Elect. Comm. in Probab. **15** (2010), 124–133

ELECTRONIC COMMUNICATIONS in PROBABILITY

MEASURABILITY OF OPTIMAL TRANSPORTATION AND STRONG COUPLING OF MARTINGALE MEASURES

JOAQUIN FONTBONA[1]
DIM-CMM, UMI(2807) UCHILE-CNRS, Universidad de Chile, Casilla 170-3, Correo 3, Santiago-Chile.
email: `fontbona@dim.uchile.cl`

HÉLÈNE GUÉRIN [2]
IRMAR, Université Rennes 1, Campus de Beaulieu, 35042 Rennes-France.
email: `helene.guerin@univ-rennes1.fr`

SYLVIE MÉLÉARD[3]
CMAP, Ecole Polytechnique, CNRS, route de Saclay, 91128 Palaiseau Cedex-France.
email: `sylvie.meleard@polytechnique.edu`

Submitted August 5, 2008, *accepted in final form* March 22, 2010

AMS 2000 Subject classification: 49Q20;60G57
Keywords: Measurability of optimal transport. Coupling between orthogonal martingale measures. Predictable transport process.

Abstract
We consider the optimal mass transportation problem in $\mathbb{R}^d$ with measurably parameterized marginals 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 a specified covariance measure process. This is done by pushing forward the first 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.

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

[1] SUPPORTED BY FONDECYT PROYECT 1070743, ECOS-CONICYT C05E02 AND BASAL-CONICYT
[2] SUPPORTED BY ECOS-CONICYT C05E02
[3] SUPPORTED BY ECOS-CONICYT C05E02

to the space variable and the parameter. One of our motivations is the construction of couplings between martingale measures (in the sense of Walsh [15]). This problem classically arises in the literature, especially in cases where the martingale measures are compensated Poisson point measures or space-time white noises, 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 for constructing such couplings is to use the Skorokhod representation theorem. However, this does not give any quantitative estimate.
Here, we aim at a method to construct what we call "strong coupling" between martingale measures. By this we mean that, given a martingale measure, we search to construct in the same probability space a second one with specified covariance measure process. We will do this by pushing forward the given martingale measure through the optimal transport map between the two covariance processes. The interest of such a construction is that it will provide quantitative estimates in terms of the Wasserstein distance between the covariance measures. But in order to make this construction rigorous, the existence of a *predictable* version of the above described transport map is needed. To our knowledge, available measurability results on the mass transportation problem with respect to some parameter require a topological structure on the space of parameters, or concern measurability of *transference plans*, but not of *transport maps* (see e.g. [10], or Corollaries 5.22 and 5.23 in [14]). Our main result provides the existence of the jointly-measurable optimal transport map which is required.

Let us establish notations and recall basic facts. We denote the space of Borel probability measures in $\mathbb{R}^d$ by $\mathscr{P}(\mathbb{R}^d)$ endowed with the usual weak topology, and by $\mathscr{P}_p(\mathbb{R}^d)$ the subspace of probability measures having finite $p-$order moment. Given $\pi \in \mathscr{P}(\mathbb{R}^{2d})$, we write

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

if $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ are respectively its first and second marginals. Such π is referred to as a "transference plan" between μ and ν.
Let $c : \mathbb{R}^{2d} \to \mathbb{R}_+$ 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

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

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 $\mathscr{P}(\mathbb{R}^{2d})$ of minimizers. If otherwise, $I(\pi) = +\infty$ for all $\pi <^{\mu}_{\nu}$, then by convention we set $\Pi^*_c(\mu,\nu) = \emptyset$.

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

We recall that the condition that μ does not give mass to sets with Hausdorff dimension smaller than or equal to $d-1$ is optimal both for existence and uniqueness of T, see Remark 9.5 in [14]. Moreover, if $\Pi_c^*(\mu,\nu) \neq \emptyset$, then the latter condition implies $H(\mu,\nu,c)$ 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 $H(\mu,\nu,c)$ 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 $\mu,\nu \in \mathscr{P}_p(\mathbb{R}^d)$, 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 measure space and consider a measurable function $\lambda \in E \mapsto (\mu_\lambda,\nu_\lambda) \in (\mathscr{P}(\mathbb{R}^d))^2$ 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 $\Sigma \otimes \mathscr{B}(\mathbb{R}^d)$ 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 $\Sigma \otimes \mathscr{B}(\mathbb{R}^d)$ with respect to $m(d\lambda)\mu_\lambda(dx)$.

