BULLETIN DE LA S. M. F.

TAOUFIQ ABDELLAOUI
HENRI HEINICH

Caractérisation d'une solution optimale au problème de Monge-Kantorovitch

Bulletin de la S. M. F., tome 127, nº 3 (1999), p. 429-443
<http://www.numdam.org/item?id=BSMF_1999__127_3_429_0>

Bull. Soc. math. France,
127, 1999, p. 429–443.

CARACTÉRISATION D'UNE SOLUTION OPTIMALE AU PROBLÈME DE MONGE-KANTOROVITCH

PAR TAOUFIQ ABDELLAOUI ET HENRI HEINICH (*)

RÉSUMÉ. — Soient P et Q deux probabilités définies sur la tribu borélienne d'un espace métrique séparable complet M et $c(x, y)$ une fonction continue de $M \times M$ dans $\mathbb{R}^+$. On considère la fonctionnelle

$$d_c(P, Q) = \inf\{E(c(X, Y))\ ;\ X \text{ de loi } P,\ Y \text{ de loi } Q\}.$$

Lorsque P est « c-continue » et Q discrète, nous montrons qu'un couple (X, Y) réalise le minimum si et seulement si $Y = f(X)$ où f est une fonction optimale pour le couple (P, Q). Nous établissons l'unicité de f et nous montrons que la condition de c-cyclique monotonie P-*p.s.* est nécessaire et suffisante pour que f soit optimale. Pour un espace de Hilbert et $c(x, y) = \theta(x-y), \theta$ strictement convexe, nous obtenons, lorsque P vérifie une condition (*) et Q quelconque, l'existence, l'unicité P-*p.s.*, la c-cyclique monotonie P-*p.s.* et l'expression de la fonction optimale.

ABSTRACT. — CARACTERIZATION OF AN OPTIMAL SOLUTION TO MONGE-KANTOROVITCH'S PROBLEM. — Let P and Q be two probabilities on a complete separable metric space M and c a continuous function on $M \times M$. We consider

$$d_c(P, Q) = \inf\{E(c(X, Y))\ ;\ X \text{ has the law } P,\ Y \text{ has the law } Q\}.$$

We show that, when P is "c-nonatomic" and Q atomic, a pair (X, Y) verifies the relation $d_c(P, Q) = E(c(X, Y))$ if and only if $Y = f(X)$, where f is the unique optimal function for (P, Q). We also prove that the condition of c-cyclical monotonicity is necessary and sufficient for f to be optimal. For a Hilbert space we suppose that $c(x, y) = \theta(x - y)$, θ strictly convexe and P verifies a condition (*). We give the expression of the unique optimal function f for (P, Q).

(*) Texte reçu le 20 juillet 1998, révisé le 7 janvier 1999, accepté le 12 février 1999.
T. ABDELLAOUI, Université Hassan II, Faculté des Sciences Casa I, 8 route d'el Jadida, BP 5366, Casablanca (Maroc). Email : abdellaoui@facsc-achok.ac.ma.

H. HEINICH, UPRES-A. CNRS 6085. INSA de Rouen, place Émile Blondel, 76131 Mont St Aignan CEDEX (France). Email : heinich@insa-rouen.fr.

Classification AMS : 60B05, 60B11, 90C08.

Mots clés : problème de Monge-Kantorovitch, probabilités à marginales fixées, cyclique monotonie, sous-différentiel.

1. Introduction, notations et définitions

Soient (M, d) un espace métrique séparable et complet, $\mathcal{B}$ sa tribu borélienne, P et Q deux probabilités définies sur $\mathcal{B}$. Les variables considérées dans la suite sont définies sur l'espace de Lebesgue $([0,1], \mathcal{B}_1, \lambda)$ où λ est la mesure de Lebesgue et $\mathcal{B}_1$ la tribu borélienne complétée. On se donne une fonction continue c définie de $M \times M$ et à valeurs dans $\mathbb{R}^+$ qui s'interprète comme une fonction de coût. La fonctionnelle de Monge-Kantorovitch est

$$(1) \qquad d_c(P,Q) = \inf\Big\{\int_{M\times M} c(x,y)\,\mathrm{d}\mu\,;\ \mu \in \mathcal{M}(P,Q)\Big\},$$

où $\mathcal{M}(P,Q)$ est l'ensemble des probabilités sur l'espace produit dont les marginales sont respectivement P et Q. En fait, nous devons considérer le sous-espace $\mathcal{M}_c(P,Q)$ formé par les probabilités $\mu \in \mathcal{M}(P,Q)$ telles que $\int c(x,y)\,\mathrm{d}\mu(x,y) < +\infty$; par simplication, cela sera sous entendu.

En termes de variables aléatoires (*v.a.*), la formulation (1) est équivalente à

$$(2) \qquad d_c(P,Q) = \inf\Big\{\int_{[0,1]} c(X,Y)\,\mathrm{d}\lambda\,;\ X \text{ de loi } P,\ Y \text{ de loi } Q\Big\}.$$

Par commodité de notations, on écrit

$$(X,Y) \in \mathcal{M}(P,Q)$$

si la loi du couple appartient à $\mathcal{M}(P,Q)$. Un couple (X,Y) est un *c-couple optimal* si

$$(X,Y) \in \mathcal{M}(P,Q) \quad \text{et} \quad d_c(P,Q) = \int c(X,Y)\,\mathrm{d}\lambda.$$

On désigne par $\mathrm{opt}_c(P,Q)$ l'ensemble des c-couples optimaux. Il est classique que $\mathrm{opt}_c(P,Q)$ est non vide, voir par exemple [2].

