

École Normale Supérieure de Cachan  
Centre des Mathématiques et de leurs Applications

Mémoire de stage

Directeur de stage : Filippo SANTAMBROGIO

# Théorie générale du transport et applications

Cécile CARRÈRE, Didier LESEVRE, Paul PEGON



# Table des matières

<b>1</b>	<b>Introduction à la théorie du transport</b>	<b>9</b>
1.1	Introduction à la problématique du transport optimal . . . . .	9
1.1.1	Le problème de Gaspard MONGE (1781) . . . . .	9
1.1.2	Premières difficultés du problème de MONGE . . . . .	10
1.1.3	Un exemple un peu plus élaboré . . . . .	11
1.1.4	La généralisation du problème par Leonid KANTOROVICH (1940) . . . . .	11
1.1.5	Minimisation et calcul variationnel . . . . .	13
1.2	Relaxation du problème de MONGE . . . . .	14
1.3	Dualité de MONGE–KANTOROVICH . . . . .	15
1.3.1	Maximum au problème dual . . . . .	16
1.3.2	Potentiels de KANTOROVICH . . . . .	17
<b>2</b>	<b>Quelques coûts particuliers</b>	<b>18</b>
2.1	Le cas du transport discret . . . . .	18
2.1.1	Un exemple discret . . . . .	18
2.1.2	Le problème du transport dans le cas discret . . . . .	18
2.2	Cas d'un coût découpé $c(x, y) = f(x) + g(y)$ . . . . .	19
2.3	Cas d'un coût $c(x, y) = h(x - y)$ où $h$ est strictement convexe . . . . .	20
2.3.1	Existence de transports optimaux . . . . .	20
2.3.2	Unicité du transport optimal . . . . .	20
2.3.3	Le théorème de BRENIER pour $h(x) = \frac{1}{2}x^2$ . . . . .	21
2.4	Reformulation à l'aide de la <i>c-concavité</i> . . . . .	21
2.4.1	Cas où $c$ est continue . . . . .	21
2.4.2	Cas où $c$ est semi-continue inférieurement . . . . .	23
<b>3</b>	<b>Cas de la distance <math>c(x, y) = d(x, y)</math></b>	<b>25</b>
3.1	Premières observations . . . . .	25
3.1.1	Cas $c(x, y) = h(x - y)$ et $h$ convexe à valeurs réelles . . . . .	25
3.2	Retour sur le problème de MONGE . . . . .	26
3.3	Rayons de transport . . . . .	26
3.3.1	La dualité de MONGE – KANTOROVICH . . . . .	26
3.3.2	Les rayons de transport . . . . .	26
3.3.3	De la différentiabilité sur les rayons de transport . . . . .	27
3.3.4	Des rayons partitionnant le support . . . . .	27
3.4	Transports optimaux pour la distance . . . . .	27
3.4.1	La perte de l'unicité de l'optimum . . . . .	27
3.4.2	Le choix d'un transport optimal . . . . .	27
3.4.3	Une caractérisation de $O(\mu, \nu)$ . . . . .	28
3.4.4	Notions de $\lambda$ -convexité . . . . .	29
3.4.5	Conclusion . . . . .	30
3.5	Le cas du coût $c(x, y) = c_\infty(x, y) = \max_{\text{Supp}(\gamma)}  x - y $ . . . . .	30

3.5.1	Retour sur les cas des transports impossibles . . . . .	30
<b>4</b>	<b>Le problème de Beckmann</b>	<b>31</b>
4.1	Le problème de BECKMANN . . . . .	31
4.1.1	Le problème physique de BECKMANN . . . . .	31
4.1.2	Reformulation du problème de minimisation . . . . .	32
4.2	L'intervention des mesures vectorielles . . . . .	33
4.3	Dans l'intimité du problème dual . . . . .	33
4.3.1	Retour sur le problème de BECKMANN . . . . .	33
4.3.2	Liens avec les critères de MONGE et de KANTOROVICH . . . . .	34
4.3.3	L'intervention de $v_\gamma$ et de $\sigma_\gamma$ . . . . .	34
4.4	Généralisation du problème de BECKMANN . . . . .	35
4.5	Analyse de $\sigma_\gamma$ . . . . .	35
4.6	estimations $\mathcal{L}^p$ . . . . .	39
4.7	Le problème de BECKMANN en dimension 1 . . . . .	40
<b>5</b>	<b>Cadre topologique et métrique</b>	<b>41</b>
5.1	Distances de WASSERSTEIN . . . . .	41
5.2	Étude de la topologie . . . . .	43
5.3	Lien avec l'équation de conservation de la charge . . . . .	45
5.4	Quelques « rappels » . . . . .	50
5.4.1	Chemins et longueurs en espace métrique . . . . .	50
5.4.2	Géodésiques . . . . .	50
5.4.3	Espaces de longueurs . . . . .	50
5.4.4	Une caractérisation des géodésiques à vitesse constante . . . . .	51
5.5	Géodésiques et transport optimal . . . . .	51
5.5.1	Géodésiques et plans de transferts optimaux . . . . .	52
5.5.2	Géodésiques à vitesses constantes et plans de transferts optimaux . . . . .	52
5.6	Convexité géodésique et fonctionnelles . . . . .	54
5.6.1	Généralisation de la notion de convexité . . . . .	54
5.6.2	Quelques fonctionnelles centrales . . . . .	54
5.6.3	De la convexité des fonctionnelles . . . . .	54
5.6.4	Inégalité de BRUNN – MINKOWSKI . . . . .	56
5.6.5	Autre application . . . . .	57
<b>6</b>	<b>Application à des problèmes d'équations aux dérivées partielles</b>	<b>58</b>
6.1	BÉNAMOÛ – BRENIER . . . . .	58
6.1.1	Le problème de la minimisation de la longueur . . . . .	58
6.1.2	La formulation de BÉNAMOÛ – BRENIER du problème . . . . .	58
6.2	Un peu de dualité... . . . .	59
6.2.1	Reformulations du problème et simplification des critères . . . . .	59
6.2.2	L'algorithme de BRENIER . . . . .	60
6.3	Cours 7 . . . . .	62
6.4	Introduction aux flots gradients . . . . .	62
6.4.1	Formulation du problème . . . . .	62
6.4.2	Discrétisation temporelle . . . . .	62
6.5	Retour sur les flots gradient . . . . .	63
6.6	Un cas précis : l'équation de la chaleur . . . . .	64

<b>7</b>	<b>Annexes</b>	<b>69</b>
7.1	Théorie de la mesure . . . . .	69
7.1.1	Convergence faible . . . . .	69
7.1.2	Lemme d'ULAM . . . . .	69
7.1.3	Suites tendues . . . . .	69
7.2	Analyse fonctionnelle . . . . .	70
7.2.1	Semi-continuité inférieure . . . . .	70
7.2.2	Théorème d'ASCOLI . . . . .	70
7.2.3	Transformation de LEGENDRE . . . . .	70
7.2.4	Fonctions convexes . . . . .	71
7.2.5	Fonctions <i>c-concaves</i> . . . . .	72
7.2.6	Théorème de HAHN-BANACH . . . . .	74
7.2.7	Relaxation d'une fonction . . . . .	74
7.2.8	Espaces de SOBOLEV . . . . .	75
7.2.9	Théorème de FENCHEL – ROCKAFELLAR . . . . .	75
7.2.10	Théorème de RADEMACHER . . . . .	76
7.3	Le transport dans le cas discret . . . . .	76
7.3.1	Théorème de CHOQUET . . . . .	76
7.3.2	Théorème de BIRKHOFF . . . . .	77
7.4	Quelques mots sur la théorie des distributions . . . . .	78
<b>8</b>	<b>Références</b>	<b>80</b>



# Introduction

Bla bla usuel...

La première partie est consacrée à une présentation générale de la théorie du transport. Le sujet est introduit à travers le problème de la recherche d'un transport optimal historiquement posé par Gaspard MONGE qui semble, même dans des cas relativement simples, difficile à traiter voire même voué à l'échec. Puis nous présentons le point de vue de la théorie de la mesure, introduit deux siècles plus tard par Léonid KANTOROVICH, qui permet de traiter le problème avec beaucoup plus d'aisance en recherchant des mesures optimales plutôt que des applications. Il convient alors de se demander si la généralisation proposée par KANTOROVICH n'est pas trop éloignée du problème initial, et quels sont les liens entre les deux problèmes : nous verrons que le problème de KANTOROVICH est la relaxation de celui de MONGE, autrement dit qu'il est, dans un certain sens, le meilleur problème généralisant l'énoncé initial. Notamment, les coûts optimaux sont identiques. Suivra alors un résultat central en théorie du transport et d'une redoutable efficacité dans les applications qui suivront est la dualité de MONGE – KANTOROVICH, qui est la reformulation du problème de minimisation considéré en un problème de maximisation d'un critère bien plus simple à utiliser.

Une seconde partie est consacrée à quelques cas particuliers relativement aisés – en comparaison des cas que nous traiterons par la suite – à traiter. Ainsi le cas du transport dans le cas d'espaces discrets finis est totalement traité et les résultats obtenus sont conformes à ce que l'intuition et la tendance naturelle suggérerait. Le cas des coûts découplés ne dépendant pas du trajet mais seulement des points de départ et d'arrivée se révélera très amical et nous verrons que n'importe quelle solution est optimale, nous permettant ainsi d'ignorer les composantes découplées du coût par la suite, celles-ci n'influant pas sur la recherche puisque n'ajoutant aucune contrainte. Pour finir, le cas de coûts strictement convexes en la distance sera entièrement traité, avec une fin heureuse puisque il existe alors un unique transport optimal, dont nous connaissons la forme. Le cas particulier du carré de la distance se comporte très bien et l'application des résultats précédents constitue le théorème de BRENIER. Quelques exemples plus concrets, ayant ou non des solutions, seront également étudiés.

Le troisième partie est dédiée à un cas nettement moins évident, à savoir celui de la distance. Dans ce cas, les résultats précédents sur les coûts strictement convexes ne s'appliquent plus, et il faut faire un périple très différent de ceux qui ont précédé pour arriver à nos fins. On introduit pour cela la notion de  $c$ -monotonie cyclique, qui nous permettra d'obtenir quelques résultats dans le cas de coûts convexes en la distance. Le cas de la distance est très riche, et nous construisons par approximation des transports optimaux qui sont également optimaux pour les coûts  $c_2$  et  $c_\infty$ . Ce dernier coût sera l'occasion de la dernière étude de cette partie.

Le problème du MONGE – KANTOROVICH est fortement lié à un problème de minimisation de flot et de conservation de la masse, proposé par BECKMANN. On travaillera, dans la quatrième partie, le lien existant entre ces deux problèmes *a priori* peu liés, de manière à avancer dans le problème de BECKMANN à l'aide des outils et des résultats développés dans l'étude précédente. On prouvera notamment que ces deux problèmes sont essentiellement les mêmes à l'aide de la dualité de MONGE – KANTOROVICH.

La cinquième partie est l'occasion d'introduire naturellement les distances de WASSERSTEIN sur l'espace des mesures de probabilité, qui sont les coûts minimum de transport entre deux mesures. Après avoir prouvé qu'il s'agit bien de distances, il convient alors d'étudier les propriétés de la topologie induite par chacune d'elles. Les problèmes de géodésiques interviennent ici naturellement, et nous détaillons leurs liens avec les problèmes de transport optimal : nous verrons que les transports optimaux et les géodésiques à vitesses constantes sont très fortement

liés. Enfin, nous présenterons le problème de BÉNAMOU – BRENIER, qui n'est que la reformulation d'un problème de minimisation de longueur d'un chemin, autrement dit un problème de recherche de géodésiques. La théorie du transport permet, une fois encore, de simplifier grandement les critères étudiés et on arrive finalement à approcher une solution optimale grâce à l'algorithme de BRENIER.

La dernière partie est dédiée à l'étude de problèmes faisant intervenir de nombreux résultats, outils et points de vue déjà développés. Ainsi certains problèmes de type flot-gradient seront présentés et traités par des méthodes de type schéma d'EULER. Notamment, le cas particulier de l'équation de la chaleur est étudié en détails. Un second problème, celui de l'étude de l'équation de conservation de la charge, est également présenté et étudié dans le cadre des espaces de WASSERSTEIN.

Compte tenu de la grande richesse de la théorie et de la large base théorique sur laquelle elle s'appuie, qui est tout aussi spécialisée que variée, des annexes détaillées regroupent tous les ingrédients utiles pour que cette étude puisse être dégustée avec passion et sans encombres. Ainsi des résultats centraux d'analyse fonctionnelle, de théorie de la mesure, d'analyse convexe ou de théorie de l'intégration y sont consignés.

# Chapitre I

## Introduction à la théorie du transport

### 1.1 Introduction à la problématique du transport optimal

#### 1.1.1 Le problème de Gaspard Monge (1781)

Le problème du transport optimal est introduit par Gaspard MONGE en 1781, alors professeur à l'école du génie militaire de Mézières, dans son *Mémoire sur la théorie des déblais et de remblais*, alliant ainsi comme à son habitude un sens aigu du concret et des problèmes pratiques avec une grande capacité d'abstraction dans la formulation et l'étude de ces problèmes.



Le problème est de minimiser le coût d'un transport de sable d'une dune pour combler un fossé, ou encore d'un transport de pierres d'une excavation pour construire une fortification. On peut modéliser la distribution de sable sur la dune par une mesure  $\mu$  et la distribution que le sable doit épouser dans le fossé par une mesure  $\nu$ , qui ont éventuellement des densités respectives  $f$  et  $g$ . Le transport est modélisé par une application  $T$  qui décrit la position finale  $T(x)$  du sable qui se trouvait originellement en  $x$ , et un *transport optimal* correspond à la minimisation du coût de ce transport, *i.e.* de la fonctionnelle intégrale définie par :

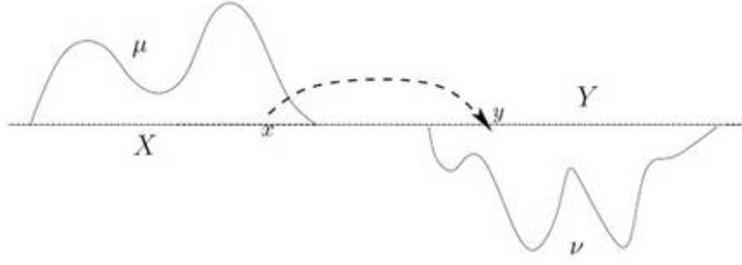


FIGURE 1.1 – Modélisation du problème de transport de MONGE

$$I : (T : \Omega \longrightarrow \Omega) \longmapsto \int_{\Omega} |T(x) - x| f(x) dx = \int_X |T(x) - x| d\mu \text{ où } T\#\mu = g$$

Un transport implique la conservation de la masse, donc les masses totales des mesures  $\mu$  et  $\nu$  doivent nécessairement être égales : on se limite donc, quitte à normaliser, à ne considérer que des mesures de probabilités. Le problème de MONGE se généralise en un problème de minimisation plus général, avec un coût de transport pouvant être différent de la simple distance. Le problème revient alors à minimiser le critère

$$(\mathcal{M}) \quad T \longmapsto \int_{\Omega} c(x, T(x)) d\mu$$

où la fonction  $T$  est une application de transport :  $T\#\mu = \nu$ .

### 1.1.2 Premières difficultés du problème de Monge

Le premier problème rencontré est celui de l'existence du minimum, *i.e.* d'un transport effectivement optimal. Considérons un exemple simple pour constater le comportement peu amical du problème : on suppose  $T$  injective et suffisamment régulière, et on suppose que  $\mu$  et  $\nu$  ont des densités respectives  $f$  et  $g$  :

$$\begin{aligned} \forall \phi : Y \longrightarrow \mathbb{R}^n, \quad \int \phi(y) g(y) dy &= \int \phi d\nu = \int \phi d(T\#\mu) = \int (\phi \circ T) d\mu \\ &= \int \phi(T(x)) f(x) dx \underset{y=T(x)}{=} \int \frac{\phi(y) f(T^{-1}(y))}{\det(\mathcal{J}T(T^{-1}(y)))} dy \end{aligned}$$

Ce qui donne alors  $\det(\mathcal{J}T) = \frac{f}{g \circ T}$ . On considère alors une suite  $(T_n)_n$  minimisant le critère intégral, *i.e.* tel que  $I(T_n)$  converge vers l'infimum, et on aimerait réussir à en extraire une sous-suite convergente dans une topologie pour laquelle  $I$  est continue, de manière à conclure à l'existence d'un minimum. Le problème vient du fait que la condition obtenue sur  $T$  n'est pas linéaire, et ne présente donc pas les bonnes propriétés que l'on aurait pu espérer. En effet, la fonctionnelle intéressante à étudier est

$$I_{\phi}(T_n) = \int (\phi \circ T_n) d\mu$$

Or rien ne nous permet de passer ici à la limite, donc de conclure. On aimerait que  $\phi(T_n)$  converge : il n'y a en général pas de convergence forte, on peut donc penser à se limiter à des fonctions  $\phi$  continues, auquel cas le théorème de convergence dominée s'appliquerait, ce qui revient à travailler sur la topologie engendrée par les fonctions continues. Or cette topologie est trop fine, et on perdrait alors les propriétés de compacité que l'on désire pour conclure et avancer par la suite : la topologie intéressante est la topologie faible, engendrée par les formes linéaires continues.