Theorem 1.1 generalizes Theorem 1.2 in [3], where a predictable version of the quadratic transport map between a time-varying law and empirical samples of it was constructed. This allowed us to exhibit the convergence rate of a Brownian motion driven interacting particle system towards a white-noise driven nonlinear process. More precisely, we considered n independent copies $(X^i)_{i=1}^n$ of the process in $\mathbb{R}^d$:

$$X_t^i = X_0^i + \int_0^t \int_{\mathbb{R}^d} \sigma(X_s^i - y) W^i(ds,dy) + \int_0^t \int_{\mathbb{R}^d} b(X_s^i - y) P_s(y) dy ds \tag{1}$$

where σ and b satisfy usual Lipschitz assumptions, $X_t^i \sim P_t(y)dy$ and W^i is an $\mathbb{R}^d$ valued space-time white noise on $[0,T] \times \mathbb{R}^d$ with independent coordinates of covariance measures $P_t(y)dy \otimes dt$. By pushing forward each $W^i(dt,dx)$ through the predictable version of the optimal transport map from P_t to $\frac{1}{n}\sum_{j=1}^n \delta_{X_t^j}$, we constructed in the same probability space n^2 independent $\mathbb{R}^d-$Brownian motions (B^{ik}), suitably correlated with the martingale measures (W^i). This provided us a coupling between the processes (X_t^i) and the particles $(X_t^{i,n})$ governed by

$$X_t^{i,n} = X_0^i + \frac{1}{\sqrt{n}} \int_0^t \sum_{k=1}^n \sigma(X_s^{i,n} - X_s^{k,n}) dB_s^{ik} + \frac{1}{n} \int_0^t \sum_{k=1}^n b(X_s^{i,n} - X_s^{k,n}) ds, \ i = 1,\ldots,n,$$

yielding the estimate $\mathscr{W}_2^2(law(X^{1,n}), law(X^1)) \leq C_{T,d} n^{\frac{-2}{d+4}}$ for the Wasserstein-2 distance $\mathscr{W}_2$ on the space of probability measures on $C([0,T],\mathbb{R}^d)$. That is indeed the convergence rate to 0 of the (squared) expected Wasserstein-2 distance $W_2^2(\frac{1}{n}\sum_{j=1}^n \delta_{X_t^j}, P_t)$ in $\mathscr{P}_2(\mathbb{R}^d)$, as n goes to infinity. (For details and precise statements, see Theorem 1.1 and Corollary 6.2. in [3])

The proof of Theorem 1.1 is developed in Section **2.**. We firstly establish a type of measurable dependence on λ of the support of the optimizers. From this result, we can define measurable (w.r.t. λ) 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 (the empirical measure) had finite support. In Section **3** we address the construction of strong couplings between orthogonal martingale measures.

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. chapter 14 of [9].

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

$$\Psi(\mu,\nu) := Cl\left(\bigcup_{\pi\in\Pi_c^*(\mu,\nu)} supp(\pi)\right), \tag{2}$$

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

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 topology in $\mathscr{P}(\mathbb{R}^d)$ and $\mathscr{P}(\mathbb{R}^{2d})$ is the usual weak one.)
We observe that Ψ writes as the topological closure of a set-valued composition,

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

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

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

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 Teichmann (Theorem 3 in [11]) exactly states that inverse images through S of closed sets in $\mathscr{P}(\mathbb{R}^{2d})$ are closed sets in $(\mathscr{P}(\mathbb{R}^d))^2$. 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 set valued mapping.
On the other hand, the inverse image by U of an open set θ of $\mathbb{R}^{2d}$ is

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

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

$\square$

Corollary 2.3. *Let (E,Σ) be a measurable space, and $\lambda \in E \mapsto (\mu_\lambda,\nu_\lambda) \in (\mathscr{P}(\mathbb{R}^d))^2$ a measurable function. We consider the function Ψ defined by (2) 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 $\Sigma\otimes\mathscr{B}(\mathbb{R}^d)$. 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 (see [9], Proposition 14.11(c)). The latter mapping being also closed valued, we conclude that

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

is a measurable set, which finishes the proof.

