ECOLE POLYTECHNIQUE

Measurability of optimal transportation and strong coupling of martingale measures

Joaquin Fontbona, Hélène Guérin, Sylvie Méléard

Measurability of optimal transportation and strong coupling of martingale measures

Joaquin Fontbona,* Hélène Guérin † Sylvie Méléard‡

Abstract

We consider the optimal mass transportation problem in $\mathbb{R}^d$ with measurably parameterized marginals, for general cost functions and under conditions ensuring the existence of a unique optimal transport map. We prove a joint measurability result for this map, with respect to the space variable and to the parameter. The proof needs to establish the measurability of some set-valued mappings, related to the support of the optimal transference plans, which we use to perform a suitable discrete approximation procedure. A motivation is the construction of a strong coupling between orthogonal martingale measures. By this we mean that, given a martingale measure, we construct in the same probability space a second one with specified covariance measure. This is done by pushing forward one martingale measure through a predictable version of the optimal transport map between the covariance measures. This coupling allows us to obtain quantitative estimates in terms of the Wasserstein distance between those covariance measures.

Keywords: Measurability of optimal transport. Coupling between orthogonal martingale measures. Predictable transport process.

Mathematics Subject Classification (2000): 49Q20. 60G57.

1   Introduction

We consider the optimal mass transportation problem in $\mathbb{R}^d$ with measurably parameterized marginals, for general cost functions and under conditions ensuring the existence of a unique optimal transport map. The aim of this note is to prove a joint measurability result for this map, with respect to the space variable and to the parameter. One of our motivations, developed at the end, is the construction of a strong coupling between martingale measures. That is, given a martingale measure, we shall construct in the same probability space a second one with specified covariance measure process. This will be done by pushing forward the given martingale measure through the optimal transport map between the covariance measures. To make this construction rigorous, we need the existence of a *predictable* version of this transport map, which will be a consequence of our main result.

We denote the space of Borel probability measures in $\mathbb{R}^d$ by $\mathcal{P}(\mathbb{R}^d)$, and by $\mathcal{P}_p(\mathbb{R}^d)$ the subspace of probability measures having finite $p-$order moment.

Given $\pi \in \mathcal{P}(\mathbb{R}^{2d})$, we write

$$\pi <_\nu^\mu$$

if $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ are respectively its first and second marginals. Such π is refereed to as a "transference plan" between μ and ν.

Let $c : \mathbb{R}^d \to \mathbb{R}_+$ be a continuous function. The mapping

$$\pi \to I(\pi) := \int_{\mathbb{R}^{2d}} c(x, y) \pi(dx, dy)$$

is then lower semi continuous.

The Monge-Kantorovich or optimal mass transportation problem with cost c and marginals μ, ν consists in finding

$$\inf_{\pi <_\nu^\mu} I(\pi).$$

It is well known that the infimum is attained as soon as it is finite, see [13], Ch.1. In this case, we denote by $\Pi_c^*(\mu, \nu)$ the subset of $\mathcal{P}(\mathbb{R}^{2d})$ of minimizers. If otherwise, $I(\pi) = +\infty$ for all $\pi <_\nu^\mu$, then by convention we set $\Pi_c^*(\mu, \nu) = \emptyset$.

We shall say that *Assumption $H(\mu, \nu, c)$* holds if

a) *μ does not give mass to sets with Hausdorff dimension smaller than or equal to $d-1$.*

b) *there exists a unique optimal transference plan $\pi \in \Pi_c^*(\mu, \nu)$, and it has the form*

$$\pi(dx, dy) = \mu(dx) \otimes \delta_{T(x)}(dy)$$

for a $\mu(dx) - a.s.$ unique mapping $T : \mathbb{R}^d \to \mathbb{R}^d$.

Such T is called an *optimal transport map between μ and ν for the cost function c.*

Hypothesis a) in $H(\mu, \nu, c)$ is optimal both for existence and uniqueness of an optimal transport map, see Remark 9.5 in [14]. We recall that if $\Pi_c^*(\mu, \nu) \neq \emptyset$, a) implies b) in the following situations (see Gangbo and McCann [4]):

i) $c(x, y) = \tilde{c}(|x - y|)$ with $\tilde{c} : \mathbb{R}_+ \to \mathbb{R}_+$ strictly convex, superlinear and differentiable with locally Lipschitz gradient.

ii) $c(x, y) = \tilde{c}(|x - y|)$ with $\tilde{c}$ strictly concave, and μ and ν are mutually singular.