On constate lors de l'étude de ce cas d'école que deux problèmes viennent entraver nos espoirs de résolution. Tout d'abord, la condition obtenue sur  $T$  n'est aucunement linéaire, le comportement des solutions n'est donc pas aussi simple.

Mais surtout le problème du transport peut ne pas avoir de solution : outre le cas trivial où les masses totales des deux mesures sont différentes et auquel cas la propriété, tant physique que mathématique, de conservation de la masse n'est pas vérifiée, l'existence d'atomes pour la masse  $\mu$  entraîne l'existence d'atomes pour la masse  $\nu$ , car en effet :

$$\forall a \in X, T\#\delta_a = \delta_{T(a)}$$

Si la mesure  $\nu$  ne possède pas d'atome, le problème est donc sans solution. Le problème de MONGE, après son énoncé et les quelques réflexions menées par ce dernier, est resté près de deux siècles sous la poussière, plus par manque d'outils et de modélisation fiable que d'utilité et d'applications...

### 1.1.3 Un exemple un peu plus élaboré

On sait que l'ensemble des transports est dense dans l'espace des mesures de  $\pi(\mu, \nu)$  (dire qu'on le voit après, ou le faire avant...), donc que l'on peut approcher tout plan de transfert par des transports, et donc même s'il n'y a pas de transport on a :

$$\inf_T \int |x - T(x)| d\mu = \min_\gamma \int |x - y| d\gamma$$

Considérons par exemple

$$\mu = \mathcal{H}_{S_0}^1 \text{ et } \nu = \frac{1}{2}\mathcal{H}_{S_{-1}}^1 + \frac{1}{2}\mathcal{H}_{S_1}^1$$

Dans ce cas, on a pour tout  $x$  et  $y$  dans les supports respectifs de  $\mu$  et  $\nu$   $|x - y| \geq 1$  et  $\min \int |x - y| \geq 1$ . Le plan de transfert partageant chaque unité de masse entre les deux segments de manière égale convient et prouve que le minimum vaut bien 1.

Cependant on ne peut pas trouver de transport optimal dans ce cas. En effet, on aurait nécessairement

$$\gamma = \frac{1}{2}(id \times f_+) \#\mu + \frac{1}{2}(id \times f_-) \#\mu$$

où  $f_+ = id + e$  et  $f_- = id - e$  avec  $e = (1, 0)$  sont les densités respectives de  $\mu$  et  $\nu$ . Si  $T$  était un transport optimal, on aurait nécessairement  $|x - y| = 1$  sur le support de  $\gamma$  car  $|x - y|$  est continue et toujours supérieure à 1 et son intégrale vaut 1. Donc pour  $y = T(x)$  dans le support de  $\gamma$ ,  $y$  vaut nécessairement  $x + e$  ou  $x - e$ . On considère alors les ensembles  $A = \{x \mid T(x) = x + e\}$  et  $A^c = \{x \mid T(x) = x - e\}$ , et on a alors

$$T\#\mu = \mathcal{H}_{|A+e}^1 + \mathcal{H}_{|A^c-e}^1 \neq \nu$$

Malgré tout, on peut approcher un transport « optimal » :

$$1 \leq |T_n(x) - x| \leq \sqrt{1 + \frac{1}{n^2}} \xrightarrow{n \rightarrow \infty} 1$$

(+ solution du problème par APPELL au début du siècle ? cf. mémoire académie des sciences)

### 1.1.4 La généralisation du problème par Leonid Kantorovich (1940)

Ce n'est que dans les années 1940 que le mathématicien et économiste russe Leonid KANTOROVICH donnera à la théorie du transport un nouveau souffle de vie, en élargissant le très contraignant problème de MONGE : au lieu de rechercher à minimiser le coût du transport selon l'application de transport  $T$ , c'est vers les mesures que se tourne le regard de KANTOROVICH, ce qui lui vaudra d'ailleurs quelques années plus tard le prix NOBEL d'économie, récompensant ses « contributions à la théorie de l'allocation des ressources ». Le transport optimal est en effet bien adapté aux problématiques économiques et financières,  $\mu$  représentant par exemple une répartition de producteurs et  $\nu$  une répartition de consommateurs.



FIGURE 1.2 – Léonid KANTOROVICH (1912 – 1986)

Intuitivement, la masse présente initialement en  $x$  doit correspondre à la somme des masses « partant » de  $x$  lors du transfert, de même que la masse finale en  $y$  doit correspondre à la somme des masses « arrivant » en  $y$ , ce qui s'écrit :

$$\mu(x) = \int_Y d\gamma(x, y), \quad \nu(y) = \int_X d\gamma(x, y)$$

Ce qui correspond à une condition sur les marginales :  $\pi_1\#\gamma = \mu$ ,  $\pi_2\#\gamma = \nu$ . On se limite donc désormais à travailler sur des mesures dont les marginales sont les distributions de masses au départ et à l'arrivée :

$$\pi(\mu, \nu) := \{\gamma \in \mathcal{P}(\Omega \times \Omega) \mid \pi_1\#\gamma = \mu, \pi_2\#\gamma = \nu\}$$

On impose donc plus précisément que pour tout sous-ensembles  $A$  et  $B$  de  $\Omega$  :

$$\gamma(A \times \Omega) = \mu(A) \text{ et } \gamma(\Omega \times B) = \nu(B)$$

Le lien entre fonction de transport et plan de transfert fait l'objet de la proposition suivante.

**Proposition** Un transport  $T : X \rightarrow Y = X$  borélien induit un plan de transfert  $\gamma_T = (\text{Id} \times T)\#\mu$  concentré sur le graphe de  $T$  qui est  $\gamma_T$ -mesurable (*i.e.* est dans la tribu borélienne  $\gamma_T$ -complétée). Et réciproquement un plan de transfert  $\gamma$  concentré sur un graphe  $\gamma$ -mesurable  $\Gamma$  est induit par un transport borélien.

**Preuve** • Dans le cas où  $T$  est un transport répondant au problème, avec  $T\#\mu = \nu$ ,  $\gamma_T$  concentre bien toute la masse sur le graphe de  $T$ , et les conditions sur les marginales sont vérifiées :

$$\gamma_T = (\text{id} \times T)\#\mu \text{ qui convient bien car } \pi_1\#\gamma_T = \mu, \pi_2\#\gamma_T = T\#\mu = \nu$$

On suppose dans un premier temps que  $X$  est compact. Le théorème de LUSIN permet de trouver une suite croissante  $K_n$  de compacts de  $X$  tels que  $G_T = \text{Id} \times T$  est continu sur  $K_n$  et  $\mu(X \setminus K_n) < \frac{1}{n}$ . Alors les  $G_T(K_n)$  sont compacts donc  $\bigcup G_T(K_n)$  est mesurable et  $\Gamma \setminus \bigcup G_T(K_n) \subset (X \times X) \setminus \bigcup G_T(K_n)$  borélien de mesure nulle. Donc  $\Gamma$  est  $\gamma_T$ -mesurable. Si  $X$  n'est pas compact mais  $\sigma$ -compact (par exemple si c'est une partie de  $\mathbb{R}^n$ ), on applique le résultat aux graphes sur les compacts, la réunion étant encore  $\gamma_T$ -mesurable.

• Si  $\Gamma$  est un graphe  $\gamma$ -mesurable, il existe une fonction  $\phi : \pi_1(\Gamma) \rightarrow X$  telle que

$$\Gamma = \{(x, \phi(x)) \mid x \in \pi_1(\Gamma)\}$$

$\gamma$  étant finie, par régularité il existe une suite croissante de compacts  $K_n \subset \Gamma$  tels que  $\gamma(\Gamma \setminus K_n) \rightarrow 0$ . On a

$$\mu(\pi_1(K_n)) = \gamma(\pi_1^{-1}(\pi_1(K_n))) \geq \gamma(K_n) \rightarrow 1$$

Or  $\bigcup_n \pi_1(K_n) \subset \pi_1(\Gamma) \subset X$  avec  $\bigcup_n \pi_1(K_n)$  borélien donc  $\pi_1(\Gamma)$  est  $\mu$ -mesurable de masse 1 (*i.e.*  $\Gamma$  est le graphe d'une fonction définie sur  $\mu$ -presque tout  $X$ ). En désintégrant  $\gamma$  par rapport à  $\pi_1$ , c'est-à-dire  $\gamma = \gamma_x \otimes \mu$  où  $\gamma_x$  est une mesure sur  $Y (= X)$ .

En admettant que  $\gamma_x = \delta_{\phi(x)}$  pour presque tout  $x$  (WHY? Ambrosio fait appel à une intégrale extérieure, qu'est-ce?), on identifie  $\phi$  à un prolongement sur tout  $X$  (on l'a prolongée sur un  $\mu$ -négligeable) et on a :

$$\gamma(A \times B) = \int_A \gamma_x(B) d\mu(x) = \int_A \mathbf{1}_{\phi(x) \in B} d\mu(x) = \mu(\{x : (x, \phi(x)) \in A \times B\}) = (\text{Id} \times \phi)_\# \mu(A \times B)$$

Ce qui conclut  $\gamma = (\text{Id} \times \phi)_\# \mu$ .

L'ensemble  $\pi(\mu, \nu)$  dans lequel on recherche les solutions du problème est bien plus agréable que celui des fonctions  $T$  sur lequel on travaillait avec le problème de MONGE. Il en est une généralisation, ou plutôt une *relaxation* : on diminue les contraintes imposées pour avoir plus de liberté de mouvement. Ici, le transfert de la masse est relativement souple et soumis à la seule condition d'avoir pour marginales  $\mu$  et  $\nu$ . Le problème de MONGE impose de plus qu'il n'y ait pas de séparation d'une unité de masse en plusieurs morceaux pour le transfert (*mass splitting*), ce qui revient à dire que chaque masse en  $x$  est envoyée sur une unique position  $y$ .

Le plan  $\pi(\mu, \nu)$  n'est jamais vide car le produit tensoriel  $\mu \otimes \nu$  des deux mesures convient toujours, ce qui correspond intuitivement à un transport idiot où chaque unité de masse déplacée est répartie sur tout l'espace d'arrivée équitablement (*i.e.* on en met un peu plus là où il reste un peu plus à combler, et un peu moins là où il a peu à faire). Il est également *convexe*, ce qui donne une propriété de stabilité bien appréciable et permet d'utiliser les raisonnements et outils usuels en analyse convexe.

Nous arrivons ainsi à l'énoncé du problème central de la théorie du transport, baptisé à juste titre *problème de MONGE-KANTOROVICH*, qui modélise de manière très générale tous les problèmes d'optimisation du coût d'un déplacement. En introduisons la fonctionnelle intégrale :

$$\mathcal{MK} \pi(\mu, \nu) \ni \gamma \longmapsto \int_{\Omega \times \Omega} c(x, y) d\gamma$$

surgissent les deux questions centrales qui sont notre raison d'être :

- \* Existe-t-il un minimum  $\gamma$  ?
- \* Est-ce un transport, *i.e.* existe-t-il une fonction de transport telle que  $\gamma = \gamma_T$  ?

### 1.1.5 Minimisation et calcul variationnel

On cherche à appliquer la méthode usuelle en calcul variationnel : lorsque  $J : X \longrightarrow \mathbb{R}$  est une fonctionnelle, on cherche à décider de l'existence d'un minimum. Par définition, la limite inférieure de  $J(X)$  existe toujours, notons-la  $l$ , et il existe une suite  $(x_n)_n$  de  $X$  dont les images par  $J$  convergent vers  $l$ . Si cette suite admet une valeur d'adhérence, par exemple dans le cas d'un espace compact pour une certaine topologie, on peut supposer, quitte à extraire, que cette suite converge, disons vers  $x$ . Si les  $J(x_n)$  convergent vers  $J(x)$ , par exemple dans le cas où  $J$  est continue ou semi-continue inférieurement pour la topologie considérée,  $J(x)$  vaut nécessairement  $l$ , et le minimum existe bien.

**1<sup>er</sup> cas :  $\Omega$  est compact et  $c$  est continue sur  $\Omega \times \Omega$**

On considère dans notre cas une suite minimisante  $(\gamma_n)_n$  de  $\pi(\mu, \nu)$ . Ce sont des mesures de probabilité, elles sont donc uniformément bornées par 1, et de plus  $\Omega \times \Omega$  est compact, comme  $\Omega$ . La propriété de compacité faible pour les fonctions continues prouve que l'on peut extraire une sous-suite faiblement convergente, vers une certaine mesure  $\gamma$ . Puisque  $J : \gamma \longmapsto \int c(x, y) d\gamma$  est continue pour la topologie faible (par définition même),  $J(\gamma_n)$  converge vers  $J(\gamma)$  qui réalise donc le minimum.

**2<sup>e</sup> cas :  $\Omega$  est compact et  $c$  est semi-continue inférieurement et bornée inférieurement**

Dans ce cas, on ne peut plus affirmer la continuité de  $J$  comme précédemment, compte tenu du fait que  $c$  n'est pas supposée continue et que nous travaillons avec la topologie faible définie par les fonctions continues. La semi-continuité inférieure permet d'écrire  $c(x, y)$  comme limite croissante de fonctions continues  $c_n(x, y)$  sur  $\Omega$ . On pose alors  $J_n(\gamma)$  comme étant l'intégrale de  $c_n$  sur  $\Omega$ , qui croît par rapport à  $n$  par croissance de la suite de coûts  $c_n$ . La borne supérieure  $J$  des  $J_n$ , qui sont continues, est alors semi-continue inférieurement et bornée inférieurement par la caractérisation des telles fonctions donnée en annexe. Cela suffit donc à appliquer le raisonnement précédent, et donc à prouver l'existence d'un minimum.

**3<sup>e</sup> cas :  $\Omega$  est un espace polonais (i.e. métrique séparable et complet) et  $c$  est semi-continue inférieurement, bornée inférieurement, à valeurs dans  $\mathbb{R} \cup \{+\infty\}$**

Ne supposant plus  $\Omega$  compact, le résultat de compacité sur les mesures de probabilités ne s'applique plus. On prend une suite  $(\gamma_n)_n$  minimisante dans  $\pi(\mu, \nu)$ , et on cherche une extraction qui converge faiblement. Malheureusement, il n'y a plus de résultat de compacité permettant de conclure aisément ; pour cela, on cherche à montrer que la suite  $(\gamma_n)_n$  est tendue. La régularité des mesures de RADON prouve que pour  $\varepsilon > 0$ ,  $\mu$  et  $\nu$  étant finies, il existe deux compacts  $K$  et  $K'$  tels que :

$$\mu(K^c) \leq \frac{\varepsilon}{2}, \nu(K'^c) \leq \frac{\varepsilon}{2}$$

Pour toute mesure  $\gamma$  de  $\pi(\mu, \nu)$ , on constate que :

$$\begin{aligned} \gamma((K \times K')^c) &\leq \gamma(K^c \times \Omega) + \gamma(\Omega \times K'^c) \\ &= \mu(K^c) + \nu(K'^c) \leq \varepsilon \end{aligned}$$

La suite  $(\gamma_n)_n$  est tendue, donc on peut en extraire une sous-suite faiblement convergente et la démarche variationnelle précédente s'applique pour obtenir l'existence d'un minimum.

Ainsi, dans de nombreux cas usuels et relativement généraux, l'infimum du critère  $(MK)$  est un minimum.

## 1.2 Relaxation du problème de Monge

Il reste à se poser la question de savoir si KANTOROVICH a trop élargi le problème en cherchant à minimiser sur les mesures  $\gamma$  plutôt que sur les transports  $T$  comme MONGE l'avait fait. En effet, on sait que le minimum du problème de KANTOROVICH est inférieur à l'infimum du problème de MONGE (qui a l'inconvénient de pouvoir ne pas être un minimum).

