

HOKKAIDO UNIVERSITY

Title	Optimal Transportation Problem by Stochastic Optimal Control
Author(s)	Mikami, Toshio; Thieullen, Michele
Citation	Hokkaido University Preprint Series in Mathematics, 690, 1-17
Issue Date	2005-02-03
DOI	https://doi.org/10.14943/83841
Doc URL	https://hdl.handle.net/2115/69495
Type	departmental bulletin paper
File Information	re690.pdf



Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Optimal Transportation Problem by Stochastic Optimal Control *

Toshio Mikami† Michèle Thieullen‡
Hokkaido University Université Paris VI

February 3, 2005

Abstract

We solve optimal transportation problem using stochastic optimal control theory. Indeed, for a super linear cost at most quadratic at infinity, we prove Kantorovich duality theorem by a zero noise limit (or vanishing viscosity) argument.. We also obtain a characterization of the support of an optimal measure in Monge-Kantorovich minimization problem (MKP) as a graph. Our key tool is a duality result for a stochastic control problem which naturally extends (MKP).

Keywords: optimal transportation, Monge-Kantorovich problem, Monge problem, duality, stochastic control, Hamilton-Jacobi-Bellman pde, value function, vanishing viscosity, semi-convex functions.

*This work was done during the visit of the second author (M. Thieullen) to the University of Hokkaido. She would like to thank this university for its hospitality. The visit was supported by the Grant-in-Aid for Scientific research No. 15340051, 16654031, JSPS.

†Department of Mathematics, Hokkaido University, Sapporo 060-0810, Japan; mikami@math.sci.hokudai.ac.jp. Partially supported by the Grant-in-Aid for Scientific research No. 15340047, 15340051, 16654031, JSPS.

‡Corresponding author, Laboratoire de Probabilités et Modèles Aléatoires, Boite 188, 4, Place Jussieu, Université Paris VI, 75252 Paris cedex 05, France

1 Introduction.

Optimal transportation theory consists in the study of the following two minimization problems where P_0 and P_1 are given Borel probability measures on $\mathbf{R}^d$ and $c : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}^+ \cup \{+\infty\}$ is measurable (c is called the cost function). In the Monge-Kantorovich problem, one considers

$$\inf\Big\{\int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)\mu(dxdy)\Big\}, \tag{1.1}$$

on the set of probability measures μ on $\mathbf{R}^d \times \mathbf{R}^d$ with marginals P_0 and P_1 (namely such that $\mu(A \times \mathbf{R}^d) = P_0(A)$ and $\mu(\mathbf{R}^d \times B) = P_1(B)$). In the Monge problem the object of study is

$$\inf\Big\{\int_{\mathbf{R}^d} c(x,g(x))P_0(dx)\Big\}, \tag{1.2}$$

and the infimum is taken over all measurable maps $g : \mathbf{R}^d \mapsto \mathbf{R}^d$ such that the image of P_0 by g is P_1.
In this paper the cost function has the form

$$c(x,y) = L(y-x) \tag{1.3}$$

with $L(u) : \mathbf{R}^d \to [0,\infty)$ convex in u and accordingly we will denote respectively by $T_{MK}(P_0,P_1)$ and $T_M(P_0,P_1)$ the two minimization problems on the corresponding set described above:

$$T_{MK}(P_0,P_1) = \inf\Big\{\int_{\mathbf{R}^d\times\mathbf{R}^d} L(y-x)\mu(dxdy)\Big\}, \tag{1.4}$$

$$T_M(P_0,P_1) = \inf\Big\{\int_{\mathbf{R}^d} L(g(x)-x)P_0(dx)\Big\}. \tag{1.5}$$

Our aim in the present paper is to show that for general L, stochastic optimal control theory can be used efficiently to solve both problems T_{MK} and T_M. This is not clear a priori since classical stochastic control is not well suited to face problems where both marginals (the initial as well as the terminal laws) are fixed. However the idea is natural since one can show that

$$T_{MK}(P_0,P_1) = \inf\Big\{E\int_0^1 L(\frac{d\xi_s}{ds})ds\Big\} \tag{1.6}$$

on the set of $\mathbf{R}^d$- valued absolutely continuous processes (ξ_s) such that ξ_0 (resp. ξ_1) has law P_0 (resp. P_1). One can also consider more general processes with a *small* martingale (or diffusion) part which is interpreted as a small viscocity coefficient. Here, along this line, and using stochastic optimal control, we are able to prove Kantorovich duality by zero noise limit (or vanishing viscosity) and also to recover that the support of μ optimal for T_{MK} is a graph (see precise statement in section 3). Moreover in the quadratic case (when $L(u) = |u|^2$), our present method greatly simplifies the arguments already used by one of us in [10], [11].

Historically the first mass transportation problem to be set was (1.5) with $L(u) = |u|$; this difficult problem remained without solution for a long time and then was approached by Kantorovich via the relaxed form (1.4). One reason for the interest in Monge-Kantorovich problem, at least in probability and statistics, has been that $\sqrt{T(P_0, P_1)}$ defines a distance on the set of probability measures, called Wasserstein distance, and this distance metrizes convergence in distribution (cf. [3], [13]). More recently a geometric approach related to fluid mechanics and pdes has revived the subject (cf. [1], [5], [7]) and established another connection with probality via log-Sobolev inequalities (cf. [15])). There are essentially two types of results concerning problems (1.4) and (1.5). The first type are dual forms of (1.4) provided either by convex analysis arguments (cf. [8], [9]) or with Hamilton-Jacobi pde (cf. [15]). The second type of results are concerned with uniqueness of μ minimizing (1.4) and characterization of its support (cf. [1], [7]). Precise statement will be given respectively in Sections 2 and 3.