Condition b) also holds if

iii) $c(x, y) = \tilde{c}(|x-y|)$ with $\tilde{c}$ strictly convex and superlinear, and moreover μ is absolutely continuous with respect to Lebesgue measure.

When $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, fundamental examples are the cost function $c(x, y) = |x - y|^p$ with $p \geq 2$ for case i), $p > 1$ for case iii), and $p \in (0, 1)$ for case ii).

Our main result is

Theorem 1.1. *Let (E, Σ, m) be a $\sigma-$finite measurable space and consider a measurable function $\lambda \in E \mapsto (\mu_\lambda, \nu_\lambda) \in \mathcal{P}(\mathbb{R}^d)$ such that for $m-$almost every λ, $H(\mu_\lambda, \nu_\lambda, c)$ holds, with optimal transport map $T_\lambda : \mathbb{R}^d \to \mathbb{R}^d$. Then, there exists a function $(\lambda, x) \mapsto T(\lambda, x)$ which is measurable with respect to $\Sigma \otimes \mathcal{B}(\mathbb{R}^d)$ and such that $m(d\lambda)-$almost everywhere,*

$$T(\lambda, x) = T_\lambda(x) \quad \mu_\lambda(dx)\text{-almost surely.}$$

In particular, $T_\lambda(x)$ is measurable with respect to the completion of $\Sigma \otimes \mathcal{B}(\mathbb{R}^d)$ with respect to $m(d\lambda)\mu_\lambda(dx)$.

Theorem 1.1 generalizes Theorem 1.2 in [3], where we constructed a predictable version of a quadratic transport map, between a time-varying law and empirical samples of it.

To our knowledge, other measurability results on the mass transportation problem require a topological structure on the space of parameters, or concern transference plans but not transport maps (see e.g. [10], or Corollaries 5.22 and 5.23 in [14]).

The proof of Theorem 1.1 is developed in the following section. We firstly establish a type of measurable dependence of the support of the optimizers on λ. From this result, we can define measurable partitions of $E \times \mathbb{R}^d$ induced by a dyadic partition of $\mathbb{R}^d$, and construct bi-measurable discrete approximations of $T(\lambda, x)$. This approximation procedure was not needed in the simpler case studied in [3], where one of the marginals was an empirical measure (thus with finite support).

2 Proof of Theorem 1.1

Let us first state an intermediary result concerning measurability properties of minimizers in the general framework. Its formulation and proof require some notions of set-valued analysis, see e.g. Appendix A of [9].

Theorem 2.1. *The function assigning to (μ, ν) the set of $\mathbb{R}^{2d}$*

$$\Psi(\mu, \nu) := Adh \left(\bigcup_{\pi \in \Pi_c^*(\mu, \nu)} supp(\pi) \right), \tag{1}$$

is measurable in the sense of set-valued mappings. That is, for any open set θ in $\mathbb{R}^{2d}$, its inverse image $\Psi^{-1}(\theta) = \{(\mu, \nu) \in (\mathcal{P}(\mathbb{R}^d))^2 : \Psi(\mu, \nu) \cap \theta \neq \emptyset\}$ is a Borel set in $(\mathcal{P}(\mathbb{R}^d))^2$.

Remark 2.2. *In the case of a set-valued mapping taking closed-set values, measurability is equivalent to the fact that inverse images of closed sets are measurable (see [9]).*

Proof. The idea of the proof is similar to the one of Theorem 1.3 in [3], where we considered the quadratic cost and the measurable structure induced by the Wasserstein topology. In the present case, the spaces $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{R}^{2d})$ are endowed with the usual weak topology. We observe that Ψ writes as the adherence of a set-valued composition,

$$\Psi(\mu, \nu) = Adh(U \circ S(\mu, \nu)) := Adh \left(\bigcup_{\pi \in S(\mu, \nu)} U(\pi) \right),$$

where S and U are the set-valued mappings respectively defined by

$$S(\mu, \nu) := \Pi_c^*(\mu, \nu) \quad \text{and} \quad U(\pi) := supp(\pi).$$

Measurability of Ψ is equivalent to $U \circ S$ being measurable. The latter will be true as soon as S is measurable and $U^{-1}(\theta)$ is open for every open set θ (see [9]).