On rappelle les fonctionnelles intégrales sur lesquelles porte le travail :

$$\begin{aligned} J(\gamma) &= \int_{\Omega} c d\gamma \\ J'(\gamma) &= \begin{cases} \int c(x, T(x)) d\gamma & \text{si } \gamma = \gamma_T \\ +\infty & \text{sinon} \end{cases} \end{aligned}$$

Le problème de MONGE correspond à la minimisation de  $J'$  :

$$\inf\{J'(\gamma) \mid \gamma \in \pi(\mu, \nu)\}$$

Le problème de KANTOROVICH correspond à la minimisation de  $J$  :

$$\inf\{J(\gamma) \mid \gamma \in \pi(\mu, \nu)\}$$

$J$  est un problème plus général, et on aimerait en faire une extension (une « relaxation ») de  $J'$  qui soit semi-continue inférieurement, mais qui reste proche de  $J$ . La relaxée  $H$  de  $J$  permet de régulariser un peu le problème, la question est alors de savoir si l'on a  $H = J$ , ce qui serait l'idéal.

Pour nous en rendre compte, notons  $H : \pi(\mu, \nu) \rightarrow \mathbb{R} \cup \{+\infty\}$  la relaxée en question. C'est une extension du problème de MONGE  $J'$ , ce qui signifie que  $H$  et  $J'$  sont confondues sur l'ensemble des mesures de transport  $\gamma_T$ . Puisque  $H$  est supposée semi-continue inférieurement, si la suite de mesures  $(\gamma_n)_n$  converge vers  $\gamma$ , alors on a  $H(\gamma) \leq \liminf J'(\gamma_n)$ . Si les  $\gamma_n$  sont ce plus supposées être des transports, i.e. des  $\gamma_{T_n}$ , alors  $J'(\gamma_n) = J(\gamma_n)$ . Dans ce cas, la continuité de  $J$  assure que les  $J'(\gamma_n)$  convergent vers  $J(\gamma)$ .

$H$  étant définie comme la plus grande fonction inférieure à  $J'$  et semi-continuité inférieurement,  $H(\gamma) \geq J(\gamma)$  d'où  $H = J$ .

Finalement, le problème de MONGE-KANTOROVITCH est la relaxation semi-continue inférieurement du problème initial de MONGE. Il permet d'élargir le problème dans un cadre bien plus général, celui des mesures de probabilité quelconques, tout en conservant la propriété intéressante

$$\text{Si } \gamma_n \xrightarrow{n \rightarrow \infty} \gamma \implies \liminf_{n \rightarrow \infty} H(\gamma_n) = J(\gamma)$$

### 1.3 Dualité de Monge–Kantorovich

La contrainte  $\gamma \in \pi(\mu, \nu)$  n'est pas très simple à manier, et nous ne savons pas caractériser de telles mesures de manière à simplifier l'écriture du problème. On cherche naturellement à se ramener à un problème de minimisation sur des mesures quelconques, ou vérifiant une conditions plus agréable à manipuler. Puisqu'il s'agit d'un problème de minimisation, une technique classique est de trouver une quantité  $\alpha(\gamma)$  qui n'altère pas le problème (i.e. qui vaille 0 sur  $\pi(\mu, \nu)$ ) et qui permette d'absorber les éléments ne vérifiant pas la condition d'appartenance à  $\pi(\mu, \nu)$  (i.e. valant une valeur suffisamment grande sur ces éléments pour que ce ne puisse être le minimum).

L'appartenance de  $\gamma$  à  $\pi(\mu, \nu)$  se traduit par l'égalité :

$$\alpha = \alpha(\gamma) = \sup_{\phi, \psi} A(\phi, \psi) = \sup_{\phi, \psi} \left[ \int \phi d\mu + \int \psi d\nu - \int (\phi(x) + \psi(y)) d\gamma \right] = 0$$

De plus, dès que  $A$  prend une valeur non nulle (i.e. dès que  $\gamma$  n'appartient pas à  $\pi(\mu, \nu)$ ), on peut concentrer toute la masse possible en une masse de DIRAC en ce point et l'accroître d'autant que l'on veut, ce qui prouve que  $\alpha$  est infini. Nous avons obtenu notre élément « absorbant » qui permet de généraliser la recherche à des mesures quelconques :

$$\alpha = \begin{cases} 0 & \text{si } \gamma \in \pi(\mu, \nu) \\ +\infty & \text{sinon} \end{cases}$$

En permutant les bornes supérieures et inférieures, nous obtenons le *problème dual* du premier :

$$\min_{\gamma} \left( \int c d\gamma + \alpha \right) \quad \text{dual de } \sup_{\phi, \psi} \left( \int \phi d\mu + \int \psi d\nu + \inf_{\gamma} \left( \int (c(x, y) - \phi(x) - \psi(y)) d\gamma \right) \right)$$

La question qui se pose est désormais de trouver un lien entre les problèmes primal et dual. Dans l'idéal, les deux problèmes seraient équivalents sous certaines conditions.

Si pour tous  $x$  et  $y$  on a  $c(x, y) - \phi(x) - \psi(y)$  qui est positif, alors  $\inf_{\gamma} = 0$ , atteint en prenant la mesure nulle sur l'espace. S'il existe  $x$  et  $y$  tels que  $c(x, y) - \phi(x) - \psi(y)$  soit strictement négative, alors on peut concentrer une masse arbitrairement grande en  $(x, y)$ , et  $\inf_{\gamma} = -\infty$ . Cela dit, le maximum recherché s'écrit alors :

$$\sup_{\phi, \psi} \left( \begin{cases} 0 & \text{si } \gamma \in \pi(\mu, \nu) \\ -\infty & \text{sinon} \end{cases} + \int \phi d\mu + \int \psi d\nu \right) = \sup_{\phi, \psi | \phi \oplus \psi \leq c} \left( \int \phi d\mu + \int \psi d\nu \right)$$

Nous introduisons alors un nouveau problème avec la minimisation par rapport à  $\gamma$  du critère intégral

$$\int c(x, y) d\gamma + \sup_{\phi, \psi} \left[ \int \phi d\mu + \int \psi d\nu - \int (\phi(x) + \psi(y)) d\gamma \right]$$

On tombe ainsi sur un problème dual du premier, échangeant un problème de minimisation par rapport à  $\gamma$  avec un problème de maximisation par rapport aux  $(\phi, \psi)$  vérifiant  $\forall x, y, \phi(x) + \psi(y) \leq c(x, y)$ . On note d'ailleurs que l'on a toujours

$$\min_{\gamma} \geq \sup_{\phi, \psi | \phi \oplus \psi \leq c}$$

Un théorème de FENCHEL – ROCKAFELLAR affirme que

$$\min_{\gamma} = \sup_{\phi, \psi | \phi \oplus \psi \leq c}$$

Autrement dit, le coût minimal correspond à la rentabilité maximale. En effet, on peut interpréter  $\int c d\gamma$  comme le coût du transport et  $\int \phi d\mu + \int \psi d\nu$  comme la rentabilité... (expliciter).

### 1.3.1 Maximum au problème dual

Le problème dual ainsi obtenu possède également de très bonnes propriétés, notamment si  $\Omega$  est compact et  $c$  est continue sur  $\Omega \times \Omega$ , alors il existe un couple  $(\phi, \psi)$  optimal, *i.e.* maximisant  $\alpha$ .

On commence par constater que si  $(\phi, \psi)$  est un couple admissible, *i.e.* vérifiant  $\phi \oplus \psi \leq c$ , alors

$$\forall x, y, \psi(y) \leq c(x, y) + \phi(x) \text{ donc } \forall y, \psi(y) \leq \inf_x (c(x, y) + \phi(x)) = \phi^c(y)$$

et le couple  $(\phi, \phi^c)$  est admissible, où  $\phi^c$  est la  $c$ -transformée de  $\phi$ .

vérifiant en particulier  $\forall x, x', y, y', |c(x, y) - c(x', y)| \leq \omega(|x' - x| + |y' - y|)$ . Il vient alors ici :

$$\forall x, y \in \Omega, \phi^c(y') \leq c(x, y') - \phi(x) \leq c(x, y) + \omega(|y' - y|) - \phi(x) = \phi^c(y) + \omega(|y' - y|)$$

Ainsi, le module de continuité de  $\phi^c$  est également  $\omega$ , ce qui quantifie la continuité de  $\phi^c$  à partir d'une quantification de celle de  $c$ . En itérant le procédé,  $(\phi^c, \phi^{cc})$  est un autre couple admissible du problème. On peut toujours remplacer un couple admissible par un autre couple dont les fonctions sont continues, de même module de continuité que  $c$ , et meilleures que le couple initial.

On prend alors une suite maximisante  $(\phi_n, \psi_n)_n$  de couples de fonctions continues de modules de continuités tous égaux à  $\omega$ . Quitte à translater les  $\phi_n$ , on peut supposer qu'elles sont nulles en un même  $x_0$ . On peut se limiter au cas  $\psi_n = \phi_n^c$  car  $\psi_n \leq \phi_n^c$ , donc conserver  $\psi_n$  n'apporte rien pour le calcul de la borne supérieure. Il vient alors

$$\forall n \in \mathbb{N}, \forall x, y \in \Omega, \begin{aligned} |\phi_n(x)| &\leq |\phi_n(x_0)| + \omega(|x - x_0|) \leq \omega(|x - x_0|) \leq 2\|c\|_{\infty} \\ |\psi_n(y)| &= |\phi_n^c(y)| \leq 2\|c\|_{\infty} \end{aligned}$$

Ainsi,  $(\phi_n, \psi_n)_n$  est uniformément bornée, donc le théorème d'ASCOLI s'applique, la translation des fonctions n'ayant pas changé leur propriété d'équicontinuité, et on peut extraire deux suites  $\phi_{\tau(n)}$  et  $\psi_{\tau(n)}$  convergeant uniformément respectivement vers  $\phi$  et  $\psi$ . Le théorème de convergence sous le symbole d'intégration s'applique par convergence uniforme, et par passage à la limite :

$$\forall x, y \in \Omega, \phi(x) + \psi(y) \leq c(x, y) \text{ et } (\phi, \psi) \text{ maximise donc } \alpha$$

Conclusion : Le *sup* est un *max*.

### 1.3.2 Potentiels de Kantorovich

On a déjà vu que

$$\min_{\gamma} \left( \int h(x-y) d\gamma \right) = \max_{\phi, \psi} \left( \int \phi d\mu + \int \psi d\nu \mid \phi(x) + \psi(y) \leq c(x, y) \right)$$

On considère alors un triplet  $(\gamma, \phi, \psi)$  d'éléments optimaux pour des critères, appelés *potentiels de KANTOROVICH*. Il vient alors par l'égalité qui précède :

$$\int c d\gamma = \int h(x-y) d\gamma = \int (\phi(x) + \psi(y)) d\gamma = \int \phi d\mu + \int \psi d\nu$$

On rappelle que le support d'une mesure  $\gamma$  est le plus petit fermé contenant de la masse pour cette mesure, *i.e.* le plus petit fermé  $F$  de  $X$  tel que  $\gamma(X \setminus F) = 0$ , *i.e.* l'ensemble des  $x$  tels que pour tout rayon  $r > 0$ , la boule de centre  $x$  et de rayon  $r$  est de masse non nulle, *i.e.* l'ensemble des points concentrés sur tout voisinage.

Or on a toujours  $c(x, y) \geq \phi(x) + \psi(y)$  et l'égalité des intégrales, donc nécessairement  $c = \phi \oplus \psi$   $\gamma$ -presque partout, et cette égalité est en particulier valide sur tout le support de  $\gamma$  puisque  $c, \phi, \psi$  sont continus. Soit  $(x_0, y_0)$  dans le support de  $\gamma$ , donc vérifiant l'égalité. On a :

$$\begin{aligned} \forall x \in \Omega, \quad \phi(x) + \psi(y_0) &\leq c(x, y_0) \\ \phi(x_0) + \psi(y_0) &= c(x_0, y_0) \end{aligned}$$

Ainsi,  $c(x, y_0) - \phi(x)$  minimale en  $x = x_0$ .

## Chapitre 2

# Quelques coûts particuliers

### 2.1 Le cas du transport discret

#### 2.1.1 Un exemple discret

Une approche discrète du problème est également possible :  $n$  usines viennent de produire une caisse de matériel chacune, et il faut les répartir entre  $n$  destinations. On recherche une permutation des caisses minimisant le coût total du transport défini par

$$c : \sigma \in \mathfrak{S}_n \mapsto \sum_i c(i, \sigma(i))$$

La fonction de coût a une grande importance et peut faire varier le résultat du tout au tout. On le voit par exemple en considérant une étagère sur laquelle sont rangés  $n$  livres de même épaisseur, et en étudiant le problème du décalage du bloc des  $n$  livres d'une épaisseur vers la droite. Deux solutions immédiates peuvent être envisagées :

- \* Mettre le livre le plus à gauche tout à droite ;
- \* Décaler chaque livre d'une épaisseur vers la droite, en commençant par la droite.

Si le coût est proportionnel à la distance, les deux solutions sont optimales. S'il est proportionnel au carré de la distance, seule la seconde est optimale : le coût du transport croissant très rapidement, les déplacements de grande ampleur sont à éviter. S'il est proportionnel à la racine carrée de la distance, seule la première est optimale : le coût est principalement dû à l'action du transport et non à la distance parcourue, donc quitte à transporter un livre, autant le transporter le plus loin possible. Les déplacements à privilégier dépendent donc fortement du coût envisagé, *i.e.* de ce que l'on considère comme coûteux.

#### 2.1.2 Le problème du transport dans le cas discret

On représente le cas discret du transport par des répartitions ponctuelles de masses :

$$\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \text{ et } \nu = \frac{1}{n} \sum_{j=1}^n \delta_{y_j}$$

On peut représenter un plan de transfert par une matrice précisant la répartition de la masse initialement présente en  $x_i$  entre les  $y_j$  :

$$\gamma = \sum_{i,j} \gamma_{ij} \delta_{(x_i, y_j)}$$

La matrice  $(\gamma_{ij})_{i,j}$  que l'on identifie à  $\gamma$  doit bien évidemment être bistochastique

$$\frac{1}{n} = \mu(\{x_i\}) = \sum_j \gamma_{ij} \quad \text{et} \quad \frac{1}{n} = \mu(\{y_j\}) = \sum_i \gamma_{ij}$$

Et réciproquement toute matrice bistochastique est un plan de transfert admissible. Le problème de la minimisation du coût se ramène donc à un problème de minimisation de

$$c(\gamma) = \sum_{i,j} \gamma_{i,j} c(x_i, y_j)$$

Les résultats de CHOQUET et de BIRKHOFF permettent de prouver que les minimiseurs sont exactement les matrices de permutation. Autrement dit, nous sommes en présence d'un cas où les plans de transfert optimaux sont tous des transports.

Cette étude permet d'obtenir les plans de transfert dans le cas du transport discret avec des masses *rationnelles*, ce qui modélise bien toutes les situations rencontrées en pratique. En effet, quitte à réduire tous les rationnels au même dénominateur et à normaliser, on peut supposer que les masses en chaque point sont toutes entières. Ceci dit, on se ramène au cas de masses uniformes en scindant un point de masse  $m$  en  $m$  points de masse 1 situés au même endroit. Le résultat précédent affirme que tous les plans de transfert optimaux sont obtenus par des permutations. Cependant, puisque nous avons scindé les masses avant de faire cela, nous ne pouvons plus espérer obtenir des transports, sauf dans des cas très particuliers : chaque masse de la distribution initiale est susceptible d'être partagée entre plusieurs destinations.