Our approach is based on a new duality theorem which we have obtained recently (cf. [12]). It turns out that the present paper together with [12] provide a global treatment of optimal transportation by stochastic control. We believe that our approach may be of interest for at least two reasons. Firstly our key arguments establish a complete correspondence with the deterministic setting: Hamilton-Jacobi pde is replaced by Hamilton-Jacobi-Bellman pde with small viscosity coefficient and we rely heavily on the representation of a solution to Hamilton-Jacobi-Bellman pde as the value function of a control problem. This is the exact analogue of Hopf-Lax formula. It is this representation which enables us to prove the semiconvexity of the stochastic value function (cf. [6]) as it is the case in the deterministic setting. Thus stochastic control theory exactly contains the analogue of what is needed

in the deterministic framework. Secondly, we see an interest at the technical level: our characterization of the support of the optimum for (1.4) is valid when L is superlinear and at most quadratic at infinity which is easy to check. Superlinearity is a standard requirement in the theory of optimal transportation whereas papers on this subject (cf. [7]) assume (except for the quadratic case $L(u) = |u|^2$) that L satisfies a cone type condition which is easy to check only for radial L.

In our approach by zero noise limit there are still questions left. For instance, when $L(u) \sim |u|^2$ at infinity, as an application of our duality theorem, we were able to describe the optimal process of our stochastic control problem as the solution of a forward-backward system (cf. [12]). This process should be a small variation of the deterministic optimal trajectory obtained for (1.4)-(1.5), which is the McCann displacement interpolation between P_0 and P_1 (for details see for instance [15]). This means that the optimal process should converge to this trajectory when its diffusion part tends to zero. Our future aim is to prove such a result.

The paper is organized as follows: in section 2 we recall basic results about optimal transportation and prove Kantorovich duality by zero noise limit. In section 3, by the same method, we recover the description of the support of an optimal measure for (1.4) as a graph.

2 Monge-Kantorovich duality by zero noise limit.

We will be working under the following assumptions: $L(u) : \mathbf{R}^d \to [0, \infty)$ is convex in u,
(A.1) L is *superlinear*: for some $\delta > 1$,

$$\liminf_{|u|\to\infty} \frac{L(u)}{|u|^\delta} > 0.$$

(A.2) (i)$L \in C^3(\mathbf{R}^d)$,
(ii) $D_u^2 L(u)$ is positive definite for all $u \in \mathbf{R}^d$,
We denote by H the Legendre transform of L:

$$H(z) := \sup_{u\in\mathbf{R}^d} \{< z, u > -L(u)\}, \tag{2.1}$$

for $z \in \mathbf{R}^d$; $D_x := (\partial/\partial x_i)_{i=1}^d$ and $< \cdot, \cdot >$ denotes the inner product in $\mathbf{R}^d$.

Let us begin this section by a brief overview of duality results for Monge-Kantorovich problem.

2.1 Duality results for Monge-Kantorovich problem.

Duality plays a fundamental role in the study of Monge-Kantorovich problem. We recall below two duality results and indicate briefly the relationship between them. For details we refer the reader to [13] or [15]. In the sequel we will refer to the first result as *Kantorovich duality*. It was proved first by Kantorovich (cf. [8]) when the cost function is a distance and generalized by Kellerer (cf. [9]).

Theorem 2.1

$$T_{MK}(P_0, P_1) = \sup\Big\{\int_{\mathbf{R}^d} \psi(y) P_1(dy) - \int_{\mathbf{R}^d} \varphi(x) P_0(dx)\Big\}, \tag{2.2}$$

where the supremum is taken over all pairs $(\varphi, \psi) \in L^1(P_0) \times L^1(P_1)$ *satisfying* $\psi(y) - \varphi(x) \leq L(y - x)$.

A second duality result has been proved by Evans and provides a dynamical interpretation for mass transportation problem based on calculus of variations as follows (cf. [4], [15]):

Theorem 2.2

$$T_{MK}(P_0, P_1) = \sup\Big\{\int_{\mathbf{R}^d} \Phi(1, y) P_1(dy) - \int_{\mathbf{R}^d} \Phi(0, x) P_0(dx)\Big\}, \tag{2.3}$$

where the supremum is taken over all continuous viscosity solutions Φ *to the following Hamilton-Jacobi equation:*

$$\frac{\partial \Phi(t, x)}{\partial t} + H(D_x \Phi(t, x)) = 0 \quad ((t, x) \in (0, 1) \times \mathbf{R}^d) \tag{2.4}$$

Let us mention that Theorem 2.1 can be proved by a minimax principle (a detailed proof is given in [15]). In the quadratic case, the supremum in the right-hand side of (2.2) is unchanged if one restricts to the set of pairs of integrable functions (φ, ψ) which are Legendre transforms of each other;

moreover this property extends to non quadratic costs by extending Legendre transforms to L-transforms.. Therefore, using Hopf-Lax, φ can be seen as the value at time 0 of the solution of an Hamilton-Jacobi pde corresponding to terminal (at time 1) value ψ. This yields the statement of Theorem 2.2. For the sake of completeness we now recall Hopf-Lax formula.