The stability theorem for optimal transference plans of Schachermayer and Teichman (Theorem 3 in [11]) exactly states that inverse images through S of closed sets in $\mathcal{P}(\mathbb{R}^{2d})$ are closed sets in $(\mathcal{P}(\mathbb{R}^d))^2$. This, together with the fact that the mapping S takes closed-set values (by lower semi continuity of $I(\pi)$) imply that S is a measurable multi-application. On the other hand, the inverse image by U of an open set θ of $\mathbb{R}^{2d}$ is

$$U^{-1}(\theta) = \{\pi \in \mathcal{P}(\mathbb{R}^{2d}) : supp(\pi) \cap \theta \neq \emptyset\} = \{\pi \in \mathcal{P}(\mathbb{R}^{2d}) : \pi(\theta) > 0\}.$$

It then follows by the Portmanteau Theorem that $U^{-1}(\theta)$ is an open set in $\mathcal{P}(\mathbb{R}^{2d})$, and this concludes the proof.

$\square$

Corollary 2.3. *Let (E, Σ) be a measurable space, and $\lambda \in E \mapsto (\mu_\lambda, \nu_\lambda) \in (\mathcal{P}_2(\mathbb{R}^d))^2$ a measurable function. We consider the function Ψ defined by (1) and let F be a closed set of $\mathbb{R}^d$. Then, the set*

$$\{(\lambda, x) : (\{x\} \times F) \cap \Psi(\mu_\lambda, \nu_\lambda) \neq \emptyset\}$$

belongs to $\Sigma \otimes \mathcal{B}(\mathbb{R}^d)$. In particular, if for all $\lambda \in E$, $\Pi_c^*(\mu_\lambda, \nu_\lambda) = \{\pi_\lambda\}$ is a singleton, the set

$$\tilde{F} := \{(\lambda, x) : (\{x\} \times F) \cap supp(\pi_\lambda) \neq \emptyset\}$$

is measurable.

Proof. Without loss of generality, we assume that F is nonempty. Let us first show that for any open set θ of $\mathbb{R}^{2d}$, the set

$$G = \{z \in \mathbb{R}^d : (\{z\} \times F) \cap \theta \neq \emptyset\}$$

is open. Indeed, for $x \in G$ there exists $y \in F$ and $\varepsilon > 0$ such that $B(x, \varepsilon) \times B(y, \varepsilon) \subset \theta$. In particular, for all $z \in B(x, \varepsilon)$ one has $(z, y) \in \theta$ and so $B(x, \varepsilon) \subset G$. By definition of measurability, the set-valued mappings $(\lambda, x) \to \{x\} \times F$ and $(\lambda, x) \to \Psi(\mu_\lambda, \nu_\lambda) - (\{x\} \times F)$ are thus measurable. The latter mapping being also closed valued, we conclude that

$$\left\{(\lambda, x) : [\Psi(\mu_\lambda, \nu_\lambda) - (\{x\} \times F)] \cap \{0\} \neq \emptyset\right\}$$

is a measurable set, which finishes the proof.

$\square$

Let us now focus on the proof of Theorem 1.1

Proof of Theorem 1.1 Since any σ-finite measure is equivalent to a finite one, we can assume without loss of generality that m is finite.

For a fixed $k \geq 1$, we denote by $(A_{n,k})_{n \in \mathbb{Z}^d}$ the partition of $\mathbb{R}^d$ in dyadic half-open rectangles of size 2^{-dk}, that is

$$A_{n,k} := \prod_{i=1}^d \left[\frac{n_i}{2^k}, \frac{n_i + 1}{2^k}\right), \quad \text{where } n = (n_1, \ldots, n_d) \in \mathbb{Z}^d.$$

Consider the sets $B_{n,k} = \{(\lambda, x) \in E \times \mathbb{R}^d : (\{x\} \times A_{n,k}) \cap supp(\pi_\lambda) \neq \emptyset\}$. Notice that since $A_{n,k} = \bigcup_{j \in \mathbb{N}} \prod_{i=1}^{d} \left[\frac{n_i}{2^k}, \frac{n_i+1}{2^k} - \frac{1}{2^{k+j}}\right]$, one has

$$B_{n,k} = \bigcup_{j \in \mathbb{N}} \left\{(\lambda, x) : \left(\{x\} \times \prod_{i=1}^{d} \left[\frac{n_i}{2^k}, \frac{n_i+1}{2^k} - \frac{1}{2^{k+j}}\right]\right) \cap supp(\pi_\lambda) \neq \emptyset\right\},$$

and so $B_{n,k}$ is measurable thanks to Corollary 2.3.