$\square$

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

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

Consider the sets $B_{n,k}=\{(\lambda,x)\in E\times\mathbb{R}^d : (\{x\}\times A_{n,k})\cap\Psi(\mu_\lambda,\nu_\lambda)\neq\emptyset\}$, with Ψ defined by (2). 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

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

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 $\Sigma\otimes\mathscr{B}(\mathbb{R}^d)-$measurable function by

$$T^k(\lambda,x)=\sum_{n\in\mathbb{Z}^d}a_{n,k}\mathbf{1}_{B_{n,k}}(\lambda,x). \tag{3}$$

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

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

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

$$\mu_\lambda(dx) \text{ almost surely: } \mathbf{1}_{B_{n,k}}(\lambda, x) = \mathbf{1}_{\{x:T_\lambda(x)\in A_{n,k}\}}. \tag{4}$$

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 $k \le k'$, 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

$$\begin{aligned}
&\int_E\int_{\mathbb{R}^d} |T^k(\lambda, x) - T^{k'}(\lambda, x)|\mu_\lambda(dx)m(d\lambda)\\
&\qquad = \int_E\int_{\mathbb{R}^d} \sum_{n\in\mathbb{Z}^d}\ \sum_{n':A_{n',k'}\subset A_{n,k}} \mathbf{1}_{B_{n',k'}}(\lambda, x)|a_{n,k} - a_{n',k'}|\mu_\lambda(dx)m(d\lambda)\\
&\qquad = \int_E \sum_{n\in\mathbb{Z}^d}\ \sum_{n':A_{n',k'}\subset A_{n,k}} |a_{n,k} - a_{n',k'}|\nu_\lambda(A_{n',k'})m(d\lambda)\\
&\qquad \le \int_E \sum_{n\in\mathbb{Z}^d} 2^{-k} \sum_{n':A_{n',k'}\subset A_{n,k}} \nu_\lambda(A_{n',k'})m(d\lambda)\\
&\qquad = \int_E \sum_{n\in\mathbb{Z}^d} 2^{-k}\nu_\lambda(A_{n,k})m(d\lambda)\\
&\qquad = 2^{-k}\int_E \nu_\lambda(\mathbb{R}^d)m(d\lambda) = 2^{-k}m(E),
\end{aligned}$$

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 $\pi_\lambda(dx, dy) = \mu_\lambda(dx)\delta_{T(\lambda,x)}(dy)$. To that end, it is enough to check that

$$\int g(x)f(T(\lambda, x))\mu_\lambda(dx) = \int g(x)f(y)\pi_\lambda(dx, dy)$$

for any pair $f, g : \mathbb{R}^d \to \mathbb{R}$ of bounded Lipschitz-continuous functions. We denote by $\|f\| := \sup_{x\neq y} \frac{|f(x)-f(y)|}{|x-y|} + \sup_x |f(x)|$ and $\|g\|$ their corresponding norms.

We have for $\lambda \in \tilde{E}$ that

$$\begin{aligned}\Big|\iint g(x)f(y)\pi_\lambda(dx,dy) - &\int g(x)f(T(\lambda,x))\mu_\lambda(dx)\Big| \\ &\leq \Big|\iint g(x)f(y)\pi_\lambda(dx,dy) - \int g(x)f(T^j(\lambda,x))\mu_\lambda(dx)\Big| \\ &\quad + \|g\|\|f\| \int \Big|T^j(\lambda,x) - T(\lambda,x)\Big|\mu_\lambda(dx) \\ &:= \Delta_j + \Delta'_j.\end{aligned} \tag{5}$$

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 set $\hat{E} \in \Sigma$ of full measure and a subsequence T^{j_i} such that $\lim_{i\to\infty} \Delta'_{j_i} = 0$ for all $\lambda \in \bar{E} := \tilde{E} \cap \hat{E}$.
It is therefore enough to show that for all $\lambda \in \bar{E}$, one has $\Delta_j \to 0$ as j goes to ∞. First, we claim that for $\lambda \in \bar{E}$, one has for any Borel set $C \subseteq \mathbb{R}^d$ and all $n \in \mathbb{Z}^d, k \in \mathbb{N}$ that

$$\int \mathbf{1}_{C\times A_{n,k}}(x, T^j(\lambda,x))\mu_\lambda(dx) = \pi_\lambda\left(C \times A_{n,k}\right) \qquad \text{for all } j \geq k. \tag{6}$$