Theorem 2.3 *The viscosity solution $\Phi(t,x)$ of Hamilton-Jacobi pde (2.4) satisfying $\Phi(1,x) = f(x)$ admits the representation*

$$\varphi(t,x) = \sup\Big\{ f(\xi_1) - \int_t^1 L(\xi(s))ds \Big\} \tag{2.5}$$

over the set of C^1-trajectories $s \in [t,1] \mapsto \xi(s)$ starting from x at time t: $\xi(t) = x$.

The statement of Theorem 2.2 uses only the case $t = 0$.

Remark 2.1 *When L is strictly convex, the minimizing trajectory for*

$$\inf \int_0^1 L(\dot{\xi}(s))ds \tag{2.6}$$

over the set of C^1-trajectories connecting x to y (such that $\xi(0) = x$ and $\xi(1) = y$ is the straight line $\xi(t) = x + t(y-x)$ and so this infimum is simply $L(y-x)$.

2.2 The stochastic control problem.

The idea for considering the minimization of V_ϵ that we introduce below (cf. [12], [11]) comes first from the identity (1.6) which we already mentioned in the introduction and also from the property that Hamilton-Jacobi-Bellman pde with viscosity $\epsilon > 0$ is uniformly parabolic and therefore existence and uniqueness of very regular solutions can be satisfied, instead of simply viscosity solutions.
Let us consider, for $\epsilon > 0$,

$$\begin{aligned} V_\epsilon(P_0, P_1) \;:=\; \inf\Big\{ & E\Big[\int_0^1 L(\beta_X(t,X))dt\Big] \Big| \\ & PX(t)^{-1} = P_t(t=0,1), X \in \mathcal{A}^\epsilon \Big\}. \end{aligned} \tag{2.7}$$

The set $\mathcal{A}^\epsilon$ is the set of all $\mathbf{R}^d$-valued, continuous semimartingales $\{X(t)\}_{0\leq t\leq 1}$ on a probability space $(\Omega_X, \mathbf{B}_X, P_X)$ such that there exists a Borel measurable $\beta_X : [0,1] \times C([0,1]) \mapsto \mathbf{R}^d$ for which
(i) $\omega \mapsto \beta_X(t,\omega)$ is $\mathcal{B}(C([0,t]))_+$-measurable for all $t \in [0,1]$, where $\mathcal{B}(C([0,t]))$ denotes the Borel σ-field of $C([0,t])$,
(ii) $\{X(t) - X(0) - \int_0^t \beta_X(s,X)ds := \sqrt{\epsilon}W_X(t)\}_{0\leq t\leq 1}$ where W_X is a $\sigma[X(s) : 0 \leq s \leq t]$-Brownian motion.
Results about existence and uniqueness of a minimizer for V_ϵ are gathered in the following statement.

Theorem 2.4 *Let $\epsilon > 0$. Let us assume that $V_\epsilon(P_0, P_1) < +\infty$ and that assumption (A.2) holds. Then*
(i) $V_\epsilon(P_0, P_1)$ admits a minimizer.
(ii) If assumption (A.1) holds with $\delta = 2$, $V_\epsilon(P_0, P_1)$ admits a Markovian minimizer
(iii) If L is strictly convex and assumption (A.1) holds with $\delta = 2$, then $V_\epsilon(P_0, P_1)$ admits a unique minimizer (which is Markovian from (ii)).

Remark 2.2 *Actually the statements (ii) and (iii) will be of no use in the present paper. They were important in [12] in order to characterize the minimizer of (2.7) as the solution of a forward-backward system (which consists in the coupling of a usual sde with a backward one, cf. for instance [2]).*

2.3 Stochastic duality.

In [12] we proved the following duality theorem for the minimization problem (2.7).

Theorem 2.5 *Let $\epsilon > 0$ be fixed and $V_\epsilon(P_0, P_1)$ defined in (2.7). Let us assume that assumption (A.1), (A.2) are satisfied and*

$$V_\epsilon(P_0, P_1) < +\infty. \tag{2.8}$$

Then, the following identity holds

$$\nu_\epsilon(P_0, P_1) = V_\epsilon(P_0, P_1) \tag{2.9}$$

with $\nu_\epsilon(P_0, P_1)$ defined by

$$\nu_\epsilon(P_0, P_1) := \sup\Big\{\int_{\mathbf{R}^d} \varphi(1,y)P_1(dy) - \int_{\mathbf{R}^d} \varphi(0,x)P_0(dx)\Big\}, \tag{2.10}$$

where the supremum is taken over all classical solutions φ, to the following Hamilton-Jacobi-Bellman equation for which $\varphi(1,\cdot) \in C_b^\infty(\mathbf{R}^d)$:

$$\frac{\partial \varphi(t,x)}{\partial t} + \frac{\epsilon}{2}\triangle\varphi(t,x) + H(D_x\varphi(t,x)) = 0 \quad ((t,x) \in (0,1)\times \mathbf{R}^d) \qquad (2.11)$$

where $\triangle := \sum_{i=1}^d \partial^2/\partial x_i^2$ and for $(t,x,z) \in (0,1)\times\mathbf{R}^d\times\mathbf{R}^d$.