Denote now by $a_{n,k} \in A_{n,k}$ the "center" of the set, and define a $\Sigma \otimes \mathcal{B}(\mathbb{R}^d)$−measurable function by

$$T^k(\lambda, x) = \sum_{n \in \mathbb{Z}^d} a_{n,k} \mathbf{1}_{B_{n,k}}(\lambda, x). \tag{2}$$

For each $\lambda \in E$, let ν_λ^k be the discrete measure defined by pushing forward μ_λ through T^k, that is,

$$\nu_\lambda^k(A) = \int \mathbf{1}_{T^k(\lambda, x) \in A} \, \mu_\lambda(dx), \ A \in \mathcal{B}(\mathbb{R}^d).$$

Denote also by $\tilde{E} \in \Sigma$ a measurable set with $m(\tilde{E}^c) = 0$ and such that for all $\tilde{\lambda} \in \tilde{E}$, $H(\mu_\lambda, \nu_\lambda, c)$ holds.

By hypothesis, for each $\lambda \in \tilde{E}$ we have that

$$\mu_\lambda(dx) \text{ almost surely: } \mathbf{1}_{B_{n,k}}(\lambda, x) = \mathbf{1}_{\{x : T_\lambda(x) \in A_{n,k}\}}. \tag{3}$$

where T_λ has been defined in the statement of Theorem 1.1. This implies that

$$\nu_\lambda^k(\{a_{n,k}\}) = \int \mathbf{1}_{B_{n,k}}(\lambda, x) \mu_\lambda(dx) = \mu_\lambda(\{x : T_\lambda(x) \in A_{n,k}\}) = \nu_\lambda(A_{n,k})$$

by definition of T_λ.

We now check that $(T^k)_{k \in \mathbb{N}}$ is a cauchy sequence in $L^1(E \times \mathbb{R}^d, m(d\lambda)\mu_\lambda(dx))$. Fix $k \leq k'$, and for each $n \in \mathbb{Z}^d$ denote by $\{A_{n',k'}\}_{n'}$ the unique partition of $A_{n,k}$ in dyadic rectangles of size $2^{-dk'}$. We then have that

$$\int_E \int_{\mathbb{R}^d} |T^k(\lambda, x) - T^{k'}(\lambda, x)| \mu_\lambda(dx) m(d\lambda)$$

$$= \int_E \int_{\mathbb{R}^d} \sum_{n \in \mathbb{Z}^d} \sum_{n' : A_{n',k'} \subset A_{n,k}} \mathbf{1}_{B_{n',k'}}(\lambda, x) |a_{n,k} - a_{n',k'}| \mu_\lambda(dx) m(d\lambda)$$

$$= \int_E \sum_{n \in \mathbb{Z}^d} \sum_{n' : A_{n',k'} \subset A_{n,k}} |a_{n,k} - a_{n',k'}| \nu_\lambda(A_{n',k'}) m(d\lambda)$$

$$\leq \int_E \sum_{n \in \mathbb{Z}^d} 2^{-k} \sum_{n' : A_{n',k'} \subset A_{n,k}} \nu_\lambda(A_{n',k'}) m(d\lambda)$$

$$\leq \int_E \sum_{n \in \mathbb{Z}^d} 2^{-k} \nu_\lambda(A_{n,k}) m(d\lambda)$$

$$\leq 2^{-k} \int_E \nu_\lambda(\mathbb{R}^d) m(d\lambda) = 2^{-k} m(E),$$

and the Cauchy property follows since $m(E) < \infty$.