Le problème qui nous intéresse est l'existence d'une fonction f, dite *optimale* pour le couple (P,Q), telle que $(X, f(X)) \in \mathrm{opt}_c(P,Q)$ pour une (donc pour toute) *v.a.* X de loi P ; on écrit $X \sim P$. Une telle fonction minimise le transport de P en Q selon la fonction de coût c.

Dans le cas où M est un espace de Hilbert, $c(x,y) = \|x-y\|^2$ et P est une probabilité fortement continue, Cuesta et Matrán ont montré l'existence d'une fonction optimale pour le couple (P,Q), *cf.* [4] et [5]. Rachev et Rüschendorf, sous les mêmes hypothèses, caractérisent un couple optimal par la cyclique monotonie, *cf.* [14]. Des résultats analogues ont été démontré par Knott et Smith, *cf.* [11]. Plus récemment, lorsque $M = \mathbb{R}^n$ et P s'annule sur les ensembles de dimension de Hausdorff $p < n$, Gangbo et MacCann ont établi l'existence d'une fonction optimale, *cf.* [9]

[10] et [13], pour une fonction de coût de la forme $c(x,y) = h(\|x-y\|)$ où h est strictement convexe ou strictement concave. Pour une lecture détaillée, nous référons le lecteur au livre de Rachev [15], au *survey* [6] et sa bibliographie.

Cet article s'organise en une introduction et deux parties.

Nous donnons une solution au problème posé, l'existence de la fonction optimale f lorsque P et Q sont deux probabilités définies sur un espace métrique séparable et complet (M,d), et dont l'une est c-continue et l'autre discrète. C'est autour de la caractérisation de f par sa c-cyclique monotonie que s'organise la première partie.

Le second paragraphe est consacré aux espaces de Hilbert séparables $\mathbb{H}$. En reliant la c-cyclique monotonie à la c-convexité, et lorsque P vérifie une condition (*), nous obtenons, pour c de la forme $c(x,y) = \theta(x-y)$, θ strictement convexe, la fonction de transport f de P en toute autre probabilité Q, minimisant le coût c. Cette fonction optimale f s'écrit

$$f(x) = x - (\nabla\theta)^{-1} \circ \nabla F(x) \quad P\text{-}p.s.$$

Donnons auparavant quelques définitions. Afin de simplifier les écritures, notons :

$$\gamma(x,a,b) = c(x,b) - c(x,a).$$

Une fonction f de M dans M est *c-cycliquement monotone d'ordre k*, si pour toute suite cyclique de longueur k $(x_1, x_2, \dots, x_k = x_0)$, on a

$$\sum_{i=1}^{k} \gamma\big(x_i, f(x_i), f(x_{i-1})\big) \geq 0.$$

Une fonction f est *c-cycliquement monotone* (plus rapidement *c-c.m.*), si elle est *c-c.m.* d'ordre k pour tout k.

Il est aisé de voir qu'une fonction f est *c-c.m.* si et seulement si, pour toute suite finie $(x_1, \cdots, x_k)$ et toute permutation $\tau \in \mathcal{S}(k)$ où $\mathcal{S}\ (k)$ est l'ensemble des permutations de $\{1, \dots, k\}$, on a :

$$\sum_{1}^{k} \gamma\big(x_i, f(x_i), f(x_{\tau(i)})\big) \geq 0.$$

L'introduction de probabilités conduit aux définitions suivantes :

Une fonction f est dite *c-cycliquement monotone P-p.s.*, (*c-c.m. P-p.s.*), si elle est *c-c.m.* en dehors d'un ensemble P négligeable.

Les notions précédentes de *c-c.m.*, et de *c-c.m. P-p.s.*, s'adaptent naturellement à un couple (X,Y). Par exemple, un couple (X,Y) est *c-c.m. λ-p.s.*, s'il existe un ensemble négligeable N, tel que

$$\forall k > 1,\ \forall \tau \in \mathcal{S}(k),\ \forall (t_1,\dots,t_k),\ t_i \notin N,\quad \sum_{i=1}^{k} \gamma\big(X(t_i), Y(t_i), Y(t_{i-1})\big) \geq 0.$$

Enfin, une probabilité P est dite *c-continue* si pour tout couple $(a,b) \in M^2$ tel que $a \neq b$, et tout $\alpha \in \mathbb{R}$, on a

$$P\big(x \mid \gamma(x,a,b) = \alpha\big) = 0.$$

Une telle probabilité est continue, $P(x) = 0$ pour tout $x \in M$. La réciproque est fausse en général ; par exemple pour $M = \mathbb{R}^2$ et $c(x,y) = \|x-y\|^2$, prendre une probabilité continue telle que $P\{x \mid \gamma(x,a,b) = 0\} = 1$.

2. Fonction optimale pour les espaces métriques

Dans cette partie, M est un espace métrique séparable avec sa tribu borélienne $\mathcal{B}$. Nous allons montrer l'existence d'une fonction optimale et la caractériser.

Le lemme suivant établit une première relation entre l'optimalité et la c-cyclique monotonie.