Remark 2.3 *Let us notice that our theorem exactly parallels the duality result (2.3) proved by Evans.*

2.4 Zero noise limit.

We will rely on Theorem 2.5 to prove Kantorovich duality (cf. Theorem 2.1) by a zero noise limit argument. We will use the notation

$$\mathcal{T}(P_0,P_1) = \sup\Big\{\int_{\mathbf{R}^d}\psi(y)P_1(dy) - \int_{\mathbf{R}^d}\varphi(x)P_0(dx)\Big\}, \qquad (2.12)$$

as in Theorem 2.1.

The first part of the following statement is our key tool to compare the case $\epsilon > 0$ with the case $\epsilon = 0$.

Theorem 2.6 *Let us assume that $T_{MK}(P_0,P_1) < +\infty$ and that assumptions (A.1)-(A.2) hold. Let us recall that V_ϵ (resp. ν_ϵ) has been defined in (2.7) (resp. (2.10)). We denote by $g_\epsilon \star P_1$ the convolution of P_1 with the Gaussian kernel $g_\epsilon(x) = (2\pi\epsilon)^{-\frac{d}{2}}\exp(-\frac{|x|^2}{2\epsilon})$. Then*
(1) The following sequence of inequalities holds true:

$$\nu_\epsilon(P_0, g_\epsilon\star P_1) \le \mathcal{T}(P_0,P_1) \le T_{MK}(P_0,P_1) \le \liminf_{\epsilon\to 0} V_\epsilon(P_0, g_\epsilon\star P_1) \quad (2.13)$$

(2) As a consequence we recover duality for Monge-Kantorovich problem:

$$T_{MK}(P_0,P_1) = \mathcal{T}(P_0,P_1) \qquad (2.14)$$

Proof of Theorem 2.6 Let us begin with $\mathcal{T}(P_0,P_1) \le T_{MK}(P_0,P_1)$ which is the easiest. Let us take (φ,ψ) such that $\psi(y) - \varphi(x) \le L(y-x)$ and μ with marginals P_0 and P_1. Then

$$\begin{aligned}\int_{\mathbf{R}^d}\psi(y)P_1(dy) - \int_{\mathbf{R}^d}\varphi(x)P_0(dx) &= \int_{\mathbf{R}^d\times\mathbf{R}^d}(\psi(y)-\varphi(x))\mu(dxdy)\\ &\le \int_{\mathbf{R}^d\times\mathbf{R}^d} L(y-x)\mu(dxdy).\end{aligned}$$

The desired inequality follows. We then prove that $\nu_\epsilon(P_0, g_\epsilon \star P_1) \leq \mathcal{T}(P_0, P_1)$. Let $\varphi(t,x)$ denote a solution to HJB pde (2.11) with $\varphi(1,\cdot) \in C_b^\infty(\mathbf{R}^d)$. Let us define

$$\begin{aligned} \overline{\varphi}(t,x) &:= E\varphi(t, x+\sqrt{\epsilon}W_t) \\ X_t^{x,y} &:= x + t(y-x) + \sqrt{\epsilon}W_t \end{aligned}$$

By definition of $\overline{\varphi}$ it holds

$$\begin{aligned} & \int_{\mathbf{R}^d} \varphi(1,y) g_\epsilon \star P_1(dy) - \int_{\mathbf{R}^d} \varphi(0,x) P_0(dx) \\ = & \int_{\mathbf{R}^d} \overline{\varphi}(1,y) P_1(dy) - \int_{\mathbf{R}^d} \overline{\varphi}(0,x) P_0(dx) \end{aligned}$$

and also $\overline{\varphi}(1,y) - \overline{\varphi}(0,x) = E(\varphi(1, X_1^{x,y}) - \varphi(0, X_0^{x,y}))$. Using Ito formula and the fact that φ solves (2.11), we obtain

$$E(\varphi(1, X_1^{x,y}) - \varphi(0, X_0^{x,y})) = E\int_0^1 (< y-x, \nabla\varphi > - H(\nabla\varphi))(s, X_s^{x,y}) ds \tag{2.15}$$

which implies $E(\varphi(1, X_1^{x,y}) - \varphi(0, X_0^{x,y})) \leq L(y-x)$. Therefore

$$\int_{\mathbf{R}^d} \varphi(1,y) g_\epsilon \star P_1(dy) - \int_{\mathbf{R}^d} \varphi(0,x) P_0(dx) \leq \mathcal{T}(P_0, P_1) \tag{2.16}$$

and the desired inequality follows.
We finally prove that $T_{MK}(P_0, P_1) \leq \liminf_{\epsilon\to 0} V_\epsilon(P_0, g_\epsilon \star P_1)$. Let us first notice that

$$0 \leq \liminf_{\epsilon\to 0} V_\epsilon(P_0, g_\epsilon \star P_1) < +\infty \tag{2.17}$$