Let us denote by T the limit in $L^1(E \times \mathbb{R}^d, m(d\lambda)\mu_\lambda(dx))$ of the sequence T^k. Theorem 1.1 will be proved by verifying that for all λ in a set of Σ of full m-measure set, one has $\pi_\lambda(dx, dy) = \mu_\lambda(dx)\delta_{T(\lambda,x)}(dy)$. Hence, it is enough to check that

$$\int \mathbb{I}_{C \times A_{n,k}}(x, T(\lambda, x))\mu_\lambda(dx) = \pi_\lambda(C \times A_{n,k})$$

for any semi-open rectangle C with dyadic extremes and all $n \in \mathbb{Z}^d, k \in \mathbb{N}$. We have for $\lambda \in \tilde{E}$ and any $j \in \mathbb{N}$ that

$$\left| \pi_\lambda(C \times A_{n,k}) - \int \mathbf{1}_{C \times A_{n,k}}(x, T(\lambda, x))\mu_\lambda(dx) \right|$$

$$\leq \left| \pi_\lambda(C \times A_{n,k}) - \int \mathbf{1}_{C \times A_{n,k}}(x, T^j(\lambda, x))\mu_\lambda(dx) \right| \tag{4}$$

$$+ \int \left| \mathbf{1}_{C \times A_{n,k}}(x, T^j(\lambda, x)) - \mathbf{1}_{C \times A_{n,k}}(x, T(\lambda, x)) \right| \mu_\lambda(dx)$$

$$:= \Delta_j + \Delta_j'.$$

We approximate $\mathbf{1}_{A_{n,k}}$ by a Lipschitz continuous function $f_{\lambda,\varepsilon}$ such that $\|f_{\lambda,\varepsilon}\|_\infty \leq 1$ and $\mu_\lambda(\{y : f_{\lambda,\varepsilon}(y) \neq \mathbf{1}_{A_{n,k}}\}) \leq \varepsilon$ (this is possible thanks to $H(\mu_\lambda, \nu_\lambda, c)$, a)). Hence, the second term Δ_j' on the r.h.s. of (4) is bounded by

$$4\varepsilon + L_{\lambda,\varepsilon} \int \left| T^j(\lambda, x) - T(\lambda, x) \right| \mu_\lambda(dx),$$

where $L_{\lambda,\varepsilon}$ is the Lipschitz constant of $f_{\lambda,\varepsilon}$. Since $\int \left| T^j(\lambda, x) - T(\lambda, x) \right| \mu_\lambda(dx)$ converges in $L^1(m(d\lambda))$ to 0, there is a subsequence T^{j_i} and a set $\hat{E} \in \Sigma$ of full measure such that the convergence holds for all $\lambda \in \hat{E}$. Consequently, for all $\lambda \in \bar{E} := \tilde{E} \cap \hat{E}$ we get that

$$\limsup_{i \to \infty} \Delta_{j_i}' = \limsup_{i \to \infty} \int \left| \mathbf{1}_{C \times A_{n,k}}(x, T^{j_i}(\lambda, x)) - \mathbf{1}_{C \times A_{n,k}}(x, T(\lambda, x)) \right| \mu_\lambda(dx) \leq 2\varepsilon,$$

and since the l.h.s. does not depend on ε, this means that $\lim_{i \to \infty} \Delta_{j_i}' = 0$.

The proof will be achieved be verifying that for fixed $\lambda \in \bar{E}$, one has $\Delta_j = 0$ for all large enough j. For such λ, fix a Borel set D_λ of $\mathbb{R}^d$ of full μ_λ measure where (3) is everywhere true. Then,

$$\int \mathbf{1}_{C \times A_{n,k}}(x, T^j(\lambda, x))\mu_\lambda(dx) = \int \mathbf{1}_{(D_\lambda \cap C) \times A_{n,k}}(x, T^j(\lambda, x))\mu_\lambda(dx)$$

$$= \int \mathbf{1}_{(D_\lambda \cap C) \times A_{n,k}}\left(x, \sum_{m \in \mathbb{Z}^d} a_{m,j}\mathbf{1}_{B_{m,j}}(\lambda, x) \right) \mu_\lambda(dx)$$

$$= \int \mathbf{1}_{(D_\lambda \cap C) \times A_{n,k}}\left(x, \sum_{m: a_{m,j} \in A_{n,k}} a_{m,j}\mathbf{1}_{A_{m,j}}(T_\lambda(x)) \right) \mu_\lambda(dx).$$