## 2.2 Cas d'un coût découplé $c(x, y) = f(x) + g(y)$

Ce cas correspond intuitivement à une situation où les efforts à fournir pour le transport se concentrent au départ et à l'arrivée. Comme vu précédemment, le problème de minimisation revient au problème de maximisation de

$$\int \phi d\mu + \int \psi d\nu$$

où  $\phi(x) + \psi(y) \leq c(x, y)$ . On peut se limiter à la recherche sur les couples  $(\phi^c, \phi^{cc})$ , comme signalé précédemment. Les expressions s'obtiennent alors facilement compte tenu du découplage des variables :

$$\begin{aligned} \forall x \in \Omega, \quad \phi^c(x) &= \inf_y (f(x) + g(y) - \psi(y)) = f(x) + \inf_y (g(y) - \psi(y)) \\ \forall y \in \Omega, \quad \phi^{cc}(x) &= \inf_x (f(x) + g(y) - \phi^c(x)) = \inf_x (f(x) - \phi^c(x)) + g(y) \end{aligned}$$

Les solutions sont donc de la forme

$$\begin{cases} \phi : x \mapsto f(x) + c \\ \psi : y \mapsto g(y) - c \end{cases}$$

où  $c$  est une constante. Or, avec ces notations, la quantité  $\int \phi d\mu + \int \psi d\nu$  ne dépend pas de la constante  $c$ , les mesures  $\mu$  et  $\nu$  étant des probabilités. Ainsi, tous les couples optimaux sont donc de cette forme. Nous connaissons ainsi la valeur du minimum du transport, et même si nous n'avons pas résolu le problème initial, à savoir trouver des mesures optimales, la connaissance de ce minimum permet d'espérer l'approximation (algorithmique par exemple) de l'optimum par des transports, même s'il n'y a pas de transport optimal dans ce cas !

Dans le cas de variables découplées, toutes les solutions conviennent donc, et le problème se ramène à la seule recherche d'une mesure qui est une mesure de transport sur l'espace. Ce résultat justifie les raisonnements ultérieurs, par exemple dans le cas  $(x-y)^2$ , de ne travailler que sur les termes couplés en  $x$  et  $y$ . Les termes découplés n'influent donc pas sur le problème d'optimisation, ils ne font que modifier la valeur des optima, et ce de manière entièrement déterminée par  $\mu$  et  $\nu$ .

1. Nous montrons plus loin que les mesures de transport sont denses dans l'espace des mesures, sous la seule condition que  $\mu$  n'ait pas d'atome, ce qui justifie la possibilité d'une approximation algorithmique

### 2.3 Cas d'un coût $c(x, y) = h(x - y)$ où $h$ est strictement convexe

Le problème de MONGE est un problème limite de ce cas-ci, la distance utilisée comme coût étant convexe mais non strictement convexe.

#### 2.3.1 Existence de transports optimaux

On se place dans le cas où  $h$  est différentiable, et donc où le gradient  $\nabla h$  est injectif puisque  $h$  est strictement convexe. Le gradient de la fonction s'annule donc en  $y_0$ , et on a  $\nabla h(x_0 - y_0) = \nabla \phi(x_0)$ , soit  $x_0 - y_0 = (\nabla h)^{-1}(\nabla \phi(x_0))$ , et on pose alors :

$$y_0 = T(x_0) = x_0 - (\nabla h)^{-1}(\nabla \phi(x_0)) = (id - (\nabla h)^{-1} \circ \nabla \phi)(x_0)$$

Le résultat fondamental est que pour un espace  $\Omega$  dont le bord est de mesure nulle, pour  $\mu$  et  $\nu$  deux mesures absolument continues par rapport à la mesure de Lebesgue  $\lambda$  sur  $\mathbb{R}^n$ , et pour un coût  $c(x, y) = h(x - y)$  où  $h$  est strictement convexe, il existe une unique mesure  $\gamma$  optimale, et c'est une mesure de transport :  $\gamma = \gamma_T$  où  $T : x \mapsto x - \nabla h^*(\nabla \phi(x))$ , définie presque partout.

Le choix d'un bord de mesure nulle permet de négliger la frontière de  $\Omega$ , donc de s'assurer de pouvoir prendre un point intérieur à  $\Omega$  en lequel la différentielle de  $T$  existe et sur lequel on peut affirmer que la minimisation de la fonction implique l'annulation de sa différentielle.

On prend  $\gamma$  optimal et  $(x_0, y_0)$  dans le support de  $\gamma$ , avec de plus  $x_0 \notin \partial\Omega$ . Les résultats précédents prouvent que :

$$y_0 = x_0 - \nabla h^*(\nabla \phi(x_0))$$

et  $\nabla \phi$  existe presque partout car  $\phi(x) = \inf_y h(x - y) - \psi(y)$  et  $h$  est localement lipschitzienne car strictement convexe, et puisque  $\phi$  a le même module de continuité de  $h$ , elle est localement lipschitzienne et, par le théorème de RADEMACHER,  $\phi$  est différentiable presque partout. Donc  $\gamma$  est concentré sur le graphe de  $T$  (puisque le bord de  $\Omega$  est  $\mu$ - et  $\nu$ -négligeable, donc le bord de  $\Omega \times \Omega$  est  $\gamma$ -négligeable) où  $T : x \mapsto x - \nabla h^*(\nabla \phi(x))$ . Donc il est induit par le transport  $T$  d'après une proposition vue précédemment.

#### 2.3.2 Unicité du transport optimal

Pour l'unicité, cela résulte de la convexité des contraintes,  $\gamma$  étant unique comme mesure et  $T$  étant unique presque-partout par ce qui précède. Si  $\gamma = \gamma_T$  et  $\gamma' = \gamma_{T'}$  sont deux mesures optimales, alors par convexité de  $\pi(\mu, \nu)$ ,  $\gamma'' = \frac{1}{2}(\gamma + \gamma')$  est également solution et, puisque  $(x_0, T(x_0))$  et  $(x_0, T'(x_0))$  sont dans le support de  $\gamma''$ , on a  $T(x_0) = T'(x_0)$   $\mu$ -presque partout. En effet, le support étant la réunion des supports, si  $T(x_0) \neq T'(x_0)$  sur un ensemble non négligeable, le support de  $\gamma''$  ne serait pas un graphe (la fibre de  $x_0$  contiendrait les deux points distincts  $T(x_0)$  et  $T'(x_0)$ ), alors qu'on a prouvé que c'en est nécessairement un.

Nous pouvons également énoncer une réciproque du théorème : si  $\phi$  et  $\psi$  sont deux fonctions mesurables telles que  $\phi = \phi^c : x \mapsto \inf_y (h(x - y) - \phi(y))$ , si  $T$  est définie par  $T : x \mapsto x - \nabla h^*(\nabla \phi(x))$ , et si  $\mu$  est absolument continue par rapport à la mesure de LEBESGUE  $\lambda$  sur  $\mathbb{R}^n$ , alors  $T$  est un transport optimal pour  $(\mu, T\#\mu)$ .

Pour le prouver, on commence par construire  $\gamma = \gamma_T$  comme usuellement. Il suffit de dire que  $\int c d\gamma_T = \int \phi d\mu + \int \psi d\nu$ , donc que  $h(x - y) = \phi(x) + \psi(y)$   $\gamma_T$ -presque partout, et que  $h(x - T(x)) = \phi(x) + \psi(T(x))$   $\mu$ -presque partout, car  $\gamma_T$  concentre toute la masse sur le graphe de  $T$ .

On introduit alors  $\psi(x_0) = \inf_y (h(x_0 - y) - \phi(y))$ , et l'égalité est vérifiée pour un  $\bar{y}$  optimal. Pour tout  $x$ , on a l'inégalité  $\phi(x) + \psi(\bar{y}) \leq h(x - \bar{y})$ , et l'égalité est vérifiée en  $x_0$ , d'où  $\bar{y} = x - \nabla h^*(\nabla \phi(x_0)) = T(x_0)$ , d'où  $\phi(x_0) + \psi(T(x_0)) = h(x_0 - T(x_0))$ . Ce sont donc des potentiels de KANTOROVICH, d'où l'optimalité de  $\gamma = \gamma_T$ .

### 2.3.3 Le théorème de Brenier pour $h(x) = \frac{1}{2}x^2$

On considère la fonction  $h : x \mapsto \frac{1}{2}x^2$ , de différentielle  $\nabla h = id = (\nabla h)^{-1}$ . Dans ce cas des résultats récents permettent de donner un transport optimal :

$$T : x \mapsto x - \nabla \phi(x) = \nabla \left( \frac{id^2}{2} - \phi \right)$$

où  $\phi(x) = \inf_y \frac{1}{2}|x - y|^2 - \psi(y)$ . La fonction  $\frac{id^2}{2} - \phi$  est convexe comme supremum de minorantes affines, qui sont convexes :

$$\frac{id^2}{2} - \phi = \dots = \inf()$$

Puisque la transformation de LEGENDRE est une involution sur les fonctions convexes, on tire que si on peut définir, pour une certaine fonction  $\chi$  convexe :

$$\begin{aligned} \forall x \in \Omega, \quad \phi(x) &= \frac{1}{2}x^2 - \chi(x) \\ \psi(y) &= \chi^*(y) + \frac{1}{2}|y|^2 \end{aligned}$$

On a alors un transport optimal pour  $(\mu, T\#\mu)$  :

$$T : x \mapsto x - \nabla h^*(\nabla \phi(x)) = \nabla \chi(x)$$

Donc dans le cas du coût quadratique, on a un transport optimal et on a sa forme explicite.

On aurait pu écrire le minimum de  $\int \frac{1}{2}|x - y|^2 d\gamma$  puis développer et remarquer que seule l'intégrale croisée fait intervenir réellement la mesure  $\gamma$ , donc est ramenée à minimiser cette seule intégrale (les autres ne dépendant que de  $\mu$  et  $\nu$ ). En fait, pour la fonction distance au carré, il n'est pas nécessaire d'avoir l'absolue continuité, qui impose de ne pas mettre de masse sur les négligeables, mais il suffit de ne pas mettre de masse aux points où il n'y a pas de différentiabilité.

$\forall \mu, \nu$  si  $\mu$  est absolument continue par rapport à la mesure de Lebesgue  $\lambda$  sur  $\mathbb{R}^n$ , il existe une unique fonction convexe  $\chi$  telle que  $\nabla \chi \# \mu = \nu$ , donc une unique fonction  $\chi$  telle que  $T = \nabla \chi$ .

## 2.4 Reformulation à l'aide de la *c-concavité*

On introduit les notions de *c-concavité* et *c-monotonie cyclique* en annexe à l'aide desquelles on va pouvoir revisiter de manière naturelle la formule de dualité et les résultats sur les coûts convexes.

### 2.4.1 Cas où $c$ est continue

**Théorème** Soit  $c : \Omega \times \Omega \rightarrow \mathbb{R}$  continue,  $\gamma$  optimisant le critère de KANTOROVICH  $\int c d\gamma \mid \gamma \in \Pi(\mu, \nu)$ . Alors le support de  $\gamma$  est *c*-cycliquement monotone.

**Démonstration** Présentons les choses intuitivement dans le cas discret d'un espace comprenant un nombre fini de points, au départ comme à l'arrivée. On suppose que ce sont tous des atomes, *i.e.* qu'ils ont tous une contribution significative au coût total du transport, que l'on peut écrire

$$\sum_{i=1}^n c(x_i, y_i)$$

en supposant que le transport se fait du point  $x_i$  au point  $y_i$ . Supposer ce plan de transfert optimal signifie qu'on ne peut améliorer le coût total du transport en réorganisant les points d'arrivée, *i.e.* qu'on ne peut trouver de permutation  $\sigma \in \mathfrak{S}$  telle que

$$\sum_{i=1}^n c(x_i, y_{\sigma(i)}) < \sum_{i=1}^n c(x_i, y_i)$$

Mais ici nous ne sommes pas dans un cas discret, cette inégalité sur quelques points n'impliquant *a priori* en rien un transport non optimal. Pour cela, il faut prouver que cette inégalité se transmet à un ensemble plus « grand », qui soit de mesure strictement positive, de manière à ce que l'inégalité puisse avoir un impact sur le coût total et prouver de ce fait la non-optimalité du transport. Dans le cas où  $c$  est continue et où les points sont dans le support, ce prolongement est naturel : l'inégalité est stricte et se prolonge donc aux voisinages des points, qui sont de mesures strictement positives puisque les points sont dans le support. En modifiant  $\gamma$  sur ces seuls voisinages, on obtient un meilleur plan de transfert, ce qui ne peut être si  $\gamma$  est optimal.

Explicitons cela : on veut montrer que pour tout ensemble fini de points  $(x_i, y_i)_{i \in [1, K]}$  dans le support de  $\gamma$  et toute permutation  $\sigma \in \mathfrak{S}$ , on a  $\sum_{i=1}^K c(x_i, y_i) \leq \sum_{i=1}^K c(x_i, y_{\sigma(i)})$ .

Par l'absurde, supposons qu'il existe de tels  $(x_i, y_i)_i$  tels que  $\sum_{i=1}^K c(x_i, y_i) > \sum_{i=1}^K c(x_i, y_{\sigma(i)})$ . Par continuité de  $c$ , il existerait un  $r$  tel que :

$$\forall x'_i \in B(x_i, r), \forall y'_i \in B(y_i, r), \sum_{i=1}^K c(x'_i, y'_i) > \sum_{i=1}^K c(x'_i, y'_{\sigma(i)})$$

On note  $V_i = \mathcal{B}(x_i, r) \times \mathcal{B}(y_i, r)$  et  $W_i = \mathcal{B}(x_i, r) \times \mathcal{B}(y_{\sigma(i)}, r)$  des voisinages des points considérés.

Sachant que, pour tout  $i$ ,  $\gamma(V_i) > 0$  par définition du support, on cherche à construire une « meilleure » mesure  $\delta$  en échangeant de la masse entre les  $V_i$  et les  $W_i$ , ce que l'on fait en raisonnant en termes de variables aléatoires. On sait en effet qu'il existe un espace mesuré  $Z$  par une mesure  $\lambda$  et  $g_i : Z \rightarrow V_i$  telles que  $g_i \# \lambda = \frac{\gamma(V_i)}{\gamma(V_i)}$  *i.e.*  $g_i$  est de loi  $\gamma$  normalisée sur  $V_i$ .

Par exemple  $Z = V_1 \times \dots \times V_K$ ,  $\lambda = \frac{\gamma(V_1)}{\gamma(V_1)} \otimes \dots \otimes \frac{\gamma(V_K)}{\gamma(V_K)}$  et  $g_i$  les projections canoniques.

On note  $g_i = (g_i^x, g_i^y)$ ,  $\alpha = \min(\gamma(V_i)_{i \in [1, K]})$  et  $h_i = (g_i^x, g_{\sigma(i)}^y)$ , puis on pose

$$\delta = \gamma - \sum_{i=1}^K \alpha g_i \# \lambda + \sum_{i=1}^K \alpha h_i \# \lambda$$

Il reste à vérifier que (i)  $\delta$  est admissible et (ii)  $\int c d\delta < \int c d\gamma$ .

(i)  $\delta$  est positive car on enlève de la masse sur les  $V_i$ , sur lesquels  $\delta = \gamma - \frac{\alpha}{\gamma(V_i)}\gamma$ . Or  $\alpha \leq \gamma(V_i)$ , ce qui garantit la positivité lorsque les  $V_i$  sont disjoints. Si les  $(x_i, y_i)$  sont distincts, c'est possible en prenant  $r$  assez petit. Si ce n'est pas le cas, pour garantir la positivité de la mesure, il faut éviter d'enlever trop de masse à  $\gamma$ , il suffit pour cela de changer  $\alpha$  en  $\frac{\alpha}{K}$ .

Reste à vérifier les marginales : TODO.

(ii) Il suffit de montrer

$$\begin{aligned} \sum_{i=1}^K \int c d h_i \# \lambda < \sum_{i=1}^K \int c d g_i \# \lambda &\iff \sum_{i=1}^K \int c d (g_i^x, g_{\sigma(i)}^y) d\lambda < \sum_{i=1}^K \int c d (g_i^x, g_i^y) d\lambda \\ &\iff \sum_{i=1}^K \int c(g_i^x, g_{\sigma(i)}^y) - c(g_i^x, g_i^y) d\lambda > 0 \end{aligned}$$

Ceci est vrai car  $g_i^x \in \mathcal{B}(x_i, r)$ ,  $g_i^y \in \mathcal{B}(y_i, r)$  et  $g_{\sigma(i)}^y \in \mathcal{B}(y_{\sigma(i)}, r)$  sont construits pour qu'on ait justement cette inégalité sur les coûts.