Indeed positivity comes from definition of V_ϵ; from what we have just proved, $\nu_\epsilon(P_0, g_\epsilon \star P_1) \leq T_{MK}(P, P_1)$ and the duality theorem 2.5 implies that

$$\nu_\epsilon(P_0, g_\epsilon \star P_1) = V_\epsilon(P_0, g_\epsilon \star P_1) \tag{2.18}$$

There exists a sequence (ϵ_n) which converges to 0 such that $V_{\epsilon_n}(P_0, g_{\epsilon_n} \star P_1)$ converges to $\liminf_{\epsilon\to 0} V_\epsilon(P_0, g_\epsilon \star P_1)$. Since for each ϵ, V_ϵ admits a minimizer (cf. Theorem 2.4), there exists a sequence of processes X^n such that $X^n \in \mathcal{A}^{\epsilon_n}$ for all n and

$$\lim_{n\to+\infty} E\int_0^1 L(b^n(X_s^n)) ds = \liminf_{\epsilon\to 0} V_\epsilon(P_0, g_\epsilon \star P_1) \tag{2.19}$$

The superlinearity of L ($L(u) \geq |u|^{\delta}$ with $\delta > 1$) implies that the sequence (X^n) is tight and the limit of any converging subsequence is an absolutely continuous process (cf. [16]). Let $X_t = X_0 + \int_0^t b_X(s)ds$ be such a limit. By the convexity property of L

$$E \int_0^1 L(b_X(s))ds \geq E(L(X_1 - X_0)) \tag{2.20}$$

We conclude using Fatou's Lemma and the fact that the law of X_1 is equal to P_1 since it is the weak limit of a subsequence of $g_\epsilon \star P_1$. Q.E.D.

3 Main result of transport theory using optimal stochastic control.

3.1 Main result of transport theoy.

We first need to recall the definition of L-concave functions.

Definition 3.1 *$\gamma : \mathbf{R}^d \to \mathbf{R} \cup \{-\infty\}$ is L- concave if there exists $\beta : \mathbf{R}^d \to \mathbf{R} \cup \{-\infty\}$ with $\beta \not\equiv -\infty$ such that*

$$\forall x \in \mathbf{R}^d \quad \gamma(x) = \inf_{y \in \mathbf{R}^d} (L(y - x) - \beta(y)) \tag{3.1}$$

Moreover for $r > 0$, $\theta \in]0, \pi[$ and $p \in \mathbf{R}^d$, let $K(p, z, \theta, r)$ be the truncated cone defined by

$$K(p, z, \theta, r) := \{x \in \mathbf{R}^d; \quad |x - p||z| \cos(\frac{\theta}{2}) \leq < z, x - p > \leq r|z|\} \tag{3.2}$$

L is said to satisfy Condition (K) if, when $|p|$ is large enough, there exists $z \in \mathbf{R}^d$ such that L restricted to $K(p, z, \theta, r)$ admits its maximum at p.

Theorem 3.1 *Let L be superlinear (L satisfies assumption (A.1)), strictly convex and satisfying condition (K). Let us assume that $T_{MK}(P_0, P_1) < +\infty$ and that P_0 is absolutely continuous w.r.t. Lebesgue measure on $\mathbf{R}^d$. Let H denote the Legendre transform of L. Then there exists a unique μ minimizing (1.4). The support of μ is the graph of the mapping*

$$g(x) = x - \nabla H(\nabla \phi(x)) \tag{3.3}$$

where ϕ is L-concave.

This theorem is sometimes called the "main result" of optimal transportation (cf. [7]).

3.2 Main result via stochastic control.

We will not be working under condition (K) but rather under the assumption

(A.3) $\exists \quad C > 0, \quad D_u^2 L \leq C$

because of its connection to semiconvexity. The other important tool is the representation of a classical solution of Hamilton -Jacobi-Bellman pde as a value function (cf. [6]). We recall below these properties.

Definition 3.2 *Let f be a function defined on a convex subset of $\mathbf{R}^d$ with values in $\mathbf{R} \cup \{+\infty\}$. The function f is semi-convex if there exists $C > 0$ such that $x \mapsto f(x) + C\frac{|x|^2}{2}$ is convex. When this is true, we say that f is semi-convex with constant C.*

Remark 3.1 *This definition is equivalent to the requirement*

$$\forall (x, z) \quad f(x+z) + f(x-z) - 2f(x) \geq -C|z|^2 \tag{3.4}$$

Proposition 3.1 *Let us assume that assumptions (A.1) and (A.2) hold and let $f \in C_b^\infty(\mathbf{R}^d)$. Then the unique solution φ of the HJB pde (2.11) satisfying $\varphi(1, \cdot) = f$ admits the following representation:*

$$\begin{aligned} \varphi(t,x) \quad = \quad & \sup_{X \in \mathcal{A}_\epsilon} \Big\{ E[f(X_1)|X_t = x] \\ & -E\Big[\int_t^1 L(\beta_X(s,X))ds \Big| X_t = x\Big]\Big\}, \end{aligned} \tag{3.5}$$

where $\mathcal{A}_\epsilon$ was defined in section 2.2. Moreover the optimal control is markovian and given by the function $\nabla H(\nabla\varphi(s, \cdot))$.

This formula obviously provides a direct stochastic analog of Hopf-Lax formula (2.5). It has played a fundamental role in the derivation of Theorem 2.5 and will also be crucial in the sequel through the following corollary. For extensions of this corollary we refer the reader to [6].