LEMME 2.1. — *Soient P et Q deux probabilités définies sur M. Si $(X,Y) \in \operatorname{opt}_c(P,Q)$, alors (X,Y) est c-c.m. λ-p.s.*

Une preuve succincte de ce lemme se trouve à l'origine dans [1]. Plus récemment, Gangbo et MacCann en donnent, dans [10], une version plus détaillée, adaptée à des espaces localement compacts... Pour cette raison, donnons une preuve relative à notre cadre.

La preuve se fait par négation. Montrons qu'il existe $\epsilon > 0$ et n boréliens $(A_1, A_2, \cdots, A_n = A_0)$ disjoints, de même mesure de Lebesgue, et tels que :

$$\forall t_i \in A_i, \quad \sum_{i=1}^{n} \gamma\big(X(t_i), Y(t_i), Y(t_{i-1})\big) \leq -\epsilon.$$

En effet, soit h l'application de $[0,1]^n$ et à valeurs dans M^{3n} définie par :

$$(t_1, t_2, .., t_n) \longmapsto \big((X(t_1), Y(t_1), Y(t_n)), (X(t_2), Y(t_2), Y(t_1)), \dots, (X(t_n), Y(t_n), Y(t_{n-1})\big)$$

et soit g l'application de M^{3n} et à valeurs dans $\mathbb{R}$ définie par :

$$g\big((u_i, v_i, w_i)_{i=1}^n\big) = \sum_{i=1}^n \gamma(u_i, w_i, v_i).$$

La négation implique l'existence de $m > 0$ tel que

$$\lambda^{\otimes n}\Big\{(t_1, t_2, \dots, t_n = t_0)\ ;\ g \circ h(t_1, t_2, \dots, t_n) < -\frac{1}{m}\Big\} > 0.$$

Si μ est la mesure image de $\lambda^{\otimes n}$ par h, on déduit que $\mu\{g < -m^{-1}\} > 0$. L'application g étant continue, il existe n ouverts $(O_1, \dots, O_n = O_0)$ tels que $(O_1 \times \dots \times O_n) \subseteq \{g < -m^{-1}\}$ et $\mu(O_1 \times \dots \times O_n) > 0$. Quitte à diminuer les ouverts O_i, on peut supposer que $O_i = O_i^1 \times O_i^2 \times O_i^3$. Il suffit alors de prendre des boréliens

$$A_i \subset X^{-1}(O_i^1) \cap Y^{-1}(O_i^2 \cap O_{i-1}^3),$$

avec $\lambda(A_i) = \lambda(A_j)$, pour obtenir les ensembles recherchés.

La contradiction s'obtient en considérant un automorphisme ponctuel de Lebesgue, τ, tel que $\tau(A_i) = A_{i-1}$ et τ est l'identité sur B, complémentaire de $\bigcup A_i$. Alors $(X \circ \tau, Y) \in \mathcal{M}(P, Q)$ et

$$\sum_1^n c\big(X \circ \tau(t_i), Y(t_i)\big) + \epsilon \le \sum_1^n c\big(X(t_i), Y(t_i)\big), \quad t_i \in A_i,$$
$$c(X \circ \tau, Y)1_B = c(X, Y)1_B.$$

En intégrant, on obtient la contradiction souhaitée. □

REMARQUE. — Dans le cas hilbertien et pour $c(x, y) = \|x - y\|^2$, on retrouve le théorème 2.3 de [4].

Montrons, pour certaines probabilités, que tout couple optimal est obtenu par une fonction optimale.

LEMME 2.2. — *Soient P et Q deux probabilités définies sur M, avec P c-continue et $Q = \sum_i \alpha_i \delta_{a_i}$. Il existe une partition, $(B_i)_{i \in \mathbb{N}}$, avec*

$$P(B_i) = \alpha_i \quad \textit{pour tout } i,$$

telle que la fonction $f = \sum_i a_i 1_{B_i}$ est optimale pour le couple (P, Q).

Inversement, si $Y \sim Q$, il existe une v.a. X de loi P telle que $Y = f(X)$ et $(X, Y) \in \mathrm{opt}_c(P, Q)$. De plus, la fonction f ainsi définie est c-c.m. P-p.s.

Preuve. — Pour $(X,Y) \in \operatorname{opt}_c(P,Q)$ et $Y = \sum a_i 1_{I_i}$, avec le théorème 8.4.1 et le lemme 8.4.2 de [7], $X(I_i)$ appartient à la tribu borélienne de M, complétée par rapport à P. Considérons les parties $J \subset \mathbb{N}$, telles que $P(A_J = \bigcap_{j \in J} X(I_j)) > 0$. Pour une telle partie, on a

$$\lambda\big(X^{-1}(X(I_i) \cap X(I_j))\big) > 0 \quad \text{pour tout } (i,j) \in J^2.$$

Le lemme 2.1 assure que $\lambda^{\otimes 2}$-*p.s.*,

$$c\big(X(t),Y(t)\big) + c\big(X(s),Y(s)\big) \leq c\big(X(t),Y(s)\big) + c\big(X(s),Y(t)\big).$$

Soient x et x' appartenant à A_J, et $(i,j) \in J^2$; alors il existe $(t,s) \in I_i \times I_j$, tel que $x = X(t)$, $x' = X(s)$. Donc, $P^{\otimes 2}$-*p.s.*,

$$c(x,a_i) + c(x',a_j) = c(x',a_i) + c(x,a_j).$$

Ainsi, pour tout couple $(i,j) \in J^2$ et pour P-presque tout $x \in A_J$:

$$(*) \qquad c(x,a_i) = c(x,a_j) + \delta_{i,j},$$

où $\delta_{i,j}$ est une constante ne dépendant que de i et j.