**Corollaire** Si  $\gamma$  est optimal pour un coût  $c$  continu, il existe une fonction  $\phi$   $c$ -concave et continue telle que

$$\text{Supp } \gamma \subseteq G_\phi^c = \{(x, y) | \phi(x) + \phi^c(y) = c(x, y)\}$$

**Conséquence** On vient donc de trouver  $\phi, \psi = \phi^c$  vérifiant  $\phi \oplus \psi \leq c$  (par définition de  $\phi^c$ ) tels que

$$\int c(x, y) d\gamma = \int [\phi(x) + \phi^c(y)] d\gamma = \int \phi d\mu + \int \psi d\nu$$

Comme  $\int c(x, y) d\gamma \leq \int \phi d\mu + \int \psi d\nu$  pour un  $\gamma$  optimal, on vient en fait de (re)montrer la formule de dualité

$$\min_\gamma \int c(x, y) d\gamma = \sup_{\phi, \psi \in C^0 | \phi \oplus \psi \leq c} \left( \int \phi d\mu + \int \psi d\nu \right)$$

De plus, le supremum est un maximum puisque  $(\phi, \phi^c)$  est minimal.

### 2.4.2 Cas où $c$ est semi-continue inférieurement

**Théorème** Si  $c$  est semi-continue inférieurement sur  $\Omega$  compact, la formule de dualité reste vraie :

$$\min_\gamma \int c(x, y) d\gamma = \sup_{\phi, \psi \in C^0 | \phi \oplus \psi \leq c} \left( \int \phi d\mu + \int \psi d\nu \right)$$

**Démonstration** Puisque  $c$  est semi-continue inférieurement, on sait qu'il existe une suite croissante  $(c_k)_k$  de coûts continus convergeant vers  $c$ . On a déjà

$$\min_\gamma \int c_k d\gamma = \max_{\phi \oplus \psi \leq c_k} \left\{ \int \phi d\nu + \int \psi d\mu \right\} \leq \sup_{\phi \oplus \psi \leq c} \left\{ \int \phi d\nu + \int \psi d\mu \right\}$$

Commençons par montrer un lemme :  $m_k = \min_\gamma \int c_k d\gamma \xrightarrow{k \rightarrow +\infty} \min_\gamma \int c d\gamma = m$ , ce qui donne  $\min_\gamma \int c d\gamma \leq \sup_{\phi \oplus \psi \leq c} \left\{ \int \phi d\nu + \int \psi d\mu \right\}$ , l'autre inégalité étant claire.

**Démonstration du lemme** On procède par double inégalité, un premier sens était clair :  $c_k \leq c$  donc  $\lim m_k \leq m$ .

D'autre part, prenons  $\gamma_k$  minimal pour  $c_k$ , et quitte à extraire ( $\gamma$  étant compacte, on a vu que  $\pi(\mu, \nu)$  est compact pour  $\xrightarrow{C^0(\Omega)}$ ),  $\gamma_k \rightarrow \gamma$ . Pour  $k \geq j$ ,  $\int c_k d\gamma_k \geq \int c_j d\gamma_k$ , puis en faisant tendre  $k \rightarrow +\infty$  :

$$\lim m_k \geq \int c_j d\gamma$$

Enfin en faisant tendre  $j \rightarrow +\infty$  on obtient le résultat :

$$\lim m_k \geq \int c d\gamma \geq m$$

**Théorème** Soient  $\Omega$  compact,  $c$  semi-continue inférieurement,  $\gamma$  optimal. Il existe  $\Gamma \subset \Omega \times \Omega$   $c$ -cycliquement monotone tel que  $\gamma(\Gamma) = 1$ .

**Démonstration** On par comme d'habitude de la dualité de MONGE — KANTOROVICH

$$\min_{\gamma} \int c d\gamma = \sup_{\phi, \psi} \int \phi d\mu + \int \psi d\nu$$

Soient alors  $(\phi_k)_k$  et  $(\psi_k)_k$  minimisantes, puis posons  $f_n(x, y) = c(x, y) - \phi_n(x) - \psi_n(y)$  qui est positive et  $\int f_n d\gamma \xrightarrow{n \rightarrow +\infty} 0$  en décroissant, i.e.  $f_n$  tend vers 0 dans  $L^1(\Omega \times \Omega; \gamma)$ . Quitte à extraire une sous-suite, on peut supposer que  $f_n$  converge vers 0  $\gamma$ -presque partout.

On pose ensuite

$$\Gamma = \left\{ (x, y) \in \Omega \times \Omega \mid f_n(x, y) \xrightarrow{n \rightarrow +\infty} 0 \right\}$$

Par définition,  $\gamma(\Gamma) = 1$ . Il reste à voir que  $\Gamma$  est c-cycliquement monotone. Pour  $\sigma \in \mathfrak{S}_k$  et  $(x_i, y_i)_{i=1 \dots K} \in \Gamma$ , on a

$$\sum c(x_i, y_i) = \lim \sum (\phi_n(x_i) + \psi_n(y_i)) = \lim \sum (\phi_n(x_i) + \psi_n(y_{\sigma(i)}))$$

Or  $\phi_n(x_i) + \psi_n(y_{\sigma(i)}) \leq c(x_i, y_{\sigma(i)})$ , ce qui conclut la démonstration.

# Chapitre 3

## Cas de la distance $c(x, y) = d(x, y)$

### 3.1 Premières observations

**Théorème**  $\phi$  est  $c$ -concave si et seulement  $\phi$  est 1-lipschitzienne.

**Démonstration** Le sens direct est évident en remarquant que  $\phi$  est un infimum de fonctions 1-lipschitziennes (inégalité triangulaire renversée). Pour l'autre sens, on montre que  $\phi(x) = \inf_y (d(x, y) + \phi(y))$ . En effet d'une part  $\forall y, \phi(x) \leq \phi(y) + d(x, y)$ , ce qui donne  $\phi(x) \leq \inf_y (d(x, y) + \phi(y))$ , et  $y = x$  donne l'autre inégalité.

**Théorème** Si  $\phi$  est 1-lipschitzienne,  $\phi^c = -\phi$ .

**Démonstration**  $-\phi$  étant 1-lipschitzienne, la démonstration précédente montre que  $(-(-\phi))^c = -\phi$ , d'où le résultat.

D'après la formule de dualité, on a

$$\min_{\gamma} \int c(x, y) d\gamma = \max_{\phi \text{ 1-lip}} \left( \int \phi d\mu + \int \psi d\nu \right) = \max_{\phi \text{ 1-lip}} \int \phi d(\mu - \nu)$$

TODO : Traiter dans un cas simple ?

#### 3.1.1 Cas $c(x, y) = h(x - y)$ et $h$ convexe à valeurs réelles

???

$\mu$  et  $\nu$  sont deux mesures de probabilité,  $\mu$  ne possède pas d'atomes et  $c(x, y) = h(x - y)$  où  $h$  est convexe.

**Théorème** Sous ces hypothèses, il existe un unique transport  $T$  croissant tel que  $T\#\mu = \nu$ . La mesure  $\gamma_T$  associée à ce transport est un minimiseur de  $\int h(x - y) d\gamma$  sur  $\Pi(\mu, \nu)$ , et il est unique si  $T$  est strictement convexe. En terme de fonctions de répartition,  $T$  vérifie  $F(x) = G(T(x))$ .

**Démonstration** On traite le cas où  $h$  est strictement convexe d'abord. Soit  $\gamma$  optimal, alors un théorème précédent dit que le support de  $\gamma$  est  $c$ -cycliquement monotone. Si  $(x_0, y_0), (x_1, y_1) \in \text{Support } \gamma$ , on veut montrer que  $x_1 > x_0 \Leftrightarrow y_1 \geq y_0$  et l'on procède par l'absurde en supposant  $y_1 < y_0$ .

TODO

## 3.2 Retour sur le problème de Monge

On traite désormais le cas de la distance euclidienne :  $c(x, y) = d(x, y) = |x - y|$ , autrement dit la distance associée au problème de MONGE. Il s'agit d'un coût convexe mais non strictement convexe, et les résultats obtenus précédemment avec le théorème de BRENIER ne s'appliquent plus *a priori*. En effet, dans ce cas, la différentielle obtenue en écrivant la minimisation de  $\phi(x) = |x - y_0| - \phi(y_0)$  s'écrit :

$$\nabla\phi(x_0) = \frac{x_0 - y_0}{|x_0 - y_0|}$$

et ce n'est clairement pas une différentielle injective, comme c'était le cas avec l'hypothèse de stricte convexité. On ne peut donc pas en extraire  $y_0$  en fonction de  $x_0$ , ce qui donnait l'expression explicite du transport.

## 3.3 Rayons de transport

### 3.3.1 La dualité de MONGE – KANTOROVICH

On rappelle que la dualité de MONGE-KANTOROVICH permet d'écrire, dans le cas d'un coût qui est une distance, cas dans lequel la  $c$ -concavité équivaut à la 1-lipschitzianité, ou encore à une  $c$ -transformation qui n'est que la multiplication par  $-1$  :

$$\min_{\pi(\mu, \nu)} \int c d\gamma = \int \phi d\mu + \int \phi^c d\nu = \max \int \phi d(\mu - \nu)$$

Gardons également à l'esprit qu'un ensemble  $c$ -cycliquement monotone est un ensemble de correspondances que l'on ne peut améliorer par permutations. Si le coût est continu, on est sûrs que le support d'une mesure optimale est un ensemble  $c$ -cycliquement monotone. S'il est seulement semi-continu inférieurement, le support d'une mesure optimale est contenu dans un ensemble concentrant toute la masse et qui est  $c$ -cycliquement monotone, mais on ne peut plus garantir sa fermeture.

### 3.3.2 Les rayons de transport

Si  $\gamma$  est optimal et si  $u$  est un potentiel de KANTOROVICH, on a par définition

$$\int |x - y| d\gamma = \int (ux - uy) d\gamma$$

Or on sait que le potentiel  $u$  est 1-lipschitzien, ce qui fait que  $ux - uy \leq |x - y|$  pour tous  $x, y$ . L'égalité des intégrales et la continuité impose alors l'égalité  $ux - uy = |x - y|$  au moins sur le support de  $\gamma$ . Observons alors ce qu'il se passe sur un segment  $[x, y]$  où les points  $x$  et  $y$  vérifient  $ux - uy = |x - y|$ . On considère  $z = (1 - t)x + ty$ .

$$\begin{aligned} ux - uz &\leq |x - z| = t|x - y| \\ uz - uy &\leq |z - y| = (1 - t)|x - y| \end{aligned}$$

Or on a alors

$$\begin{aligned} |x - y| = ux - uy &= (ux - uz) + (uz - uy) \\ &\leq t|x - y| + (1 - t)|x - y| = |x - y| \end{aligned}$$

donc les deux inégalités sont nécessairement des égalités, et on obtient finalement que  $u$  est linéaire de pente 1 sur tout le segment. On travaille donc désormais sur les *rayons de transport*, qui sont les segments ouverts maximaux sur lesquels  $u$  est *linéaire* de pente 1.

### 3.3.3 De la différentiabilité sur les rayons de transport

Prouvons tout d'abord que ces segments sont disjoints, en prouvant que  $u$  est différentiable sur ces segments et que sa différentielle  $u$  est constante et de valeur la pente entre  $x$  et  $y$ .

En effet, prenons  $x_0$  à l'intérieur du rayon de transport. Soit  $e$  de norme 1 porté par le rayon, et  $h$  perpendiculaire au rayon. On a

$$\begin{aligned} u(x_0 + h) - u(x_0 + te) &= u(x_0 + h) - u(x_0) - t \geq -|x_0 + h - x_0 - te| = -\sqrt{t^2 + h^2} \\ u(x_0 + h) - u(x_0 - te) &= u(x_0 + h) - u(x_0) + t \leq |x_0 + h - x_0 + te| = \sqrt{t^2 + h^2} \end{aligned}$$

d'où  $|u(x_0 + h) - u(x_0)| \leq t(-1 + \sqrt{1 + \frac{h^2}{t^2}}) \leq \frac{h^2}{2t} = O(h^2) = o(h)$ . On peut toujours prendre un  $t$  assez petit pour que les  $x_0 + te$  et  $x_0 - te$  soient dans le rayon compte tenu que l'on considère les rayons ouverts, il vient alors

$$u(x_0 + te + h) - u(x_0) - \langle e | te + h \rangle = o(h)$$

ce qui prouve que le gradient de  $u$  en  $x_0$  est  $e$ .  $\square$

### 3.3.4 Des rayons partitionnant le support

Les rayons ne peuvent donc se croiser car la différentielle est unique et donne la direction du rayon, on ne peut donc avoir deux directions distinctes la portant. De la même manière, ils ne peuvent se couper entre l'extrémité de l'un et l'intérieur de l'autre, car ce dernier aurait une différentielle portée par deux directions. Par contre, rien n'empêche l'intersection au niveau des extrémité, aucun des deux rayons ne portant de différentielle en ces points.

Considérons désormais une famille de segments  $s \in \mathcal{S}$  contenant le support de  $\gamma$ . Certains peuvent être des singletons, et ils sont d'intérieurs disjoints. Ils peuvent se croiser aux points où la différentielle de  $u$  n'existe pas, qui constituent un ensemble négligeable par le théorème de RADEMACHER, on peut donc se restreindre à l'ensemble  $\Omega \setminus N$  des points où  $u$  admet une différentielle.

On a ainsi partitionné le support en rayons de transport, ce qui justifie l'étude séparée sur chaque rayon et rend possible un recollement final.

## 3.4 Transports optimaux pour la distance

### 3.4.1 La perte de l'unicité de l'optimum

Dans le cas de la *dimension 1*, nous avons déjà vu que pour une mesure  $\mu$  sans atome et pour un coût strictement convexe, on avait l'existence et l'unicité d'un transport optimal, qui est nécessairement monotone. Si le coût est seulement convexe, il n'y a plus nécessairement unicité, mais on est également sûrs que le transport monotone est un transport optimal.

### 3.4.2 Le choix d'un transport optimal

Cette absence d'unicité va nous contraindre à choisir un transport parmi tous ceux optimisant le critère de MONGE, par exemple en imposant une contrainte supplémentaire. On va choisir  $\gamma$  optimisant  $c_2$ , dans le cas où  $\Omega$  est compact. Puisque  $O \subset \pi \subset P$  et que  $\pi$  est alors compact pour la convergence faible, on est assurés de l'existence d'une solution.

Choisissons l'unique  $\gamma_\varepsilon$  qui optimise le critère de transport pour le coût  $c_1 + \varepsilon c_2$ , qui est cette fois-ci strictement convexe. Quitte à extraire, on peut supposer que  $\gamma_\varepsilon \rightarrow \bar{\gamma}$ .

**Théorème.**  $\bar{\gamma}$  minimise  $c_2$  et est dans  $O(\mu, \nu)$ .

*Preuve.* Tout d'abord, par minimalité de  $\gamma_\varepsilon$  pour le coût  $c_1 + \varepsilon c_2$  :

$$\forall \gamma \in \pi(\mu, \nu), \int c_1 d\gamma_\varepsilon \leq \int (c_1 + \varepsilon c_2) d\gamma_\varepsilon \leq \int (c_1 + \varepsilon c_2) d\gamma$$

Or,  $\int (c_1 + \varepsilon c_2) d\gamma_\varepsilon \geq \int c_1 d\gamma_\varepsilon$  et le membre de droite converge vers  $\int c_1 d\bar{\gamma}$  lorsque  $\varepsilon$  devient arbitrairement petit. Donc  $\bar{\gamma}$  est minimal pour  $c_1$ , et est donc dans  $O(\mu, \nu)$ .

On raisonne de la même manière pour montrer que  $\bar{\gamma}$  optimise le coût  $c_2$ . On écrit

$$\begin{aligned} \forall \gamma \in O(\mu, \nu), \int c_1 d\gamma_\varepsilon + \varepsilon \int c_2 d\gamma_\varepsilon &= \int (c_1 + \varepsilon c_2) d\gamma_\varepsilon \\ &\leq \int (c_1 + \varepsilon c_2) d\gamma \text{ par minimalité de } \gamma_\varepsilon \text{ pour } c_1 + \varepsilon c_2 \\ &= \int c_1 d\gamma + \varepsilon \int c_2 d\gamma \\ &\leq \int c_1 d\gamma_\varepsilon + \varepsilon \int c_2 d\gamma \text{ par minimalité de } \gamma \text{ pour } c_1 \end{aligned}$$

Ce qui donne

$$\int c_2 d\gamma_\varepsilon \leq \int c_2 d\gamma$$

Il vient alors par passage à la limite :

$$\int c_2 d\bar{\gamma} \leq \int c_2 d\gamma$$

Ainsi,  $\bar{\gamma}$  est optimal pour  $c_2$  dans  $O$ . Nous prouverons plus tard que ce transport est unique et qu'il dérive d'un transport absolument continu par rapport à la mesure de LEBESGUE  $\lambda$ .  $\square$

On a ainsi prouvé l'existence d'un transport optimal pour  $c_1$  et  $c_2$ .