Corollary 3.1 *Let us denote by $\varphi(t,x)$ the value function defined by (3.5). Under assumption (A.3), $\varphi(0,\cdot)$ is semi-convex with constant C.*

Proof Let $(X_t^x; t \in [0,1])$ be optimal for (3.5) in $\mathcal{A}_\epsilon$; namely, $X_t^x = x + \int_0^t \nabla H(\nabla\varphi(s, X_s^x))ds + \sqrt{\epsilon}W_t$ and

$$\varphi(0,x) = E[\varphi(1, X_1^x) - \int_0^1 L(\nabla H(\nabla\varphi(s, X_s^x)))ds] \tag{3.6}$$

Let us set $X_t^1 := X_t^x + (1-t)z$ as well as $X_t^2 := X_t^x - (1-t)z$; these processes both belong to $\mathcal{A}_\epsilon$ and satisfy $X_0^1 = x + z$ and $X_0^2 = x - z$ as well as $X_1^1 = X_1^2 = X_1^x$. For simplicity let us set $\beta_t^x := \nabla H(\nabla\varphi(t, X_t^x))$. From the definition of φ in (3.5) it follows that

$$\begin{aligned} &\varphi(0,x+z) + \varphi(0,x-z) - 2\varphi(0,x) \geq \\ &E\int_0^1 (2L(\beta_t^x) - L(\beta_t^x + z) - L(\beta_t^x - z))dt \end{aligned} \tag{3.7}$$

The conclusion follows from assumption (A.3). Q.E.D.

We now consider μ optimal for the Monge-Kantorovich problem: μ has marginals P_0 and P_1 and satisfies

$$T_{MK}(P_0, P_1) = \int_{\mathbf{R}^d \times \mathbf{R}^d} L(y-x)\mu(dxdy) \tag{3.8}$$

For $\epsilon > 0$ and φ_ϵ a solution of (2.11) let us set, as in the previous section,

$$\Psi_\epsilon(t,x) := E(\varphi(t, x + \sqrt{\epsilon}W_t)) \tag{3.9}$$

We obtain a first description for the support of μ.

Proposition 3.2 *Let us assume that assumptions (A.1-(A.2) hold true. There exists a sequence (ϵ_n) converging to 0 and a sequence (φ_{ϵ_n}) of solutions of (2.11) such that $\mu(S) = 1$ with S given by*

$$S := \{(x,y); \lim_{n\to+\infty} \Psi_{\epsilon_n}(1,y) - \Psi_{\epsilon_n}(0,x) = L(y-x)\} \tag{3.10}$$

Proof For each $\epsilon > 0$, the definition of $\nu_\epsilon(P_0, g_\epsilon \star P_1)$ as in (2.10) and that of Ψ_ϵ in (3.9), we can choose $\varphi_\epsilon(t,x)$, solution of (2.11), such that

$$\nu_\epsilon(P_0, g_\epsilon \star P_1) - \epsilon \leq \int_{\mathbf{R}^d} \varphi_\epsilon(1,y) g_\epsilon \star P_1(dy) - \int_{\mathbf{R}^d} \varphi_\epsilon(0,x) P_0(dx). \tag{3.11}$$

We have seen in the previous section that ψ_ϵ satisfies

$$\Psi_\epsilon(1,y) - \Psi_\epsilon(0,x) \leq L(y-x) \tag{3.12}$$

as well as

$$\begin{aligned} & \int_{\mathbf{R}^d} \varphi_\epsilon(1,y) g_\epsilon \star P_1(dy) - \int_{\mathbf{R}^d} \varphi_\epsilon(0,x) P_0(dx) \\ = & \int_{\mathbf{R}^d} \Psi_\epsilon(1,y) P_1(dy) - \int_{\mathbf{R}^d} \Psi_\epsilon(0,x) P_0(dx) \end{aligned}$$

Therefore the two inequalities below hold

$$\nu_\epsilon(P_0, g_\epsilon \star P_1) - \epsilon \leq \int_{\mathbf{R}^d \times \mathbf{R}^d} (\Psi_\epsilon(1,y) - \Psi_\epsilon(0,x)) \mu(dxdy) \leq T_{MK}(P_0, P_1) \tag{3.13}$$

Moreover from inequality (3.12) it follows that $L(y-x) - (\Psi_\epsilon(1,y) - \Psi_\epsilon(0,x))$ is non negative. By letting ϵ go to 0 in Theorem 2.6 we conclude that

$$\lim_{\epsilon \longrightarrow 0} \int_{\mathbf{R}^d \times \mathbf{R}^d} |\Psi_\epsilon(1,y) - \Psi_\epsilon(0,x) - L(y-x)| \mu(dxdy) = 0. \tag{3.14}$$

This convergence in $L^1(\mu)$ implies a. s. convergence w.r.t. μ for a subsequence (ϵ_n). Q.E.D.