La c-continuité de P assure que $P(A_J) = 0$ si cardinal$(J) > 1$. Donc les $B_i = X(I_i)$ forment P-*p.s.* une partition de M.

La fonction f définie par $f = \sum_i a_i 1_{B_i}$ est optimale pour (P,Q). En effet, il est évident que f est de loi Q, relativement à P et que $Y = f(X)$.

Seconde partie : partons d'une variable aléatoire $Y = \sum_i a_i 1_{I_i}$ avec $\lambda(I_i) = \alpha_i$, les I_i formant une partition de $[0,1]$, de loi Q. Avec la première partie, soit f une fonction optimale : $(X_0, f(X_0)) \in \operatorname{opt}_c(P,Q)$. Il existe un automorphisme τ, préservant la mesure et induit par une application ponctuelle de $[0,1]$ sur $[0,1]$, tel que $\tau((f \circ X_0)^{-1}\{a_i\}) = I_i$. La variable aléatoire $X = X_0 \circ \tau$ est de loi P et $f(X) = Y$. On conclut que $(X,Y) \in \operatorname{opt}_c(P,Q)$.

La c-cyclique monotonie P-*p.s.* est une conséquence du lemme 2.1. □

Remarque. — Soient X une *v.a.* sur $[0,1]$ de loi P, c-continue, et A un borélien de $[0,1]$ tel que $\lambda(A) > 0$. Notons P_{X_A} la loi de la restriction de X à A pour la trace de λ sur A. On obtient :

Si $(X,Y) \in \operatorname{opt}_c(P,Q)$, *alors* $(X_A, Y_A) \in \operatorname{opt}_c(P_{X_A}, Q_{Y_A})$.

Le résultat principal de cette partie est le théorème suivant.

THÉORÈME 2.3. — *Soient P et Q deux probabilités définies sur M, avec P c-continue et Q discrète. Alors :*

(a) *Il existe une unique (P-p.s.) fonction f optimale pour (P, Q).*

(b) *Une fonction f de loi Q par rapport à P est optimale pour (P, Q) si et seulement si elle est c-cycliquement monotone P-p.s.*

(c) *De plus, les assertions suivantes sont équivalentes :*

1)$(X, Y) \in \mathrm{opt}_c(P, Q)$

2)$(X, Y) \in \mathcal{M}(P, Q)$ *et* (X, Y) *est c-c.m. λ-p.s.*

3)$Y = f(X) P$*-p.s.*

La preuve de ce théorème se fait par étapes, en approchant la probabilité Q par des probabilités Q_n à support fini.

LEMME 2.4. — *Soient P et Q deux probabilités définies sur M, avec P c-continue et $Q = \sum_1^n \alpha_i \delta_{a_i}$. Les deux assertions suivantes sont équivalentes :*

(i) $(X, Y) \in \mathrm{opt}_c(P, Q)$;

(ii) $(X, Y) \in \mathcal{M}(P, Q)$ *et* (X, Y) *est c-c.m. λ-p.s.*

Preuve. — Montrons que (ii) implique (i). Soient (X, Y) un couple vérifiant (ii) et $f = \sum_1^n a_i 1_{B_i}$ une fonction optimale pour (P, Q) par le lemme 2.2.

Nous allons montrer que, pour tout i, l'ensemble $I_i = Y^{-1}(a_i)$ est égal, λ-*p.s.*, à $X^{-1}(B_i)$. Ceci suffit pour établir (i) car alors $f(X) = Y$, λ-*p.s.* De plus, ceci montre l'unicité de la fonction optimale.

Par négation, supposons que pour un i donné, I_i soit différent de $X^{-1}(B_i), \lambda$ *p.s.* Comme $\lambda(I_k) = \lambda(X^{-1}(B_k)) = \alpha_k$ pour tout k, il existe un autre indice, dépendant de i, soit $j(i) \neq i$, tel que

$$\lambda\big(I_i \cap X^{-1}(B_{j(i)})\big) > 0.$$

Par conséquent, $I_{j(i)}$ et $X^{-1}(B_{j(i)})$ sont λ-*p.s.* différents. En itérant ce procédé, on détermine ainsi une suite d'indices, notés $\beta(i)$, telle que

$$\lambda\big(D_i = I_{\beta(i)} \cap X^{-1}(B_{\beta(i+1)})\big) > 0.$$

Cette suite prend ses valeurs dans $\{1, \cdots, n\}$ où n est le cardinal du support de Q. On en extrait une suite cyclique finie, que nous notons de la même manière, $\beta(i)$, avec $i \in \{1, \cdots, k\}$ et $k \leq n$, telle que

$$\lambda(D_i = I_{\beta(i)} \cap X^{-1}\big(B_{\beta(i+1)})\big) > 0, \quad i \in \{1, \cdots, k\} \text{ et } \beta(k+1) = \beta(1).$$

Il existe alors un ensemble de suites cycliques $(t_{k+1} = t_1, \dots, t_k)$, avec $t_k \in D_k$, de mesure $\lambda^{\otimes k}$ positive. Par conséquent, avec l'assertion (ii), en prenant $i+1$ à la place de $i-1$ et $\beta(k+1) = \beta(1)$, on a :

$$\sum_1^k \gamma\big(X(t_i), Y(t_i), Y(t_{i+1})\big) = \sum_1^k \gamma\big(X(t_i), a_{\beta(i)}, a_{\beta(i+1)}\big) \geq 0.$$