### 3.4.3 Une caractérisation de $O(\mu, \nu)$

**Théorème.**  $\gamma$  est dans  $O(\mu, \nu)$  si, et seulement si,  $\gamma$  est concentré sur  $\{(x, y) \mid ux - uy = |x - y|\}$  si  $u$  est un potentiel de KANTOROVICH associé au coût  $c$ . Autrement dit, la partition en rayons de transport caractérise les plans de transfert optimaux pour la distance.

*Preuve.* Nous avons déjà prouvé que la condition est nécessaire, par définition du potentiel  $u$  et des marginales de  $\gamma$ . Réciproquement, si  $\gamma$  est concentré sur un tel ensemble, on a

$$\int u d\mu - \int u d\nu = \int (ux - uy) d\gamma = \int |x - y| d\gamma \text{ et } ux - uy \leq |x - y|$$

On absorbe la condition ainsi obtenue en modifiant quelque peu le coût. Le nouveau problème est l'optimisation du critère intégral

$$\int c(x, y) d\gamma \text{ où } c(x, y) = \begin{cases} |x - y|^2 & \text{si } ux - uy = |x - y| \\ +\infty & \text{sinon} \end{cases}$$

où  $\gamma$  appartient à  $\pi(\mu, \nu)$ .

Si  $c$  est semi-continue inférieurement et  $(x_n, y_n)$  converge vers  $(x, y)$ , alors  $ux_n - uy_n = |x_n - y_n|$  converge vers  $ux - uy = |x - y|$  et  $|x_n - y_n|^2$  converge vers  $|x - y|^2$ . Puisque les rayons de transport concentrent la masse et est fermé, il contient le support de  $\gamma$  qui est le plus petit fermé concentrant la masse (WHY ??? :  $\exists \Gamma$   $c$ -cycliquement monotone, et contenu dans le fermé  $\{(x, y) \mid ux - uy = |x - y|\}$  ???). Ainsi la masse est essentiellement concentrée sur les rayons de transport. On note aussi que la condition de  $c$ -monotonie cyclique n'est pas utile si les points permutés ne sont pas dans un rayon de transport.

On exclut désormais les bords des rayons de transport, en travaillant sur  $E_h = \{x | x \in S \text{ et } d(x, \partial S) \geq \frac{1}{h}\}$ . Si  $x$  est dans  $E_h$ , on a :

$$ux = \inf_y (|x - y| + uy)$$

car  $u$  est lipschitzienne donc  $ux - uy \leq |x - y|$ , et il y a égalité pour  $y = x$ . Sur un rayon de transport, il y a égalité partout par linéarité, donc on peut supprimer un voisinage de la singularité  $y = x$  de la distance euclidienne :

$$ux = \min_{x \notin B(x, \frac{1}{h})} (|x - y| + uy)$$

Soit  $c_h$  une fraction rationnelle  $C^\infty$  valant  $|z|$  hors de la boule et supérieure à  $|z|$  à l'intérieur. On a alors

$$ux = \inf_y (|x - y| + uy) \leq u_h = \min_x (c_h(x - y) + uy) \leq \min_{x \notin B(x, \frac{1}{h})} (|x - y| + uy) = ux$$

PARTIE SUR  $u_h$  +

### 3.4.4 Notions de $\lambda$ -convexité

On définit les fonctions  $\lambda$ -convexes comme un élargissement de la classe des fonctions convexes. Intuitivement, l'idée est de ne plus imposer une dérivée seconde positive, mais seulement supérieure à  $-\lambda$ . Plus précisément,  $f$  est  $\lambda$ -convexe si  $f(x) + \lambda \frac{x^2}{2}$  est convexe. Ici,  $c_h$  étant suffisamment régulière, nous pouvons trouver une « borne de convexité ». La fonction  $c_h$  est  $\frac{1}{h}$ -convexe. Or  $u_h$  est  $\frac{1}{h}$ -concave.

Il convient alors naturellement de se demander quelles sont les propriétés des fonctions qui sont à la fois  $\lambda$ -concaves et  $\lambda$ -convexes.

**Théorème.** *Les fonctions à la fois  $\lambda$ -convexes et  $\lambda$ -concaves sont les fonctions qui sont de classe  $C^1$  et dont le gradient est  $\lambda$ -lipschitzien.*

*Preuve?* La démonstration de ce résultat est relativement simple. Une fonction  $\lambda$ -concave et  $\lambda$ -convexe est une fonction vérifiant

$$\langle -\lambda|x - y|^2 \nabla f(x) - \nabla f(y) |x - y \rangle \leq \lambda|x - y|^2$$

Si  $f$  est de classe  $C^2$ , alors cela équivaut à  $\|D^2 f\| \leq \lambda$ , i.e. à la  $\lambda$ -lipschitzianité de  $\nabla f$ . Si  $f$  est moins régulière, on peut la rendre de classe  $C^2$  en la convolant avec un noyau de convolution régulier. Les propriétés considérées étant des critères continus, le passage à la limite conserve la  $\lambda$ -lipschitzianité du gradient, et donne le résultat pour des fonctions  $f$  quelconques.  $\square$

Si la contrainte est satisfaite, et si tous les  $(x_i, y_j)$  sont dans le même rayon de transport, on obtient une propriété de croissance :

$$\langle x_2 - x_1 | y_2 - y_1 \rangle \geq 0$$

### DESINTEGRATION DE MESURES

Notons  $E_h^+$  l'ensemble des  $x$  pour lesquels  $u$  et  $u_h^+$  sont confondus, et  $E_h^-$  celui des  $x$  pour lesquels  $u$  et  $u_h$  sont confondus. Si on ne veut pas se limiter aux  $E_h$ , on utilise le théorème suivant : chaque  $E_h^+$  est réunion dénombrable d'ensembles  $A_k^h$  sur lesquels  $\nabla u$  est lipschitzienne.

Sur ces ensembles, la fonction  $f(t, y) = y + \nabla u(y)$  est lipschitzienne car les deux termes de la somme le sont.  
PAS D'ATOMES POUR MU S

### 3.4.5 Conclusion

Nous avons désormais de quoi prouver qu'il n'existe qu'un unique optimisant  $\gamma$  dans  $O(\mu, \nu)$  qui optimise le coût  $c_2$ . On décompose une mesure  $\gamma$  optimale sur les rayons de transport, qui la caractérisent presque partout. Notons ces restrictions  $\gamma_S$ . Le cas du coût  $c$  en dimension 1 affirme que les  $\gamma_S$  sont des transports, autrement dit des  $\gamma_{T_S}$ . Il suffit alors de définir  $T$  sur la partition définie par la partition en rayons de transport.  $\mu_S$  n'a pas d'atomes car  $\nabla u$  est lipschitzienne. L'unicité vient, comme toujours, de la convexité de l'ensemble des mesures optimales.  $\square$

### 3.5 Le cas du coût $c(x, y) = c_\infty(x, y) = \max_{\text{Supp}(\gamma)} |x - y|$

Les cas  $c_1$  et  $c_p$  pour  $p > 1$  ont donc été traités. Il reste à regarder le cas  $c_\infty$ , qui revient à trouver des transports qui minimisent les déplacements longs. Notons déjà que les normes  $p$  tendent en croissant vers la norme  $\|\cdot\|_\infty$ . Ainsi,  $c_\infty$  est semi-continue inférieurement pour la convergence faible comme limite supérieure de fonctions continues.

**Théorème.** Il existe un minimiseurs  $\gamma$  qui optimise le coût  $c_\infty$  sur  $\pi(\mu, \nu)$ .

Notons qu'il y a peu d'espoir d'espérer une quelconque unicité, compte tenu de la contrainte extrêmement faible et tenant très peu en compte la totalité du transport.

*Preuve.* En notant  $O_\infty(\mu, \nu)$  l'ensemble des mesures optimisant le critère  $\int c_\infty d\gamma$ , et en notant  $L_\infty$  la valeur de ce minimum, on a

$$\gamma \in O_\infty(\mu, \nu) \iff \max_{\text{Supp}(\gamma)} |x - y| = L \iff \text{Supp}(\gamma) \subseteq \{(x, y) \mid |x - y| \leq L\}$$

Comme précédemment avec le coût  $c_1$ , on va imposer une contrainte supplémentaire pour trouver un transport optimal.

**Théorème.** Il existe une unique mesure de  $O_\infty(\mu, \nu)$  qui optimise le coût quadratique  $c_2$ , et que cette mesure optimale dérive d'un transport.

*Preuve.* On recherche un minimiseur pour le critère  $\int c_2 d\gamma$  sur  $O_\infty(\mu, \nu) = c_\infty^{-1}(]-\infty, L])$  qui est fermé car  $c_\infty$  est une fonction continue.  $O_\infty(\mu, \nu)$  est donc compact comme fermé dans  $\mathbb{P}(\Omega)^2$ , et le minimum existe donc bien.

PREUVE A FINIR

#### 3.5.1 Retour sur les cas des transports impossibles

Si  $\mu = \delta_x$  et si  $\nu = \frac{1}{2}(\delta_y + \delta_z)$ , il n'y a aucun transport possible, donc sûrement pas de transport optimal pour un critère fixé...

# Chapitre 4

## Le problème de Beckmann

Nous avons précédemment traité le problème du transport optimal de MONGE-KANTOROVICH dans le cas  $c(x, y) = |x - y|$ , dans le cas des distances  $c_p$  pour  $p > 1$ , ainsi que dans le cas  $c_\infty$ . Nous considérons dorénavant des problèmes ayant leur source dans la physique ou dans l'économie, et qui se ramènent simplement et naturellement à un problème de transport, que nous pouvons traiter avec les outils, les méthodes et les résultats développés précédemment.

### 4.1 Le problème de Beckmann

**Date ?**

#### 4.1.1 Le problème physique de BECKMANN

On considère un champ de vecteurs  $v : \Omega \rightarrow \mathbb{R}^d$  de classe  $\mathcal{C}^1$ , et deux mesures  $\mu$  et  $\nu$  quelconques. Le problème de BECKMANN est celui de la minimisation du critère intégral suivant :

$$\int_{\Omega} |v(x)| dx \text{ où } \nabla \cdot v = \mu - \nu$$

Le théorème de STOKES permet alors d'écrire :

$$\int_{\Omega} \nabla \cdot v = \int_{\Omega} d(\mu - \nu) = \int_{\partial\Omega} v \cdot n$$

où  $n$  désigne le vecteur normal au bord de  $\Omega$ , orienté vers l'extérieur. Le second membre de cette expression est la composante normale de  $v$  sur le bord et désigne, physiquement, la quantité de masse sortante de l'enceinte  $\Omega$ . Le champ de vecteurs  $v$  peut donc être interprété comme un flux de matière. La quantité de matière sortant est globalement  $\mu(\Omega) - \nu(\Omega)$ . La mesure  $\mu$  modélise ainsi la distribution de matière de la source, et  $\nu$  celle de la cible.

On est ici en présence d'un choix de méthode d'observation des mouvements continus et fluides. On peut choisir une description *lagrangienne*, en suivant l'évolution d'une particule fixée et en représentant par une mesure  $\gamma \in \mathcal{P}(\Omega \times \Omega)$  le nombre de particules se trouvent en  $x$  à l'instant initial et en  $y$  à l'instant final. On peut également opter pour un point de vue *eulérien*, en observant uniquement une région de l'espace, ce qu'il s'y passe, le nombre de particules y circulant, les comportements qu'elles manifestent, etc.

### 4.1.2 Reformulation du problème de minimisation

On cherche désormais à évaluer la quantité totale du mouvement, que l'on cherche naturellement à minimiser. Modifions la contrainte pour en conserver une interprétation au sens des distributions :

$$(\mathcal{H}) \quad \forall \phi \in \mathcal{C}^1(\Omega), \quad \int_{\Omega} v \cdot \nabla \phi = - \int_{\Omega} \phi d(\mu - \nu)$$

Pour un champ de vecteurs  $v$  suffisamment régulier, cette condition est équivalente à  $\nabla \cdot v = \mu - \nu$  et à  $\nabla \cdot n = 0$  sur le bord. Cette nouvelle condition correspond à l'absence de masse traversant  $\partial\Omega$ , autrement dit on se place en régime permanent.

Introduisons une quantité caractérisant cette condition dans le problème de minimisation :

$$\sup_{\phi \in \mathcal{C}^1(\Omega)} \left( \int_{\Omega} \phi d(\mu - \nu) + \int_{\Omega} v \cdot \nabla \phi \right) = \begin{cases} 0 & \text{si } (\mathcal{H}) \\ +\infty & \text{sinon} \end{cases}$$

Le problème de minimisation de BECKMANN se réécrit alors :

$$\inf_{v|(\mathcal{H})} \int |v(x)| dx = \inf_{v|(\mathcal{H})} \left[ \int_{\Omega} |v(x)| dx + \sup_{\phi \in \mathcal{C}^1(\Omega)} \left( \int_{\Omega} \phi d(\mu - \nu) + \int_{\Omega} v \cdot \nabla \phi \right) \right]$$

En permutant sans justification la borne supérieure de la borne inférieure, on se ramènerait au *problème dual* de  $(\mathcal{B})$  :

$$\begin{aligned} & \sup_{\phi \in \mathcal{C}^1(\Omega)} \left( \int_{\Omega} \phi d(\mu - \nu) + \inf_{v|(\mathcal{H})} \int_{\Omega} (|v(x)| + v(x) \cdot \nabla \phi) dx \right) \\ &= \sup_{\phi \in \mathcal{C}^1(\Omega)} \left( \int_{\Omega} \phi d(\mu - \nu) + \inf_{v|(\mathcal{H})} \int_{\Omega} |v(x)| \left( 1 + \frac{v(x)}{|v(x)|} \cdot \nabla \phi \right) dx \right) \end{aligned}$$

Notons que cette permutation, bien qu'*a priori* illicite, a pour propriété d'être toujours au moins une inégalité :

$$\forall A(v, \phi), \quad \inf_v \sup_{\phi} A(v, \phi) \geq \sup_{\phi} \inf_v A(v, \phi)$$

La quantité  $\frac{v(x)}{|v(x)|}$  est un vecteur unitaire, donc est entièrement caractérisé par sa direction et son sens. On fixe donc une direction nécessaire pour minimiser  $|v(x)| \left( 1 + \frac{v(x)}{|v(x)|} \cdot \nabla \phi \right)$ , qui est celle opposée à  $-\nabla \phi$ , ce qui permet d'évaluer la borne inférieure

$$\inf_{v|(\mathcal{H})} \int_{\Omega} |v(x)| (1 - |\nabla \phi|) dx = \begin{cases} 0 & \text{si } |\nabla \phi| \leq 1 \\ -\infty & \text{sinon} \end{cases}$$

Puisqu'un gradient borné par 1 est équivalent à une fonction différentiable et 1-lipschitzienne, le problème dual est ramené à :

$$\begin{aligned} & \sup_{\phi \in \mathcal{C}^1(\Omega)} \left( \int_{\Omega} \phi d(\mu - \nu) + \inf_{v|(\mathcal{H})} \int_{\Omega} (|v(x)| + v(x) \cdot \nabla \phi) dx \right) \\ &= \sup_{|\nabla \phi| \leq 1} \int_{\Omega} \phi d(\mu - \nu) \\ &= \sup_{\phi \text{ 1-Lip}} \int_{\Omega} \phi d(\mu - \nu) \\ &= \min_{\gamma \in \pi(\mu, \nu)} \int_{\Omega} |x - y| d\gamma \end{aligned}$$

On réduit alors le problème de BECKMANN à une sous-classe beaucoup plus restreinte de champs de vecteurs :

$$\inf \left\{ \int |v(x)dx \mid v \in \mathcal{L}^1(\Omega), \forall \phi \in \mathcal{C}^1(\Omega), \int \nu \cdot \nabla \phi = - \int \phi d(\mu - \nu) \right\}$$

Mais l'espace  $\mathcal{L}^1$  est peu concilient, en particulier il ne présente pas de propriété de compacité faible...