In the sequel we keep the notations of Proposition 3.2. As preparation for our proof of the main result let us introduce the following function:

Definition 3.3 *Let $a \in \mathbf{R}^d$. We denote by ψ_a the function defined by*

$$\psi_a(x) := \limsup_{n \to +\infty} (\Psi_{\epsilon_n}(0,x) - \Psi_{\epsilon_n}(0,a)) \tag{3.15}$$

Proposition 3.3 *Under assumption (A.3)*
(1) ψ_a is a semi-convex function on its domain $D_a := \{x \in \mathbf{R}^d; \psi_a(x) < +\infty\}$.
(2) the set D_a is convex and independent of $a \in \pi_1(S) := \{x \in \mathbf{R}^d; \exists y \in \mathbf{R}^d \quad (x,y) \in S\}$.

Proof Proof of (1) *1st step:* it is not difficult to check first that the $\limsup f_n$ of a sequence (f_n) of semi-convex functions with the same constant C is itself

semi-convex with this same constant. Indeed, from inequality (3.4) for all (x, z) and n it holds

$$2f_n(x) - C|z|^2 \leq f_n(x+z) + f_n(x-z). \tag{3.16}$$

The conclusion is straightforward since $\limsup f_n = \inf_k \sup_{n \geq k} f_n$.
2nd step: By definition (3.9) $\Psi_{\epsilon_n}(0, \cdot) \equiv \varphi_{\epsilon_n}(0, \cdot)$. It suffices to apply Corollary 3.1.
Proof of (2) Let us first prove two facts about ψ_a: let $(a, b) \in S$. For all points u, c in $\mathbf{R}^d$

$$\psi_a(u) \geq L(b-a) - L(b-u) \tag{3.17}$$

$$\psi_a(u) \geq \psi_c(u) + L(b-a) - L(b-c) \tag{3.18}$$

It is sufficient to prove the second inequality which implies the first one. Let $(a, b) \in S$. The obvious identity

$$\begin{aligned} \Psi_{\epsilon_n}(0,u) &- \Psi_{\epsilon_n}(0,a) = \\ \Psi_{\epsilon_n}(0,u) - \Psi_{\epsilon_n}(0,c) &+ \Psi_{\epsilon_n}(0,c) - \Psi_{\epsilon_n}(1,b) + \Psi_{\epsilon_n}(1,b) - \Psi_{\epsilon_n}(0,a) \end{aligned}$$

together with (3.12) imply

$$\begin{aligned} \Psi_{\epsilon_n}(0,u) &- \Psi_{\epsilon_n}(0,a) \geq \\ \Psi_{\epsilon_n}(0,u) - \Psi_{\epsilon_n}(0,c) &+ \Psi_{\epsilon_n}(1,b) - \Psi_{\epsilon_n}(0,a) - L(b-c) \end{aligned}$$

It remains to let n go to $+\infty$ to conclude using the definition of ψ_a and ψ_c and Proposition 3.2. The first inequality is a consequence when u is taken equal to c. The domain of ψ_a is a convex set since it coincides with the domain of the convex function $\psi_a + \frac{C}{2}|\cdot|^2$. Moreover let a and a' in $\pi_1(S)$. By applying (3.18) twice, once to (a, a') and afterwards to (a', a), we deduce immediately that $D_a = D_{a'}$. Q.E.D.

Corollary 3.2 *Let us assume that $L \in C^1(\mathbf{R}^d)$. Let $(x, y) \in S$ and $i \in \{1, ..., d\}$.*
(1) If for some $h \neq 0$ and $z \in \mathbf{R}^d$, $(x + he_i, z) \in S$, then

$$L(y-x) - L(y-(x+he_i)) \leq \psi_x(x+he_i) \leq L(z-x) - L(z-(x+he_i)) \tag{3.19}$$

(2) Let us assume that $\forall i \in \{1, ..., d\}$ there exist sequences $(h_n^{(i)})$ and $(y_n^{(i)})$ such that $h_n^{(i)} \to 0$, $y_n^{(i)} \to y$ and $\forall n$, $(x + h_n^{(i)} e_i, y_n^{(i)}) \in S$. Then

$$\forall i \in \{1, ..., d\} \quad \lim_{n \to +\infty} \frac{1}{h_n^{(i)}} \psi_x(x + h_n^{(i)} e_i) = \partial_i L(y-x) \tag{3.20}$$

Proof Apply now the inequality (3.18) twice: the first time with $(a,b) = (x,y)$ and $u = c = x + he_i$, the second time with $(a,b) = (x + he_i, z)$ and $u = x+he_i$, $c = x$. Remember that by definition $\psi_x(x) = \psi_{x+he_i}(x+he_i) = 0$. This proves (1).
(2) Apply inequality (1) just proved; let $n \to +\infty$. The conclusion follows since L is assumed to be continuously differentiable. Q.E.D.