En notant $x_i = X(t_i)$, on en déduit qu'il existe un ensemble $\mathcal{C}$, de suites cycliques $(x_{k+1} = x_1, \dots, x_k)$, où $x_i \in B_{\beta(i+1)}$, de longueurs k, et de probabilité $P^{\otimes k}$ strictement positive, tel que

$$\sum_1^k \gamma\big(x_i, a_{\beta(i)}, a_{\beta(i+1)}\big) \geq 0. \tag{4}$$

Or le théorème 2.2 assure que f est c-*c.m.* P-*p.s.*, d'où

$$\sum_1^k \gamma\big(x_i, f(x_i), f(x_{i-1})\big) \geq 0 \quad p.s.$$

Choisissons la suite (x_i) dans $\mathcal{C}$ telle que $x_0 = x_k$, $f(x_i) = a_{\beta(i+1)}$. Par conséquent

$$\sum_1^k \gamma\big(x_i, a_{\beta(i+1)}, a_{\beta(i)}\big) \geq 0. \tag{5}$$

En rassemblant (4) et (5), et comme $\gamma(x,a,b) = -\gamma(x,b,a)$, nous avons, $P^{\otimes k}$-*p.s.*, sur $\mathcal{C}$

$$\sum_1^k \gamma\big(x_i, a_{\beta(i)}, a_{\beta(i+1)}\big) = 0. \tag{6}$$

Désignons par S_k l'ensemble des k-uples de M vérifiant la relation (6). L'ensemble $\mathcal{C}$ est P-*p.s.* inclus dans S_k.

Montrons que S_k est négligeable. La probabilité $P^{\otimes k}(S_k)$ se calcule par Fubini :

$$P^{\otimes k}(S_k) = \int P\Big(x \mid \gamma(x, a_{\beta(2)}, a_{\beta(1)}) = \sum_2^k \gamma\big(x_i, a_{\beta(i)}, a_{\beta(i+1)}\big)\Big)\, \mathrm{d}P^{\otimes k-1}(x_2, \dots, x_k).$$

L'hypothèse P est c-continue permet de conclure :

$$P^{\otimes k}(S_k) = 0 = P^{\otimes k}(\mathcal{C}).$$

C'est la contradiction recherchée.

Enfin, l'implication (i) $\Rightarrow$ (ii) est une conséquence du lemme 2.1. □

Retour au théorème : prouvons l'assertion (b). Si f est optimale pour le couple (P,Q), le lemme 2.1 assure qu'elle est c-$c.m.$ P-$p.s.$

Voyons la réciproque. D'après le lemme 2.4, le seul cas restant à faire est celui où le support de $Q = \sum_k \alpha_k \delta_{a_k}$, les a_k étant distincts, est de cardinal infini. Soit f une fonction c-$c.m.$ P-$p.s.$ de loi Q. On peut écrire

$$f = \sum x_i 1_{F_i}$$

où les x_i sont des points de $\{a_k\ ;\ k \in \mathbb{N}\}$ (non nécessairement distincts les uns des autres) et les F_i disjoints. Alors $\alpha_k = P(\bigcup_{i \in I_k} F_i)$ où $i \in I_k$ si et seulement si $x_i = a_k$. Pour $k \in \mathbb{N}$, notons P_k la probabilité trace de P sur $A_k = f^{-1}(\{a_1, \cdots, a_k\})$. Alors f est clairement c-$c.m.$ P_k-$p.s.$ et sa loi par rapport à P_k est la trace de Q sur $\bigcup_{i \le k} a_i$, soit

$$Q_k = \frac{1}{P(A_k)} \sum_{i=1}^{k} \alpha_i \delta_{a_i}.$$

Le lemme 2.4 assure que f est optimale pour (P_k, Q_k) et, nous l'avons vu au passage, elle est unique pour ce couple.

Montrons que

$$\lim_{k \to \infty} d_c(P_k, Q_k) = d_c(P,Q).$$

Comme

$$d_c(P_k, Q_k) = \int c\big(x, f(x)\big)\, \mathrm{d}P_k(x) = \frac{1}{P(A_k)} \int_{A_k} c\big(x, f(x)\big)\, \mathrm{d}P(x),$$

on en déduit que

$$d_c(P_k, Q_k) \longrightarrow \int c\big(x, f(x)\big)\, \mathrm{d}P(x) \ge d_c(P,Q).$$

Inversement, le lemme 2.2 permet d'écrire $d_c(P,Q) = \int c(x, g(x))\, \mathrm{d}P(x)$ pour une fonction g c-$c.m.$ P-$p.s.$ et de loi Q par rapport à P. Quitte à choisir une partition plus fine pour f et g, nous pouvons écrire

$$g = \sum_i b_i 1_{F_i}, \quad b_i \in \{a_k\ ;\ k \in \mathbb{N}\}.$$

Or g a même loi que f par rapport à P_k ; par suite (nous l'avons vu dans la preuve du lemme 2.4), elles sont égales P_k-$p.s.$ et, par conséquent,

$$d_c(P,Q) \ge P(A_k) \int c\big(x, g(x)\big)\, \mathrm{d}P_k(x) = P(A_k) d_c(P_k, Q_k).$$

En passant à la limite, nous obtenons la convergence cherchée.