Remark now that for all $j \geq k$, $y \in A_{n,k} \iff \sum_{m: a_{m,j} \in A_{n,k}} a_{m,j}\mathbf{1}_{A_{m,j}}(y) \in A_{n,k}$. Then, for all $j \geq k$,

$$\int \mathbf{1}_{C \times A_{n,k}}(x, T^j(\lambda, x))\mu_\lambda(dx) = \int \mathbf{1}_{(D_\lambda \cap C) \times A_{n,k}}(x, T_\lambda(x))\mu_\lambda(dx) = \pi_\lambda(C \times A_{n,k}).$$

□

6

3 Application: strong coupling for orthogonal martingale measures

We now develop an application of Theorem (1.1). Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space and consider M an adapted orthogonal martingale measure on $\mathbb{R}_+ \times \mathbb{R}^d$ (in the sense of Walsh [15]). Assume that its covariance measure has the form $q_t(da)dk_t$, where $q_t(\omega, da)$ is a predictable random probability measure on $\mathbb{R}^d$ with finite second moment and k_t a predictable increasing process. Let us also consider another predictable random probability measure $\hat{q}_t(\omega, da)$ on $\mathbb{R}^d$ with finite second moment.

We want to construct in the same probability space a second martingale measure with covariance measure $\hat{q}_t(\omega, da)dk_t$, in such a way that in some sense, the distance between the martingale measures is controlled by the Wasserstein distance between their covariance measures. Recall that this distance is defined for $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ by

$$W_2^2(\mu, \nu) = \inf_{\pi <_\nu^\mu} I(\pi)$$

with the quadratic cost

$$c(x, y) = |x - y|^2.$$

This distance makes the set $\mathcal{P}_2(\mathbb{R}^d)$ a Polish space, and strengthens the weak topology with the convergence of second moments (see [8]).

Theorem 3.1. *In the previous setting, assume moreover that $\mathbb{P}(d\omega)dk_t(\omega)$ a.e. q_t has a density with respect to Lebesgue measure in $\mathbb{R}^d$.*

Then, there exists in $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ a martingale measure $\hat{M}$ on $\mathbb{R}_+ \times \mathbb{R}^d$ with covariance measure $\hat{q}_t(da)dk_t$, such that for all $S > 0$ and for every predictable function $\phi : \Omega \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ that is Lipschitz continuous in the last variable with $\mathbb{E}\left(\int_0^S \int \phi^2(s, a)\left(q_s(da) + \hat{q}_s(da)\right) dk_s\right) < \infty$, one has

$$\mathbb{E}\left(\sup_{t \leq S}\left(\int_0^t \int \phi(s, a)M(ds, da) - \int_0^t \int \phi(s, a)\hat{M}(ds, da)\right)^2\right) \leq \mathbb{E}\left(\int_0^S L_s^2 \, W_2^2(q_s, \hat{q}_s) \, dk_s\right), \tag{5}$$

where $L_s(\omega)$ is a measurable version of a Lipschitz constant of $\phi(s, \omega, \cdot)$ and W_2^2 is the quadratic Wasserstein distance in $\mathcal{P}_2(\mathbb{R}^d)$.

Proof. Since $q_t(\omega, da)$ has a density for almost every (t, ω), assumption $H(q_t(\omega, da), \hat{q}_t(\omega, da), c)$ is satisfied. We can therefore apply Theorem 1.1 to $(E, \Sigma, m) = (\Omega \times \mathbb{R}_+, \mathcal{P}red, \mathbb{P}(d\omega)dk_t(\omega))$, where $\mathcal{P}red$ is the predictable σ-field with respect to $\mathcal{F}_t$. Then, there exists a predictable mapping $T : \Omega \times \mathbb{R}_+ \times \mathbb{R}^d :\to \mathbb{R}^d$ that for m-almost every (t, ω) pushes forward q_t to $\hat{q}_t$. Moreover, for a.e. (t, ω), one has

$$\int |a - T(t, \omega, a)|^2 q_t(\omega, da) = W_2^2(q_t, \hat{q}_t).$$

On can thus define a martingale measure $\hat{M}$ by the stochastic integrals

$$\int_0^t \int \psi(s, a)\hat{M}(ds, da) := \int_0^t \int \psi(s, T(s, a))M(ds, da)$$