Fix a Borel set D_λ of $\mathbb{R}^d$ of full μ_λ measure (which might depend on C, k and n) where (4) is everywhere true. Then,

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

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

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

establishing (6). Next, we introduce for each $k \in \mathbb{N}$ an approximation of $h : \mathbb{R}^d \to \mathbb{R}$ a bounded Lipschitz-continuous function by step functions

$$h^{(k)}(x) := \sum_{n\in\mathbb{Z}^d} h(a_{n,k})\mathbf{1}_{A_{n,k}}(x), \quad x \in \mathbb{R}^d.$$

One has $|h(x) - h^{(k)}(x)| \leq \|h\|d2^{-k}$ for all $x \in \mathbb{R}^d$, from which it follows that

$$\begin{aligned}\Delta_j \leq d2^{-k}(4\|f\|\|g\|) + \Big|\iint g^{(k)}(x)f^{(k)}(T^j(\lambda,x))\mu_\lambda(dx) - \int g^{(k)}(x)f^{(k)}(y)\pi_\lambda(dx,dy)\Big| \leq \\ d2^{-k}(4\|f\|\|g\|) + \sum_{m,n\in\mathbb{Z}^d} |g(a_{n,k})||f(a_{m,k})| \Big|\int \mathbf{1}_{A_{m,k}\times A_{n,k}}(x, T^j(\lambda,x))\mu_\lambda(dx) - \pi_\lambda\left(A_{m,k}\times A_{n,k}\right)\Big|\end{aligned}$$

for all $k \in \mathbb{N}$. Thanks to (6), the sum identically vanishes for all $j \geq k$. This means that $\Delta_j \to 0$ as j goes to ∞, and the proof is complete. $\square$

3 Application: strong coupling for orthogonal martingale measures

We now develop an application of Theorem (1.1), which is a generalization of what has been used in [3] to estimate the convergence rate of Landau type interacting particle systems. Let $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathbb{P})$ be a filtered 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. This distance is defined for $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ by $W_2^2(\mu, \nu) = \inf_{\pi <^{\mu}_{\nu}} I(\pi)$ with the quadratic cost $c(x, y) = |x - y|^2$. It makes the set $\mathscr{P}_2(\mathbb{R}^d)$ a Polish space, and strengthens the weak topology with the convergence of second moments (see e.g. [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 $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathbb{P})$ 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 $\phi : \mathbb{R}_+ \times \Omega \times \mathbb{R}^d \to \mathbb{R}$ 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

$$\mathbb{E}\left(\sup_{t \leq S}\left(\int_0^t \int \phi(s, \cdot, a) M(ds, da) - \int_0^t \int \phi(s, \cdot, a)\hat{M}(ds, da)\right)^2\right) \leq \mathbb{E}\left(\int_0^S (K_s^{\phi})^2\, W_2^2(q_s, \hat{q}_s)\, dk_s\right), \tag{7}$$

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

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 $(E, \Sigma, m) = (\Omega \times \mathbb{R}_+, \mathscr{P}red, \mathbb{P}(d\omega)dk_t(\omega))$, where $\mathscr{P}red$ is the predictable $\sigma-$field with respect to $\mathscr{F}_t$. Then, there exists a predictable mapping $T : \mathbb{R}_+ \times \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ that for m-almost every (t, ω) pushes forward q_t to $\hat{q}_t$. Moreover, for a.e. (t, ω), one has