## 4.2 L'intervention des mesures vectorielles

Le manque de généralité des mesures scalaires nous amène naturellement à introduire la notion de mesure vectorielle. Une application  $\mu : \Omega \longrightarrow \mathbb{R}^d$  est une mesure si elle vérifie les propriétés usuelles suivantes :

- \*  $\mu(\emptyset) = 0$
- \*  $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$  pour une union disjointe

Ce sont bien évidemment des mesures moins faciles à manipuler que les mesures positives. On aimerait associer à une mesure vectorielle quelconque  $\mu$  une mesure scalaire positive  $|\mu|$ . On ne peut pas définir  $|\mu|(A) = |\mu(A)|$  car ce ne serait plus une mesure. On la définit comme

$$\forall A \in \mathfrak{B}(\Omega), |\mu|(A) = \sup_{\cup_i A_i \subseteq A} \sum_i |\mu(A_i)|$$

L'application  $|\cdot|$  ainsi définie est une norme sur l'espace  $\mathcal{M}(\Omega, \mathbb{R}^d)$  des mesures vectorielles sur  $\Omega$ . Comme nous l'avons déjà signalé pour justifier l'utilisation de la convergence faible telle que nous l'avons définie, l'espace  $\mathcal{M}(\Omega, \mathbb{R}^d)$  est le dual de l'espace des fonctions continues sur  $\Omega$  si celui-ci est compact, ou sur l'ensemble des fonctions continues sur  $\Omega$  et nulle à l'infini de manière plus générale.

La norme duale serait alors :

$$\begin{aligned} \|\mu\|_{C(\Omega)'} &= \sup\{\langle \mu, \phi \rangle \mid \|\phi\| \leq 1\} \\ &= \sup\left\{ \int_{\Omega} \phi d\mu \mid \phi \in C(\Omega, \mathbb{R}^d), |\phi| \leq 1 \right\} \\ &= |\mu|(\Omega) \end{aligned}$$

D'où l'intérêt de la norme ainsi introduite. On dira que  $|\mu|$  est la valeur absolue de la mesure  $\mu$ .

## 4.3 Dans l'intimité du problème dual

### 4.3.1 Retour sur le problème de BECKMANN

Maintenant, on regarde le problème de la minimisation de  $|v|(\Omega)$  où  $v$  est une mesure vectorielle et vérifie la condition

$$\int_{\Omega} \nu \cdot \nabla \phi = - \int_{\Omega} \phi d(\mu - \nu)$$

On va prouver que ce problème a toujours une solution. En effet, soit  $(v_n)_n$  une suite minimisante du critère. Quitte à extraire<sup>1</sup>, on peut supposer qu'elle converge faiblement vers un certain champ vectoriel  $v$ . On a alors

$$|v|(\Omega) \leq \liminf_{n \rightarrow \infty} |v_n|(\Omega)$$

par semi-continuité inférieure de la norme (duale. VRAI?).

<sup>1</sup>. Une suite minimisante est toujours bornée à partir d'un certain rang, à moins que le critère soit toujours infini, cas peu intéressant à traiter...

### 4.3.2 Liens avec les critères de Monge et de Kantorovich

Rappelons qu'en vertu de l'inégalité rappelée lors de la permutation des bornes inférieures et supérieures pour obtenir le problème dual, on obtient la chaîne d'inégalités suivante si le champ  $v$  est admissible, *i.e.* vérifie :

$$(\mathcal{M}) = (\mathcal{D}) \leq (\mathcal{B}) \leq |v|(\Omega) \leq (\mathcal{M})$$

en rappelant les différents critères optimaux rencontrés jusque alors :

$$\begin{aligned} (\mathcal{M}) &= \int |x - y| d\gamma \text{ pour la problème de MONGE — KANTOROVICH} \\ (\mathcal{D}) &= \int \phi d\mu - \int \psi d\nu \text{ pour la problème dual de MONGE — KANTOROVICH} \\ (\mathcal{B}) &= \int v(x) dx \text{ pour le problème de BECKMANN} \end{aligned}$$

Il y a donc égalité tout au long de la chaîne d'inégalités.

### 4.3.3 L'intervention de $v_\gamma$ et de $\sigma_\gamma$

Soit  $\gamma$  optimal pour le critère  $(\mathcal{M})$ . On définit  $v_\gamma$  par la mesure donnant l'intégration des fonctions sur tous les segments de l'espace  $\Omega$  :

$$\forall \phi \in \mathcal{C}^1(\Omega), \langle v_\gamma, \phi \rangle = \int_\Omega \phi dv_\gamma = \int_{\Omega \times \Omega} d\gamma(x, y) \int_0^1 \phi(\omega_{x,y}(t)) \omega_{x,y}(t) dt = \int_\Omega d\gamma(x, y) \int_{[x,y]} \phi$$

où  $\omega_{x,y} : t \mapsto (1-t)x + ty$  est une paramétrisation affine du segment  $[x, y]$ . Pour justifier que  $v_\gamma$  est bien une mesure, il suffit de dire que  $\langle v_\gamma, \phi \rangle$  est borné par  $C\|\phi\|_\infty$ . En fait prendre pour  $C$  le minimum du critère de MONGE convient.

On est rassurés de savoir que  $v_\gamma$  est une mesure, mais il reste à vérifier qu'elle vérifie bien la contrainte du problème de BECKMANN. Prenons  $\psi \in \mathcal{C}^1(\Omega)$ . On a :

$$\begin{aligned} \int_\Omega v_\gamma \cdot \nabla \psi &= \int_{\Omega \times \Omega} d\gamma(x, y) \int_0^1 \nabla \psi(\omega_{x,y}(t)) \omega_{x,y}(t) dt \\ &= \int_{\Omega \times \Omega} [\psi \circ \omega_{x,y}]_0^1 d\gamma(x, y) \\ &= - \int_{\Omega \times \Omega} \psi d(\mu - \nu) \text{ car } \gamma \in \pi(\mu, \nu) \end{aligned}$$

Par l'égalité entre tous les critères intégraux établie précédemment, on a  $(\mathcal{M}) = |v_\gamma|(\Omega) = \|v_\gamma\|_{\mathcal{C}(\Omega)'} = \sup\{\langle v_\gamma, \phi \rangle \mid \phi \in \mathcal{C}(\Omega), |\phi| \leq 1\}$ , d'où :

$$\langle v_\gamma, \phi \rangle \leq \int |x - y| d\gamma = (\mathcal{M}) = |v|(\Omega)$$

On peut également définir  $\sigma_\gamma$  :

$$\forall \phi \in \mathcal{C}^1(\Omega), \langle \sigma_\gamma, \phi \rangle = \int_{\Omega \times \Omega} d\gamma(x, y) \int_0^1 \phi(\omega_{x,y}(t)) \left| \omega_{x,y}(t) \right| dt$$

Il s'agit en fait de  $|v_\gamma|$ . On a en tous cas immédiatement que  $|v_\gamma|$  est inférieure à  $\sigma_\gamma$ , et  $\sigma_\gamma(\Omega)$  qui est égale au minimum du critère de MONGE  $(\mathcal{M})$ . Ainsi, on a comme précédemment la chaîne d'inégalités :

$$|v_\gamma|(\Omega)(\mathcal{M}) = (\mathcal{D}) = (\mathcal{B}) \leq \sigma_\gamma(\Omega) \leq (\mathcal{M})$$

## 4.4 Généralisation du problème de Beckmann

Si on prend une application continue  $K$  :

$\Omega \rightarrow \mathbb{R}_+$ , on peut chercher à minimiser un critère intégral un peu plus général que celui initialement proposé par BECKMANN :

$$\int_{\Omega} K(x)|v(x)|dx = \int_{\Omega} Kd|v| \text{ où } \nabla \cdot v = \mu - \nu$$

Par exemple dans le cas de la modélisation d'un problème de trafic urbain,  $K$  n'a pas de raison d'être, comme dans le problème initial, constante : le transport n'est *a priori* pas aussi coûteux partout, ce que l'on comprend bien dans le cas de goulots d'étranglement. Notamment dans les applications, cette pondération sera souvent concave.

Le même raisonnement que précédemment permet de travailler sur le problème dual, en se restreignant aux fonctions  $\phi$  qui ont un gradient borné par  $K$ . On pose

$$d_K(x, y) = \inf \int d\gamma \int_0^1 (\phi \circ \omega) \dot{\omega} \text{ où } \omega(0) = x, \omega(1) = y$$

On cherche à trouver un candidat intéressant pour un  $v$  optimal. Soit  $\gamma$  optimal pour  $\int d_K(x, y)d\gamma$ . On a alors la minimisation de

$$\langle v_{\gamma}, \phi \rangle = \int d\gamma \int_0^1 (\phi \circ \omega) \dot{\omega}$$

pour  $\omega_{x,y}$  qui est une géodésique pour  $d_K$ . La question qui se pose est alors celle de l'unicité : il y a-t-il plusieurs géodésiques, ou plusieurs  $\gamma$  optimaux ? Le plus souvent,  $v_{\gamma}$  n'est pas unique.

La pondération  $K$  pourrait également être une fonction de  $v$ . Par exemple dans des problèmes d'étude de congestion de trafic routier, on étudie

$$\min \int H(|v(x)|)dx$$

L'optimisation pour la norme  $\mathcal{L}^2$  est également courante, et est beaucoup plus aisée que le cas de la norme  $\mathcal{L}^1$ , dont le comportement est bien moins amical.

## 4.5 Analyse de $\sigma_{\gamma}$

Ce problème est lié au système elliptique

$$\begin{cases} -\nabla \cdot (\sigma \nabla u) = \mu - \nu \\ |\nabla u| \leq 1 \text{ sur } \Omega \\ |\nabla u| = 1 \text{ } \lambda\text{-p.p.} \end{cases}$$

appelé *système de MONGE — KANTOROVICH*. On observe que pour un  $v$  optimal suffisamment régulier et pour un potentiel de KANTOROVICH  $u$ , on peut écrire :

$$\int |v| = \int u d(\mu - \nu) = - \int \nabla u \cdot v$$

Pour minimiser le critère,  $v$  doit être positivement colinéaire au gradient de  $u$ . Cependant, on ne peut pas écrire cette condition pour des fonctions seulement lipschitziennes. On résout ce problème en régularisant la fonction par convolution : on considère  $u_h = u \star \rho_h$  où  $\rho_h$  est un noyau de convolution régulier convergeant vers l'identité, et on a

$$\begin{cases} u_h \xrightarrow{\text{CVU}} u \\ \nabla u_h \xrightarrow{\text{p.p.}} \nabla u \end{cases}$$

Le critère intégral de MONGE pour  $u_h$  converge donc vers le minimum du problème de MONGE, qui est aussi le minimum des problèmes duaux de MONGE — KANTOROVICH et bidual de BECKMANN. Les  $u_h$  sont donc presque optimaux.

On a de plus  $\int \nabla u_h \cdot v \leq |v|(\Omega)$  car  $|\nabla u_h| \leq 1$  car  $|\nabla u| \leq 1$  et  $\nabla u_h = \nabla u * \rho_h$ .

On a alors

$$\int_{\Omega} d\sigma(1 + \nabla u_h \cdot w) \xrightarrow{n \rightarrow \infty} 0$$

où  $w$  est la direction de  $v$  :  $v = \sigma w$  et  $|w| = 1$ . Cela implique donc, puisque  $1 + \nabla u_h \cdot w$  est toujours positif car  $\nabla u_h$  est de norme inférieure à 1, que  $1 + \nabla u_h \cdot w$  tend vers 0 avec  $h$ , donc que  $\nabla u_h \cdot w$  tend vers  $-1$   $\sigma$ -presque partout, quitte à extraire. Les deux termes du produit scalaire étant de norme inférieure à 1, cette convergence ne laisse qu'une solution envisageable :

$$\nabla u_h \xrightarrow{h \rightarrow 0} -w$$

Autrement dit le gradient de  $u_h$  tend à s'orienter dans la direction opposée à  $w$ , donc dans la direction opposée à  $v$ . On introduit alors  $E = \{x \in \Omega \mid \exists \lim_{h \rightarrow 0} \nabla u_h(x) = \theta, |\theta| = 1\}$  sur lequel  $\sigma$  est concentrée, et telle que  $v = -\sigma\theta$ , comme voulu.

On vérifie ainsi le système elliptique de MONGE — KANTOROVICH pour les  $\nabla u_h$ , donc il en va de même pour la limite.

Observons néanmoins qu'il n'est pas évident que tout  $v$  optimal soit de la forme  $v_\gamma$ . De plus, il n'est pas dit que  $v_\gamma$ , si tant est qu'il existe, soit unique, il pourrait en effet varier en fonction du  $\gamma$  optimal choisi.

Dans le cas où  $\omega$  est convexe et compact, on a  $\mu$  absolument continue par rapport à la mesure de LEBESGUE  $\lambda$  si, et seulement si, il existe  $v$  optimal de la forme  $v_\gamma$  et  $\sigma_\gamma$  est absolument continue par rapport à  $\lambda$ .

Pour le prouver, fixons une mesure  $\gamma$  optimale pour le critère  $(\mathcal{M})$ . On a alors

$$\begin{aligned} \sigma_\gamma &= \int_{\Omega} = \int_{\Omega} \pi_t \# (|x - y| \gamma) \text{ où } \pi_t(x, y) = (1 - t)x + ty \\ \text{d'où } \langle \sigma_\gamma, \phi \rangle &= \int d\gamma(x, y) \int_0^1 \phi(\omega_{x,y}) |\omega_{x,y}| dt \\ &= \int d\gamma \int_0^1 \phi(\pi_t(x, y)) |x - y| dt \\ &= \int_0^1 dt \int_{\Omega \times \Omega} \phi(\pi_t(x, y)) |x - y| d\gamma \\ &= \int_0^1 dt \int_{\Omega \times \Omega} \phi d[\pi_t(|x - y| \gamma)] \end{aligned}$$

Puisque  $\Omega$  est compact,  $|x - y|$  est bornée par son diamètre et donc

$$v_\gamma \leq \sigma_\gamma \leq \text{diam}(\Omega) \int_0^1 [\pi_t \# \gamma] dt$$

On cherche donc à étudier plus précisément  $\mu_t = \pi_t \# \gamma$ . Montrons que pour presque tout  $t$ , on a  $\mu_t$  qui est absolument continue par rapport à la mesure de LEBESGUE  $\lambda$ .

### Cas où $v$ est atomique

On peut alors écrire  $v$  sous la forme

$$v = \sum_{i=1}^m a_i \delta_{y_i}$$

Prenons  $\gamma$  optimal pour le critère  $(\mathcal{M})$ . On a pour un transport optimal  $T$ , vérifiant  $T\#\mu = v$ , et donc  $T(x)$  est l'un des  $y_i$   $\mu$  presque partout. On se limite donc à considérer les régions envoyées sur chacun des  $y_i$ , notons  $R_i = T^{-1}(\{y_i\})$ .

Sur  $R_i$ ,  $T$  est constante de valeur  $y_i$ , et  $(1-t)x + ty = (1-t)x + ty_i$ , donc on a une expression simplifiée de  $\mu_t$  :

$$\mu_t = \sum_i [T_i^t = (1-t)id + ty_i] \# \mu_{R_i}$$

Et  $T_i^t$  est un transport de  $\mu_{R_i}$  à  $\delta_{y_i}$ . On développe alors :

$$\gamma = \gamma_T = (id \times T) \# \mu = \sum_i (id \times T) \# \mu_{R_i} = \sum_i (id \times y_i) \# \mu_{R_i}$$

Les  $T_i^t$  sont des difféomorphismes comme applications affines. Si l'un des  $\mu_{R_i}$  possédait une densité  $f$ , autrement dit s'il n'était pas atomique, la densité de  $(\mu_t)|_{T_i^t(R_i)}$  devrait être :

$$x \mapsto \frac{f((T_i^t)^{-1}(x))}{(1-t)^d}$$

Et on aurait  $(\mu_t)|_{T_i^t(R_i)}$  qui serait absolument continue par rapport à  $\lambda$ .

Les  $T_i^t(R_i)$  sont presque partout disjoints au cours du temps. En effet, posons  $x_i$  un point de  $R_i$ ,  $x_j$  un point de  $R_j$ ,  $z$  un point de  $T_i^t(R_i) \cap T_j^t(R_j)$  s'écrivant sous la forme  $z = (1-t)x_i + ty_i = (1-t)x_j + ty_j$ . Alors deux rayons de transport de croiseraient en leur intérieur, ce qui signifie qu'ils sont confondus par l'étude déjà présentée de ces rayons de transport. Le point  $z$  est donc sur la droite  $(y_i, y_j)$  : si l'on retire toutes ces droites, qui forme une quantité négligeable de points, toutes les régions se retrouvent disjointes. Bien évidemment, on se place en dimension supérieure à deux pour pouvoir retirer les droites sans état d'âme, le cas de la dimension 1 ayant déjà été traité auparavant.