Theorem 3.2 *Let $P_0(dx) \ll dx$ and $T_{MK}(P_0, P_1) < +\infty$. Under assumption (A.3), there exists a function Φ such that S is the graph of $Id+\nabla H \circ \nabla \Phi$.*

Proof Take $a \in \pi_1(S)$. Suppose that for $i \in \{1, \cdots, d\}$, there exists b_i for which $(a, b_i) \in S$ and such that

$$U_{i,j,\frac{1}{n}}(a, b_i) \cap S \neq \emptyset \quad (j = +1, -1, n \geq 1),$$

where

$$U_{i,+1,\frac{1}{n}}(a, b_i) := \{(x,y) \in \mathbf{R}^d \times \mathbf{R}^d | x-a = h_x \mathbf{e}_i, 0 < h_x, |x-a|^2 + |y-b_i|^2 < n^{-2}\},$$

$$U_{i,-1,\frac{1}{n}}(a, b_i) := \{(x,y) \in \mathbf{R}^d \times \mathbf{R}^d | x-a = h_x \mathbf{e}_i, 0 > h_x, |x-a|^2 + |y-b_i|^2 < n^{-2}\}.$$

Then $\partial_i \psi_a(a)$ exists from Corollary 3.2. Moreover from inequality (3.17), for any $b \in \pi_2(S)$ for which $(a,b) \in S$, it holds

$$f(x) := \psi_a(x) + L(b-x) \geq L(b-a),$$

where equality hods if $x = a$. If $\nabla \psi_a(a)$ exists, then $\nabla f(a)$ exists and is equal to 0. Hence

$$\nabla f(a) = \nabla \psi_a(a) - \nabla L(b-a) = 0.$$

We have just proved that if $\nabla \psi_a(a)$ exists and $(a,b) \in S$, then $b = a + \nabla H(\nabla \psi_a(a))$.
Suppose now that $\partial_1 \psi_a(a)$ does not exist. Then for any b_1 for which $(a, b_1) \in S$, there exist $j = +1$ or -1 and $n \geq 1$ such that

$$U_{1,j,\frac{1}{n}}(a, b_1) \cap S = \emptyset.$$

If $(\overline{a}, \overline{b}_1) \in S$, $U_{1,j,\frac{1}{n}}(\overline{a}, \overline{b}_1) \cap S = \emptyset$ and $a - \overline{a} = h\mathbf{e}_1$ $(h \in \mathbf{R})$, then

$$|a - \overline{a}|^2 + |b_1 - \overline{b}_1|^2 \geq n^{-2}.$$

Therefore such points can be at most countably many.
Hence for any $(x_2, \cdots, x_d) \in \mathbf{R}^{d-1}$, the following set contains at most countably many points:

$$\{x \in \mathbf{R} | \partial_1 \psi_{\mathbf{x}}(\mathbf{x}) \text{ does not exist for } \mathbf{x} := (x, x_2, \cdots, x_d)\}.$$

By Fubini's theorem, the proof is over. Q.E.D.

References

[1] Brenier, Y., Benamou, J.D., 1999. A numerical method for the optimal mass transport problem and related problems, in: Caffarelli, L.A., Milman, M. (Eds.), Monge Ampre equation: applications to geometry and optimization, Proc. of the NSF-CBMS Conference, Deerfield Beach, FL 1997, Contemporary Mathematics 226, Amer. Math. Soc., Providence, RI, pp. 1-11.

[2] Delarue, F., *On the existence and uniqueness of solutions to FBSDEs in a non-degenerate case.* Stochastic Process. Appl. 99, 209–286 (2002)

[3] Dudley, R.M., 1976. Probabilities and metrics. Convergence of laws on metric spaces, with a view to statistical testing, Lecture Notes Series, No. 45, Matematisk Institut, Aarhus Universitet, Aarhus.

[4] Evans, L.C., 1998. Partial differential equations, Graduate Studies in Mathematics Vol. 19, Amer. Math. Soc., Providence, RI, USA.

[5] Evans, L.C., *Partial differential equations and Monge-Kantorovich mass transfer* in: Yau, S.T. (Ed.), Current developments in mathematics, Cambridge, MA 1997, Int. Press, Boston, MA, pp. 65-126 (1999)

[6] Fleming, W.H., Soner, H.M., 1993. Controlled Markov Processes and Viscosity Solutions, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo.

[7] Gangbo, W., McCann, R., J., *The geometry of optimal transportation.* Acta Math. 177, 2, 113-161 (1996)

[8] Kantorovich, L.,V., *On the translocation of masses.* C. R. (Dokl.) Acad. Sci. USSR, 37, 199-201 (1942)

[9] Kellerer, H., G., *Duality theorem for marginal problems.* Z. Wahrsch. Verw. Gebiete, 67, 399-432, (1984)

[10] Mikami, T., *Optimal control for absolutely continuous stochastic processes and the mass transportation problem.* Elect. Comm. in Probab. 7, 199-213 (2002)

[11] Mikami, T., *Monge's problem with a quadratic cost by the zero noise limit of h-path processes.* Probab. Theory Related Fields **129**, 245-260 (2004)

[12] Mikami, T., Thieullen, M. *Duality Theorem for Stochastic Optimal Control Problem.* Hokkaido University Preprint Series in Mathematics , Series 652, June 2004. Submitted to Stoch. Proc. and their Appl.

[13] Rachev, S.T., Rüschendorf, L., 1998. Mass transportation problems, Vol. I: Theory, Vol. II: Application, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo.

[14] Rüschendorf, L., Thomsen, W., 1993. *Note on the Schrödinger equation and I-projections,* Statist. Probab. Lett. 17, 369-375.

[15] Villani, C., 2003. Topics in Optimal Transportation, Graduate Studies in Mathematics Vol. 58, Amer. Math. Soc., Providence, RI.

[16] Zheng, W.A., 1985. *Tightness results for laws of diffusion processes application to stochastic mechanics.* Ann. Inst. Henri Poincaré 21, pp. 103-124.