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

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

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

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 (7) is less than

$$\mathbb{E}\Big(\int_0^S \int |\phi(s,a)-\phi(s,T(s,a))|^2 q_s(da)dk_s\Big) \leq \mathbb{E}\left(\int_0^S (K_s^\phi)^2 \left(\int |a-T(s,a)|^2 q_s(da)\right) dk_s\right)$$
$$\leq \mathbb{E}\left(\int_0^S (K_s^\phi)^2 \, W_2^2(q_s,\hat{q}_s)\, dk_s\right),$$

by definition of T and of W_2^2. ∎

Remark 3.2. *In general, a coupling satisfying the estimate* (7) *can be obtained by constructing an orthogonal martingale measure $\tilde{M}(dt,da,da')$ on $\mathbb{R}^+\times\mathbb{R}^d\times\mathbb{R}^d$ with covariance measure $\pi_t(da,da')dk_t$, π_t being an optimal transference plan between q_t and $\hat{q}_t$. Then, the marginal martingale measures $\tilde{M}(dt,da,\mathbb{R}^d)$ and $\tilde{M}(dt,\mathbb{R}^d,da')$ are orthogonal martingale measures with the required covariances. Our argument provides a simple way of constructing $\tilde{M}(dt,da,da')$ when the realization of one of the "marginal" martingale measures is given in advance.*

An immediate consequence is

Corollary 3.3. *Let the space $(\Omega,\mathscr{F},\mathscr{F}_t,\mathbb{P})$, the martingale measure M and the random measure $q_t(da)dk_t$ be as in Theorem 3.1. Assume that $(q_t^n(\omega,da))_{n\in\mathbb{N}}$ is a sequence of predictable random probability measures on $\mathbb{R}^d$ in the same probability space, such that for all $t>0$, $\int_0^t W_2^2(q_s,q_s^n)\,dk_s$ goes to 0 in $L^1(\mathbb{P})$ when $n\to\infty$.*
Then, there exists a sequence (M_n) of orthogonal martingale measures in $(\Omega,\mathscr{F},\mathscr{F}_t,\mathbb{P})$ with covariance measures $(q_t^n(da))$ such that for all $t>0$ and all ϕ with $\|sup_{s\leq t}K_s^\phi\|_{L^\infty(\mathbb{P})}<\infty$, $\int_0^t \phi(s,a)M^n(ds,da)$ converges to $\int_0^t \phi(s,a)M(ds,da)$ in $L^2(\mathbb{P})$.

Acknowledgements We thank the anonymous referees for pointing out a gap in the proof of Theorem 1.1 and for valuable comments that allowed us to improve the presentation of this work.

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. MR0448546

[2] EL KAROUI N., MÉLÉARD S. : Martingale measures and stochastic calculus, *Probab. Th. Rel. Fields* 84 (1990), 83–101. MR1027822

[3] FONTBONA J., GUÉRIN H., MÉLÉARD S. : Mesurability of optimal transportation and convergence rate for Landau type interacting particle systems. *Probab. Th. Rel. Fields* 143 (2009), 329–351. MR2475665

[4] GANGBO W., MCCANN R.J. : The geometry of optimal transportation. *Acta Math.* 177 (1996), 113–161. MR1440931

[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. MR0293703

[6] GUÉRIN, H. : Existence and regularity of a weak function-solution for some Landau equations with a stochastic approach. *Stochastic Process. Appl.* 101 (2002), 303–325. MR1931271

[7] MÉLÉARD S., ROELLY S. : Discontinuous measure-valued branching processes and generalized stochastic equations. *Math. Nachr.* 154 (1991), 141–156. MR1138376

[8] RACHEV S.T., RUSCHENDORF L. : Mass Transportation Problems, Vol. I . *Probability and its applications* Springer (1998). MR1619171

[9] ROCKAFELLAR R.T., WETS R. J-B.: Variational Analysis. *Grundlehren der mathematischen Wissenschaften* 317 Springer (1998). MR1491362

[10] RÜSCHENDORF L. : The Wasserstein distance and approximation theorems. *Z. Wahrsch. Verw. Gebiete* 70 (1985), 117–129. MR0795791

[11] SCHACHERMAYER W., TEICHMANN, J. : Characterization of optimal transport plans for the Monge-Kantorovich-problem. *Proc. Amer. Math. Soc.* 137 (2009), 519–529. MR2448572

[12] TANAKA, H.: Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrsch. Verw. Geb.* 46 (1978), 67–105. MR0512334

[13] VILLANI, C.: Topics in Optimal Transportation. *Graduate Studies in Mathematics* Vol. 58, AMS (2003). MR1964483

[14] VILLANI, C.: Optimal transport, old and new. *Grundlehren der mathematischen Wissenschaften* 338 Springer (2009). MR2459454

[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. MR0876085