and strongly exposed points of $C(\nu)$. We will show that any $Z \in L^\infty$, the law of which is a.c. w.r.t. Lebesgue measure, strongly exposes a point of $C(\nu)$ (which must then belong to $M(\nu)$) and conversely, provided ν is absolutely continuous w.r.t. Lebesgue measure, that any point of $M(\nu)$ is strongly exposed by such a Z.

Theorem 18. *Let $\nu \in \mathcal{P}_1(\mathbb{R}^d)$, $Z \in L^\infty$ and suppose the law of Z is absolutely continuous with respect to Lebesgue measure. Then Z strongly exposes some point of $C(\nu)$, and the exposed point in fact belongs to $M(\nu)$*

Proof. Let μ be the law of Z and consider the maximal correlation problem (BP$[\mu,\nu]$). By Theorem 12, it has a unique solution T. Set $X = T(Z)$. Clearly X has law ν, and by uniqueness:

$$[X' \in M(\nu),\ X' \neq X] \Longrightarrow \langle Z, X\rangle > \langle Z, X'\rangle \tag{3.2}$$

So X is an exposed point in $M(\nu)$. Take any $Y \in C(\nu)$, so that $\mu_Y \preccurlyeq \nu$. By Proposition 13, we have $\langle Z, X\rangle \geqslant \langle Z, Y\rangle$, and if $\langle Z, X\rangle = \langle Z, Y\rangle$, then $\mu_Y = \mu_X = \nu$. So Y must belong to $M(\nu)$, and by formula (3.2), we must have $Y = X$. So X is an exposed point in $C(\nu)$ as well.

It remains to prove that it is strongly exposed. For this, take a maximizing sequence X_n in $C(\nu)$. Since $C(\nu)$ is weakly compact and $\nu_n \preceq \nu$, where ν_n is the law of X_n, there is a subsequence X_{n_k}which converges weakly to some $X' \in C(\nu)$. By Proposition , the set of all $\mu \precsim \nu$ is weak-$*$ compact, so we may assume that the laws ν_{n_k} converge weak-$*$ to some $\bar{\nu}$. Obviously X' maximizes $\langle Z, X'\rangle$, and since Z exposes X, we must have $X' = X$. So the X_{n_k}converge weakly to X, and, by Proposition 9, $\mu_X = \nu \preccurlyeq \overline{\nu}$.

On the other hand, take any convex function f with linear growth. Since $\nu_{n_k} \preccurlyeq \nu$ we have:

$$\int f(x) d\nu_{n_k} \leqslant \int f(x) d\nu$$

Letting $k \to \infty$, we get:

$$\int f(x)d\overline{\nu} \leqslant \liminf_k \int f(x)d\nu_{n_k} \leqslant \int f(x)d\nu$$

So $\nu = \overline{\nu}$, and Proposition 9 then implies that $\|X_{n_k} - X\|_1 \to 0$. Since the limit does not depend on the subsequence, the whole sequence X_n converges, and X is strongly exposed, as announced. □

Here is a kind of converse:

Theorem 19. *Fix two measures μ and ν on $\mathbb{R}^d$, the first one having finite first moment and the second one compact support. Suppose* both of them *are absolutely continuous with respect to Lebesgue measure. Then, for every X with law μ, there is a unique Z with law ν which strongly exposes X in $C(\mu)$.*

Proof. Consider the maximal correlation problem (BP$[\mu,\nu]$). It has a unique solution $T : \mathbb{R}^d \to \mathbb{R}^d$ verifying $T\sharp\mu = \nu$. Since both μ and ν are absolutely continuous with respect to Lebesgue measure, the problem (BP$[\mu,\nu]$) also has a unique solution $S : \mathbb{R}^d \to \mathbb{R}^d$ verifying $S\sharp\mu = \nu$. Clearly $S = T^{-1}$ and $T = S^{-1}$. Define $Z = S(X)$. It is then the case that the law of Z is μ and $T(Z) = T \circ S(X) = X$. Repeating the preceding proof we find that Z strongly exposes X in $C(X)$. □

Note that the condition that ν be absolutely continuous with respect to the Lebesgue measure cannot be dropped from the preceding theorem. This may be seen by a variant of a well-known example in optimal transport theory ([9], Example 4.9). On $\mathbb{R}^2$ consider the measure ν which is uniformly distributed on the interval $\{0\} \times [0,\ 1]$ while μ is uniformly distributed on the rectangle $[-1,\ 1] \times [0,\ 1]$. Then μ is absolutely continuous w.r.t. Lebesgue measure, while ν is not. Clearly the optimal transport T from μ to ν for the maximal correlation problem is given by the projection on the vertical axis. This map is not invertible.