Comme

$$d_c(P_k, Q_k) = \frac{1}{P(A_k)} \int_{X^{-1}(A_k)} c\big(X, f(X)\big)\,\mathrm{d}\lambda,$$

où X est une variable de loi P, le passage à la limite montre que $d_c(P, Q) = \int c(X, f(X))\,\mathrm{d}\lambda$. La fonction f est optimale pour le couple (P, Q) et son unicité P-*p.s.* est évidente.

Achevons la preuve des autres assertions.

• L'assertion (a) est une conséquence du lemme 2.2 et de la partie précédente.

• L'équivalence entre (c) 1) et (c) 3) résulte du même lemme et des parties (a) et (b) précédentes.

• Le lemme 2.1 montre l'implication (c) 1) $\Rightarrow$ (c) 2).

• Reste à établir que (c) 2) $\Rightarrow$ (c) 3). Soit $(X, Y) \in \mathcal{M}(P, Q)$ un couple c-*c.m.* λ-*p.s.* Il existe une suite croissante de compacts, $K_n \subset [0, 1]$, telle que $\lambda(K_n) \to 1, (X, Y)(K_n) \in \mathcal{B}(M \times M)$ et la loi Q_n de Y par rapport à λ_n, trace de λ sur K_n, a un support de cardinal inférieur à n.

Notons P_n la loi de X par rapport à λ_n. La probabilité P_n est c-continue car elle est absolument continue par rapport à P. De plus, pour λ_n, le couple (X, Y) appartient à $\mathcal{M}(P_n, Q_n)$ et est c-*c.m.* λ_n-*p.s.* Il est donc optimal par le lemme 2.4.

Avec le lemme 2.4 et la partie précédente, il existe une fonction f_n, c-*c.m.* P_n-*p.s.* telle que $Y = f_n(X)$, λ_n-*p.s.* Autrement dit, on a $Y 1_{K_n} = f_n(X) 1_{K_n}$, λ-*p.s.* En notant $A_n = X(K_n)$, on en déduit que, pour tout couple d'entiers (n, p), $f_{n+p} 1_{An} = f_n 1_{A_n} P_n$-*p.s.* Comme $P(A_n) \to 1$, la fonction f définie par $f(x) = f_n(x)$ où n est le plus petit entier tel que $x \in A_n$, vérifie $Y = f(X)$ λ-*p.s.* et elle est c-*c.m.* P-*p.s.* La fonction f vérifie l'assertion (b) du théorème 2.4, c'est l'unique fonction optimale et l'assertion (c) 3) est démontrée. $\square$

3. Cas de l'espace de Hilbert

Dans cette partie, le cadre est plus classique : nous prenons pour M un espace de Hilbert séparable $\mathbb{H}$.

Une probabilité P sur $\mathbb{H}$ vérifie la propriété (*) si, pour toute fonction convexe F de $\mathbb{H}$ dans $\overline{\mathbb{R}}$ telle que $P(F \in \mathbb{R}) = 1$, alors

$$P\big(x,\ \text{cardinal}\ \{\partial F(x)\} = 1\big) = 1.$$

Cette notion est liée aux c-c hypersurfaces de [18] et à la propriété de transport des probabilités sur $\mathbb{R}^n$ voir [3]. Nous montrons d'abord que les probabilités fortement continues, notion introduite dans [4], vérifient cette propriété. Puis nous « explicitons » la fonction optimale.

Dans la suite, nous supposons qu'il existe deux fonctions $f_1 \in L^1(P)$ et $f_2 \in L^1(Q)$ telles que

$$c(x,y) \leq f_1(x) + f_2(y) ;$$

par conséquent $d_c(P,Q) < +\infty$. Cette hypothèse permet de considérer le problème dual et, selon [12] et [15],

$$(3) \qquad d_c(P,Q) = \sup\Big\{ \int h_1 \,\mathrm{d}P + \int h_2 \,\mathrm{d}Q \,;\ h_1(x) + h_2(y) \leq c(x,y) \Big\}.$$

Nous pouvons alors utiliser les théorèmes de Rachev et de Rüschendorf [14] et [17] : « *Pour tout couple c-optimal* (X,Y), *il existe une fonction* F *c-convexe telle que* $Y \in \partial_c F(X)$ *p.s.* ».

Précisons quelques définitions concernant les fonctions c-convexes.

- Une fonction F est dite *c-convexe*, s'il existe un ensemble d'indices I tel que

$$F(x) = \sup_{i\in I}\big(c(x,a_i) + b_i\big), \quad \text{où } a_i \in \mathbb{H} \text{ et } b_i \in \mathbb{R}.$$

- Dans ce cas, le *c-sous-différentiel* de F au point x est l'ensemble

$$\partial_c F(x) = \big\{ y \in \mathbb{H} \,;\ c(z,y) - c(x,y) \leq F(z) - F(x),\ \forall z \in \mathbb{H} \big\}.$$

- La *c-conjuguée* de F est la fonction définie par

$$F^*(y) = \sup_{\{x \mid F(x) < +\infty\}} \big(c(x,y) - F(x)\big).$$

Rappelons aussi que F et F^* vérifient l'inégalité

$$F(x) + F^*(y) \geq c(x,y),$$

avec égalité si et seulement si $y \in \partial_c F(x)$.

Lorsque $c(x,y) = \langle x,y\rangle$, on retrouve les notions classiques de l'analyse convexe. Une fonction c-convexe est alors une fonction convexe au sens habituel et le c-sous-différentiel se réduit au sous-différentiel.