7

for predictable simple functions ψ. Its covariance measure is by construction $\hat{q}_t(da)dk_t$, and by Doob's inequality, the left hand side of (5) is less than

$$\mathbb{E}\Big(\int_0^S \int |\phi(s,a) - \phi(s,T(s,a))|^2 q_s(da)dk_s \Big) \leq \mathbb{E}\left(\int_0^S L_s^2 \left(\int |a - T(s,a)|^2 q_s(da) \right) dk_s \right)$$

$$\leq \mathbb{E}\left(\int_0^S L_s^2 \, W_2^2(q_s, \hat{q}_s) \, dk_s \right), \quad \text{by definition of } T \text{ and of } W_2^2.$$

$\square$

Remark 3.2. *i) The construction of strong couplings between orthogonal martingale measures arises classically in the literature, especially in cases where the martingale measure M is a compensated Poisson point measure or a space-time white noise, for which the covariance measures are deterministic (cf. Grigelionis [5], El Karoui-Lepeltier [1], Tanaka [12], El Karoui-Méléard [2], Méléard-Roelly [7], Guérin [6]). A classical approach is to use the Skorokhod representation theorem. This however prevents any hope to obtain quantitative estimates related to the associated covariance measures, what we have been able to do here thanks to the optimal transport maps.*

ii) If the probability space and the martingale measure M are not fixed in advance, a coupling satisfying the estimate (5) can be constructed from and orthogonal martingale measure $\tilde{M}(dt, da, da')$ on $\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d$ with covariance measure $\pi_t(da, da')dk_t$, where π_t is an optimal transference plan between q_t and $\hat{q}_t$. Then, $\tilde{M}(dt, da, \mathbb{R}^d)$ and $\tilde{M}(dt, \mathbb{R}^d, da')$ are indeed two orthogonal martingale measures with the required covariances and satisfying estimate (3.1). The question in this situation is however how to construct such $\tilde{M}$.

References

[1] EL KAROUI, N.; LEPELTIER, J.P. Représentation des processus ponctuels multivariés à l'aide d'un processus de Poisson, *Z. Wahrsch. Verw. Geb.* **39**, (1977), 111–133.

[2] EL KAROUI, N.; MÉLÉARD, S. Martingale measures ans stochastic calculus, *Probab. Th. Rel. Fields* **84**, (1990), 83–101.

[3] FONTBONA, J.; GUÉRIN, H.; MÉLÉARD, S. Mesurability of optimal transportation and convergence rate for Landau type interacting particle systems. *to appear in PTRF 2008*

[4] GANGBO, W.; MCCANN, R.J. The geometry of optimal transportation, *Acta Math.* **177**, (1996), 113–161.

[5] GRIGELIONIS, B. On the representation of integer valued measures by means of stochastic integrals with respect to Poisson measures, *Litov. Mat. Sb.* **11**, (1971), 93–108.

[6] GUÉRIN, H. Existence and regularity of a weak function-solution for some Landau equations with a stochastic approach, *Stochastic Process. Appl.* **101**, (2002), no. 2, 303–325.

[7] MÉLÉARD, S.; ROELLY, S. Discontinuous measure-valued branching processes and generalized stochastic equations, *Math. Nachr.* **154**, (1991), 141–156.

[8] RACHEV, S.T.; RUSCHENDORF, L. Mass Transportation Problems, Volumes I and II, Springer (1998).

[9] ROCKAFELLER, R.T.; WETS, R. J-B.: Variational Analysis, Springer (1998).

[10] RÜSCHENDORF, L. The Wasserstein distance and approximation theorems. *Z. Wahrsch. Verw. Gebiete* **70** (1985), 117–129.

[11] SCHACHERMAYER, W.; TEICHMANN, J.: Characterization of optimal transport plans for the Monge-Kantorovich-problem. *to appear in Proc. A.M.S.*

[12] TANAKA, H. Probabilistic treatment of the Boltzmann equation of Maxwellian molecules, *Z. Wahrsch. Verw. Geb.* **46**, (1978), 67–105.

[13] VILLANI, C. Topics in Optimal Transportation, Graduate Studies in Mathematics Vol. 58, AMS (2003). **8** (1998), 957-983.

[14] VILLANI, C. Optimal transport, old and new. To appear in Grundlehren der mathematischen Wissenschaften, Springer.

[15] WALSH, J.B. An introduction to stochastic partial differential equations, *École d'été de Probabilités de Saint-Flour XIV, Lect. Notes in Math.* **1180** (1984), 265-437.