Let $(\Omega, \mathcal{A}, P)$ be given by $\Omega = [0,\ 1]$ equipped with the Lebesgue measure P on the Borel σ-algebra. Define a random vector $X \in L^1(\Omega, \mathcal{A}, P; \mathbb{R}^2)$ by $X(\omega) = (0, \omega)$, so that the law of X is ν. Let us now calculate the maximal correlation between μ and ν. Let $Z_0 \in L^\infty$ have law μ. and define $X_0 = T(Z_0)$ so that X_0 has law ν. By the proof of theorem 18 we get:

$$\begin{aligned}\mathbf{mc}(\mu,\nu) &= \int_\Omega \langle X_0, Z_0\rangle\, dP = \int_{\mathbb{R}^2} \langle x, T(x)\rangle\, d\mu \\ &= \frac{1}{2}\int_{-1}^{1}\left[\int_0^1 x_2^2 dx_2\right] dx_1 = \int_0^1 x_2^2 dx_2 = \frac{1}{3}\end{aligned}$$

On the other hand, we claim that:

$$\int \langle X, Z_0\rangle\, dP < \frac{1}{3} \tag{3.3}$$

Since this holds for any Z_0 with law μ, it shows that X does not expose any point in $C(\mu)$. This is the desired counterexample.To prove (3.3), write $Z_0(\omega) = (Z_{0,1}(\omega), Z_{0,2}(\omega))$ and note that $P[Z_{0,2} \neq X_2] > 0$. Indeed, assume otherwise, so that $Z_{0,2}(\omega) = X_2(\omega) = \omega$ almost surely. Then $Z_{0,1}(\omega)$ is fully determined by $Z_{0,2}(\omega)$, meaning that, in the image of Ω by Z, the second coordinate z_1 is determined by the first z_2. This contradicts the fact that the law of Z is μ, which is absolutely continuous. Since the law of $Z_{0,2}$ is the Lebesgue measure, but $Z_{0,2}$ does not coincide with X, we have, from the uniqueness of the Brenier map:

$$\int XZ_0 dP = \int X_2 Z_{0,2} dP < \int X_2^2 dP = \frac{1}{3}$$

Let us summarize our findings: There are measures μ and ν on $\mathbb{R}^2$ with compact support, μ being absolutely continuous with respect to Lebesgue measure, and some $X \in L^\infty\left(\mathbb{R}^2\right)$ with law ν such that there is no $Z \in L^\infty\left(\mathbb{R}^2\right)$ wich exposes X in $C\left(\mu\right)$

References

[1] Brenier, Yann, "Polar factorization and monotone rearrangements of vector-valued functions", Comm. Pure and Applied Math. 64, 375-417 (1991)

[2] Elyes, Jouini, Walter Schachermayer and Nizar Touzi, "Law-invariant risk measures have the Fatou property", Adv. Math. Econ. 9, 49-71

[3] Kantorovitch. L.V. "On the transfer of masses", Doklady Akad. Nauk. CCCR (1942), 113-161

[4] Kellerer, H.G. "Duality theorems for marginal problems" Z. Wahrsch. Verw. Gebiete 67 (1984), 399-432

[5] Meyer, Paul-André, "Probabilités et potentiel", Hermann (1966)

[6] Rachev, S.T. and Rüschendorf, L. , "A characterization of random variables with minimal L^2 distance", J. Multivariate An. 32, 48-54 (1990)

[7] Strassen, V. "The existence of probability measures with given marginals" Ann. Math. Stat 36 (1965) p.423-439

[8] Villani, Cedric, "Topics in optimal transportation", Publications of the AMS, 2003

[9] Villani, Cedric "Optimal transport, old and new", Springer Verlag