Remarquons que dans le cas où la fonction $c(x,y)$ est convexe, une fonction c-convexe est convexe, ce qui permet de parler de son sous-différentiel ainsi que de son c-sous-différentiel. Pour simplifier, nous notons $\partial F(x)$ le sous-différentiel d'une fonction convexe F. Enfin, si $\partial F(x)$ est réduit à un singleton, on le note $\nabla F(x)$. Nous référons le lecteur à [8] et [16] pour une lecture détaillée.

Dans la suite nous supposons que

$$c(x,y) = \theta(x-y)$$

où θ est une fonction continue strictement convexe de $\mathbb{H}$ dans $\mathbb{R}^+$ et $\theta(0) = 0$.

Enfin, une probabilité définie sur $\mathbb{H}$ est *fortement continue*, s'il existe une base orthonormée (e_i) telle que si, V_i désigne le sous-espace vectoriel engendré par e_i, $V_i^\perp$ le sous-espace orthogonal à V_i et $P_{V_i^\perp}$ la probabilité image de P par la projection orthogonale sur l'espace $V_i^\perp$, alors pour $P_{V_i^\perp}$-presque tout $y \in V_i^\perp$, la loi conditionnelle $P_{i/y}(\,.\,) = n_i(y,\,.\,)$ sur V_i est continue.

Le lemme suivant montre que toute probabilité fortement continue vérifie la propriété (*).

LEMME 3.1. — *Soient une probabilité P fortement continue sur $\mathbb{H}$ et F une fonction convexe telle que $P\{x \mid \partial F(x) \neq \phi\} = 1$. Alors $P(x \mid \operatorname{card}\{\partial F(x)\} = 1) = 1$.*

Preuve. — Choisissons une base orthonormée (e_i) assurant la propriété « P fortement continue » et notons

$$f(x,e_i) = \lim_{\substack{\alpha\to 0\\ \alpha>0}} \frac{1}{\alpha}\big(F(x+\alpha e_i) - F(x)\big).$$

Pour tout i, on a $-f(x,-e_i) \le f(x,e_i)$. Notons aussi

$$B_i = \big\{x\,;\ -f(x,-e_i) < f(x,e_i)\big\}$$

et, pour $y \in V_i^\perp$, soit $B_{i,y}$ la section de B_i en y. La fonction $t \mapsto \phi(t) = F(y+te_i)$ est convexe, réelle; ses dérivées à droite ϕ'_d, et à gauche ϕ'_g, sont croissantes et $\phi'_d(t) \ge \phi'_g(t)$. Par conséquent, $B_{i,y}$ est dénombrable.

Par Fubini, $P(B_i) = \int_{V_i^\perp} n(y, B_{i,y})\,\mathrm{d}P_{V_i^\perp}(y)$ et, comme pour $P_{V_i^\perp}$ presque tout y, $n_i(y, .)$ est continue, on conclut que $P(B_i) = 0$.

Pour $x \notin \bigcup B_i$, vérifions que $\operatorname{card}(\partial F(x)) = 1$. Soit $y \in \partial F(x)$; pour $\alpha > 0$, les inégalités

$$\langle y, \alpha e_i\rangle + F(x) \leq F(x + \alpha e_i), \quad \langle y, -\alpha e_i\rangle + F(x) \leq F(x - \alpha e_i),$$

montrent que $-f(x, -e_i) \leq \langle y, e_i\rangle \leq f(x, e_i)$. Par conséquent $\langle y, e_i\rangle = f(x, e_i)$ pour tout i, ainsi $y = \sum_i f(x, e_i)e_i$. □

THÉORÈME 3.2. — *Soient P et Q deux probabilités définies sur $\mathbb{H}$, P vérifiant la propriété* (*). *Il existe alors une unique fonction f, c-cycliquement monotone P-p.s. optimale pour (P, Q). Cette fonction s'écrit :*

$$f(x) = x - (\nabla\theta)^{-1} \circ \nabla F(x), \quad F \text{ c-convexe.}$$

Réciproquement, si f est de loi Q et est c-cycliquement monotone P-p.s., alors elle est l'unique fonction optimale pour le couple (P, Q).

Preuve du théorème. — L'idée consiste à utiliser le lemme précédent et la stricte convexité de la fonction θ pour établir que $\partial_c F(X)$ est réduit λ-*p.s.* à un singleton.

Soit $(X, Y) \in \operatorname{opt}_c(P, Q)$; d'après le théorème 2 de [17], il existe une fonction F, c-convexe telle $Y \in \partial_c F(X)\lambda$-*p.s.* Par conséquent, $\partial_c F(x)$ est P-*p.s.* non vide. Donc

$$F(y) - F(x) \geq \theta(y - a_1) - \theta(x - a_1), \quad \forall\, y \in \mathbb{H}. \tag{6}$$

La fonction $\theta(\cdot - a_1)$ est une fonction continue, strictement convexe, donc admet un sous-différentiel en tout point. Il existe ainsi un élément $b \in \mathbb{H}$ tel que

$$\theta(y - a_1) - \theta(x - a_1) \geq \langle b, y - x\rangle, \quad \forall\, y \in \mathbb{H}. \tag{7}$$

Des deux inégalités (6) et (7), on déduit que b est un sous-différentiel de F au point x. Et, avec l'hypothèse sur P, $b = \nabla F(x)$, P-*p.s.*

Par suite, si a_1 et a_2 sont deux éléments $\partial_c F(x)$, alors les deux fonctions $\theta(\cdot - a_1)$ et $\theta(\cdot - a_2)$ admettent P-*p.s.* le même sous-différentiel. On en déduit que $\theta(x - a_1) - \theta(x - a_2)$ est constant P-*p.s.* La stricte convexité de θ permet alors de conclure que $a_1 = a_2$.

Ainsi, $\partial_c F(x)$ est réduit P-*p.s.* à un singleton, d'où l'existence d'une fonction f optimale pour (P, Q).

Montrons l'unicité de la fonction optimale. Soit g une deuxième fonction optimale pour le couple (P, Q). Comme

$$\begin{aligned}\int \theta\big(X - g(X)\big)\,\mathrm{d}\lambda &= \int \theta\big(X - f(X)\big)\,\mathrm{d}\lambda \\ &= \int F(X) + F^*\big(f(X)\big)\,\mathrm{d}\lambda \\ &= \int F(X) + F^*\big(g(X)\big)\,\mathrm{d}\lambda,\end{aligned}$$

on en déduit que $\theta(X - g(X)) = F(X) + F^*(g(X))$ λ-*p.s.*, et par conséquent, $g = f$ P-*p.s.*

Pour obtenir l'expression de la fonction f, il suffit de combiner l'inégalité (6) et (7) et d'utiliser le lemme précédent pour obtenir l'égalité

$$\nabla\theta\big(x - f(x)\big) = \nabla F(x). \tag{8}$$

Comme θ est strictement convexe, $\nabla\theta$ est une bijection de $\mathbb{H}$ vers $\nabla\theta(\mathbb{H})$, d'où l'expression de la fonction optimale, ce qui termine la preuve du théorème.

On peut aussi remarquer que $y \in \partial\theta(x)$ si et seulement si $x \in \partial\theta^*(y)$, ce qui permet d'écrire $(\nabla\theta)^{-1} = \nabla\theta^*$. $\square$

Un raisonnement analogue au précédent, permet d'établir une version du théorème 3.2 pour les fonctions $c(x, y) = \theta(x - y)$ lorsque θ est une fonction strictement concave.

Remerciements. — Nous tenons à remercier le *referee* dont les remarques pertinentes ont permis d'améliorer notablement cet article.

BIBLIOGRAPHIE

[1] Abdellaoui (T.), Heinich (H.). — *Sur la distance de deux lois de probabilités dans le cas vectoriel*, C. R. Acad. Sci. Paris, t. **319**, série I, 1994, p. 981–984.

[2] Bickel (P.J.), Freedman (D.A.). — *Some asymptotic Theory for the Bootstrap*, Ann. Statis., t. **9**, 1981, p. 1196–1217.

[3] BELILI (N.), HEINICH (H.). — *Mass Transport Problem and Derivation*, à paraître dans Appli. Mathematicae, 1999.

[4] CUESTA-ALBERTOS (J.A.), MATRÁN (C.). — *Notes on the Wasserstein Metric in Hilbert Spaces*, Ann. Prob., t. **17**, 1989, p. 1264–1276.

[5]] CUESTA-ALBERTOS (J.A.), TUERO-DIAZ (A.). — *A Characterization for the Solution of the Monge-Kantorovich Mass Transference Problem*, Statist. Probab. Lett., t. **16**, 1993, p. 147–152.

[6] CUESTA-ALBERTOS (J.A.), MATRÁN (C.), RACHEV (S.T.), RÜSCHENDORF (L.). — *Mass Transportation Problems in Probability Theory*, App. Prob. Trust., t. **21**, 1996, p. 1–39.

[7] COHN (D.L.). — *Measure Theory*. — Birkhäuser, 1980.

[8] EKELAND (I.), TEMAM (R.). — *Analyse convexe et problèmes variationnels*. — Dunod, 1974.

[9] GANGBO (W.), MCCANN (R.J.). — *Optimal Maps in Monge's Mass Transport Problem*, C. R. Acad. Sci. Paris, t. **321**, Série I, 1995, p. 1653–1658.

[10] GANGBO (W.), MCCANN (R.J.). — *The Geometry of optimal Transportation*, Acta Math., t. **177**, 1996, p. 113–161.

[11] KNOTT (M.), SMITH (C.S.). — *Note on the optimal Transportation of Distributions*, J. Opt. Theor. Appl., t. **52**, 1987, p. 323–329.

[12] KELLERER (H.). — *Duality Theorems for Marginal Problems*, Z. Wahrsh. Verw. Gebiete, t. **67**, 1984, p. 399–432.

[13] MCCANN (R.J.). — *Existence and Uniqueness of monotone measure-preserving Maps*, Duke. Math. J., t. **80**, 1995, p. 309–323.

[14] RACHEV (S.T.), RÜSCHENDORF (L.). — *A Characterization of random Variables with minimum L^2 Distance*, J. Multivariate Anal., t. **32**, 1990, p. 48–54.

[15] RACHEV (S.T.). — *Probability Metrics and the Stability of stochastic Models*. — Wiley, New York, 1991.

[16] ROCKAFELLAR (R.T.). — *Convex Analysis*. — Princeton University Press, 1972.

[17] RÜSCHENDORF (L.). — *Optimal Solutions of multivariate coupling Problems*, Appl. Mathematicae, t. **23**, 1995, p. 325–338.

[18] ZAJIČEK (L.). — *On the Differentiation of convex Functions in finite and infinite dimensional Spaces*, Czechoslovak Math. J., t. **29**, 1979, p. 340–348.