Par définition de  $\mu_t$ , on a :

$$\begin{aligned} \forall \phi \in \mathcal{C}^1(\Omega), \int_{\Omega} \phi d\mu_{t,R_i} &= \int \phi \chi_{R_i} d\mu_t \\ &= \int \phi \chi_{R_i} d(((1-t)id + tT) \# \mu) \\ &= \int \phi \chi_{R_i} d(((1-t)id + ty_i) \# \mu) \\ &= \int \phi \chi_{R_i} d(T_i^t \# \mu) \\ &= \int \phi(T_i^t(x)) \chi_{R_i}(T_i^t(x)) d\mu \\ &= \int \phi(T_i^t(x)) \chi_{(T_i^t)^{-1}}(x) f(x) dx \\ &\stackrel{y=T_i^t(x)}{=} \int \phi(y) \chi_{(T_i^t)^{-1}}(y) \frac{f((T_i^t)^{-1}(y))}{(1-t)^d} dy \\ &= \int \phi(y) \chi_{(T_i^t)^{-1}}(y) g(y) dy \end{aligned}$$

## Cas général

Rappelons que si  $\mu$  est une mesure absolument continue par rapport à la mesure de LEBESGUE  $\lambda$ , on sait que

$$\forall \varepsilon > 0, \exists \delta > 0, \forall B \in P(\Omega), (\lambda(B) \leq \delta \implies \mu(B) \leq \varepsilon)$$

Si  $v$  n'est pas atomique, considérons une suite  $v_n$  de mesures atomiques approchant  $v$ . Cette suite peut être obtenue, par exemple, en partitionnant le plan en carrés de côté  $\frac{1}{n}$ , et en concentrant la masse de chaque carré en son centre.

Par ce qui précède, les  $v_n$  étant atomiques, on peut choisir  $\gamma_n = \gamma_{T_n}$  une mesure dérivant d'un transport, optimale de  $\mu$  à  $v_n$ . Quitte à extraire, on peut supposer que cette suite  $(\gamma_n)_n$  converge vers une certaine mesure  $\gamma$ .

Considérons les deux lemmes suivants, qui vont nous permettre d'achever la preuve : la limite d'une suite de transports optimaux pour  $\mu$  et  $v_n$  est optimale pour  $\mu$  et  $v$  ; l'opérateur  $\phi\#$  est continu pour la convergence faible de mesures.

Par le premier lemme,  $\gamma$  est optimal dans  $\pi(\mu, v)$ .

Notons  $\mu_t^n = \pi_t\#\gamma_n$ . Par le second lemme, les deux quantités égales  $\mu_t^n$  et  $\pi_t\#\gamma_n$  convergent faiblement respectivement vers  $\mu_t$  et  $\pi_t\#\gamma$ , ces deux limites sont donc égales :  $\pi_t\#\gamma = \mu_t$ .

Il suffit désormais de dire que si  $\lambda(A)$  est nulle, alors il en va de même pour  $\mu_t(A)$  quel que soit  $t$ , en vertu de l'absolue continuité de ces dernières par rapport à la mesure de LEBESGUE. Soit  $A$  de mesure de LEBESGUE nulle, et soit  $\delta > 0$ . On sait qu'il existe un ouvert  $\omega$  contenant  $A$  et de mesure de LEBESGUE inférieure à  $\delta$  par définition. On a dans ce cas

$$\mu_t(\omega) \leq \liminf \left( \mu_t^n(\omega) = \int \chi_\omega d\mu_t^n \right) = \int \chi_\omega d\mu_t$$

En effet, la fonction indicatrice d'un ouvert convexe est semi-continue inférieurement, car elle vaut 1 sur celui-ci, et 0 sur les fermés l'entourant (préciser dans le cas général ?).

$$\text{On a } \mu_t^n(\omega) = \mu(\omega') \text{ et } \lambda(\omega') = \frac{\lambda(\omega)}{(1-t)^d}.$$

Si  $\lambda(\omega')$  est inférieur à  $\delta_\varepsilon$ , alors  $\mu(\omega')$  est inférieur à  $\varepsilon$  et il en va donc de même pour les  $\mu_t^n(\omega)$  à partir d'un certain rang. Il suffit alors de choisir  $\delta$  suffisamment petit, par exemple inférieur à  $(1-t)\delta_\varepsilon/2$ , pour avoir  $\mu_t(A) < \varepsilon$ . Puisque  $\varepsilon$  peut être arbitrairement petit,  $A$  est de mesure nulle pour les  $\mu_t$ , donc  $\sigma_\gamma(A) = 0$ .

Pour prouver le second lemme, si  $\psi$  est une fonction test continue, on a :

$$\int \phi d(\phi\#\gamma_n) = \int (\psi \circ \phi) d\gamma_n \xrightarrow{n \rightarrow \infty} \int (\psi \circ \phi) d\gamma = \int \psi d(\psi\#\gamma)$$

et puisque  $\int (\psi \circ \phi) d\gamma_n = \int \psi d(\psi\#\gamma)$ , on a l'égalité voulue.

Pour le premier lemme, on a  $\int |x-y| d\gamma_n$  qui converge vers  $\int |x-y| d\gamma$ , donc le minimum du problème  $((\mathcal{M}), \mu, v_n)$  converge vers celui de  $((\mathcal{M}), \mu, v)$ . La mesure  $\gamma$  réaliserait le minimum et serait admissible par le second lemme, qui prouve immédiatement qu'il a les bonnes marginales par passage à la limite.

Dans notre cas, il suffit de prendre  $\gamma$  dans  $\pi(\mu, v)$ , ainsi que les  $\gamma_n$ , qui convergent faiblement vers  $\gamma$ . On a alors

$$((\mathcal{M}, \mu, v) = \int |x-y| d\gamma = \lim_{n \rightarrow \infty} \int |x-y| d\gamma_n \geq \lim_{n \rightarrow \infty} ((\mathcal{M}, \mu, v_n)$$

Si on prend  $\gamma$  optimal,  $\int |x-y| d\gamma_n$  converge vers  $\int |x-y| d\gamma$ , et on a l'égalité (...).

Soient  $p_n : \Omega \rightarrow G_n = \{y_1, \dots, y_m\}$  les centres des carrés d'une grille de pas  $1/n$ . On a  $p_n(x)$  qui vaut  $y_i$  si  $x$  est dans la  $i$ -ième case, d'où

$$\forall i \in \llbracket 1, m \rrbracket, \forall x \in \Omega, |p_n(x) - x| \leq \frac{1}{n}$$

Si l'on prend  $\gamma_n = (id \times p_n) \# \gamma : (x, y) \mapsto (x, p_n(x))$ , il vient :

$$\pi_y \# \gamma_n = p_n(\pi_y) \# \gamma = p_n(v) = p_n \# (\pi_y \# \gamma) =: v^n$$

Les  $v^n$  sont atomiques et convergent faiblement vers  $v$ . On a :

$$\int |x - y| d\gamma_n = \int |x - p_n(y)| d\gamma \text{ et } \int [|x - p_n(y)| - |x - y|] d\gamma \leq \frac{1}{n}$$

d'où la convergence de  $\int |x - y| d\gamma_n$  vers  $\int |x - y| d\gamma$ .

La mesure  $\sigma$  représente la « densité » des rayons de transport, et est appelée la *densité de transport*.

## 4.6 estimations $\mathcal{L}^p$

Dans le cas atomique, on avait la densité de  $\mu_t(x)$  qui valait  $f_t \lambda$ , où

$$f_t : x \mapsto \frac{f((T_i^t)^{-1}(x))}{(1-t)^d}$$

On obtient alors par changement de variable affiche :

$$\int f_t^p(x) dx \stackrel{x=T_i^p(y)}{=} \int_{R_i} f_t^p(T_i^t(y)) (1-t)^d dy = \int_{R_i} \left[ \frac{f(y)}{(1-t)^d} (1-t)^d dy \right]$$

Il vient alors, en considérant la partition du support en régions  $R_i$  :

$$\int_{\Omega} f_t^p = \sum_i \int_{T_i^p(R_i)} f_t^p = \sum_i \int_{R_i} \frac{f(y)^p}{(1-t)^{(p-1)d}} dy = \frac{1}{(1-t)^{(p-1)d}} \int_{\Omega} f^p dy$$

Prouvons désormais que si  $\mu = f \lambda$  avec  $f \in \mathcal{L}^p$  et  $p < d/(d-1)$ , alors  $\sigma$  est également dans  $\mathcal{L}^p$ .

Dans le cas où  $v$  est atomique, toutes les mesures  $\mu_t$  sont dans  $\mathcal{L}^p$  et  $\mu_t = f_t \lambda$  et :

$$\|f_t\|_{\mathcal{L}^p} = \frac{1}{(1-t)^{(p-1)d}} \|f\|_{\mathcal{L}^p}$$

Il vient donc

$$\sigma_{\gamma} \leq \int \mu_t dt = \int f_t \lambda dt$$

Posons  $f_0 = \int_0^1 f_t dt$ . On a alors par inégalité de la moyenne :

$$\|f_0\|_{\mathcal{L}^p} \leq \int_0^1 \|f_t\|_{\mathcal{L}^p} dt$$

Ce qui fait que  $f_0$  est dans  $\mathcal{L}^p$  dès que  $d(1 - \frac{1}{p}) < 1$ .

Dans le cas où  $v$  n'est pas nécessairement atomique, on peut approcher faiblement  $v$  par une suite de mesures atomiques  $v_n$ , et la suite des  $\gamma_n$  optimaux pour les  $v_n$  converge alors aussi faiblement vers  $\gamma$ , ce qui fait que  $\mu_t^n$  qui converge aussi faiblement vers  $\mu_t$ . Si on note  $\mu_t^n = f_t^n \lambda$ , il vient comme précédemment

$$\|f_t^n\|_{\mathcal{L}^p} = \frac{1}{(1-t)^{(p-1)d}} \|f\|_{\mathcal{L}^p}$$

On a donc  $\mu = f_t \lambda$  avec  $\|f_t^n\|_{\mathcal{L}^p} \leq \liminf \|f_t^n\|$ .

Introduisons le lemme suivant : si  $F_p$  est une application de  $\mathcal{P}(\Omega)$  dans  $[0, +\infty]$ , et si  $\mu$  est définie par

$$\mu : x \mapsto \begin{cases} \int f^p & \text{si } \mu = f\lambda \\ +\infty & \text{sinon} \end{cases}$$

Alors  $F_p$  est semi-continue inférieurement pour la topologie faible.

Soit  $\mu_n$  convergeant faiblement vers  $\mu$ . Soit la limite inférieure de  $F_p(\mu_n)$  est infinie, et on a le résultat, soit elle est finie et il existe un compact  $K$  tel que  $F_p(\mu_n(K))$  soit inférieur à une certaine constante  $C$  pour tout  $n$ . Notons  $\mu_n = f_n \lambda$ , on a alors  $\|f\|_{\mathcal{L}^p} \leq C^{\frac{1}{p}}$ , d'où la convergence des  $f_n$  vers une certaine fonction  $f$  de  $\mathcal{L}^p$ . Il reste à prouver que  $\mu$  s'écrit également sous la forme  $f\lambda$  et que  $\|f\|_{\mathcal{L}^p} \leq \liminf \|f_n\|_{\mathcal{L}^p}$ .

Pour  $\phi$  une fonction test continue et bornée, on a :

$$\int \phi d\mu \xleftarrow{n \rightarrow \infty} \int \phi d\mu_n = \int \phi(x) f_n(x) dx \xrightarrow{n \rightarrow \infty} \int \phi(x) f(x) dx$$

Ce qui donne  $\mu = f\lambda$ , comme voulu.

Nous avons ainsi prouvé que si  $\mu$  est absolument continue par rapport à la mesure de LEBESGUE, alors il en va de même pour  $\sigma$  ; et que si  $\mu$  était dans  $\mathcal{L}^p$ , alors il en va de même pour  $\sigma$  dès que  $p < d(d-1)$ . On sait également que si  $\mu$  et  $\nu$  sont toutes les deux dans  $\mathcal{L}^p$ , alors  $\sigma$  également.

## 4.7 Le problème de Beckmann en dimension 1

Dans le cas où l'espace de travail est une droite, on cherche à minimiser le critère intégral ( $\mathcal{B}$ ) sous la condition  $v' = \mu - \nu$ , qui est bien plus agréable que la condition sur le gradient. La première constatation est que  $v$  est à variations bornées, sa dérivée étant une mesure. On peut intégrer l'équation et écrire :

$$v : x \mapsto c + \int_0^x d(\mu - \nu)$$

En particulier, tout  $v$  admissible est bornée. De plus, il y a unicité. En effet, la condition affaiblie au sens des distributions s'écrit

$$\int v \phi' = - \int \phi d(\mu - \nu) \quad \text{i.e.} \quad \int v f = - \int (\int f) d(\mu - \nu)$$

Le théorème de SMIRNOV affirme que tout  $v$  optimal est de la forme  $v_\gamma$ . L'unicité devient alors l'égalité de tous les  $v_\gamma$  pour un  $\gamma$  optimal.

Dans le cas où  $v$  est une mesure sans cycle, ce qui augmenterait le coût sans intérêt aucun,  $v$  est de la forme

$$\forall \phi, \langle v, \phi \rangle = \int d\theta(u) \int_0^1 \phi(u(t)) u'(t) dt$$

# Chapitre 5

## Cadre topologique et métrique

Nous laissons momentanément de côté les problèmes de transport pour se concentrer sur l'espace  $P(\Omega)$  des mesures de probabilités sur  $\Omega$ , et plus précisément sur ce que les résultats obtenus au cours des études précédentes peuvent nous apporter quant à la richesse de cet espace.

### 5.1 Distances de Wasserstein

Pour  $p \geq 1$  et  $\mu, \nu$  deux mesures de probabilité, on introduit la *distance de WASSERSTEIN* d'ordre  $p$  :

$$W_p(\mu, \nu) = \min \left\{ \left( \int_{\Omega} |x - y|^p d\gamma \right)^{\frac{1}{p}} \mid \gamma \in \pi(\mu, \nu) \right\}$$

On peut également définir :

$$W_{\infty}(\mu, \nu) = \min \{ \max |x - y| \mid \gamma \in \pi(\mu, \nu), (x, y) \in \text{Supp} \gamma \}$$

Nous allons tout d'abord prouver que les  $W_p$  sont bien des distances, puis étudier de plus près la topologie induite.

Tout d'abord,  $W_p$  est symétrique par symétrie du coût  $c_p$ . Concernant l'axiome de séparation, supposons que  $W_p(\mu, \nu) = 0$  pour deux mesures de probabilité  $\mu$  et  $\nu$ . Dans ce cas, il existe une mesure  $\gamma$  dans  $\pi(\mu, \nu)$  telle que  $\int_{\Omega} |x - y|^p d\gamma$  soit nulle, ce qui implique que  $x = y$   $\gamma$ -presque partout. Cela revient à dire que  $\gamma$  est concentrée sur la droite ( $x = y$ ). Par définition, cela implique :

$$\forall \phi \in \mathcal{C}^0(\Omega), \int \phi d\mu = \int \phi(x) d\gamma(x, y) = \int \phi(y) d\gamma(x, y) = \int \phi d\nu$$

Donc nécessairement  $\mu = \nu$ . Il reste à prouver l'inégalité triangulaire. Soient  $\mu, \nu$ , et  $\rho$  trois mesures de probabilité sur  $\Omega$ .

Si les  $\gamma$  associés à chacun des couples de mesures dérivent de transports, ce serait relativement aisé : notons  $T$  et  $S$  des transports optimaux respectivement de  $\mu$  à  $\rho$  et de  $\rho$  à  $\nu$ . Alors le transport  $S \circ T$  serait un transport de  $\mu$  à  $\nu$  et :

$$\begin{aligned}
W_p(\mu, \nu) &\leq \|S \circ T - id\|_{L^p(\mu)} = \|(S \circ T - T) + (T - id)\|_p \\
&\leq \|S \circ T - T\| + \|T - id\| \\
&= W_p(\mu, \rho) + \left( \int |S \circ T - T| d\mu \right)^{\frac{1}{p}} \\
&= W_p(\mu, \rho) + \left( \int |S(y) - y| d(T\#\mu) \right)^{\frac{1}{p}} \\
&= W_p(\mu, \rho) + \left( \int |S - id| d\rho \right)^{\frac{1}{p}} \\
&= W_p(\rho, \nu)
\end{aligned}$$

L'hypothèse qui sert est l'absolue continuité de  $\rho$  (WHY?), donc on pourrait conclure en raisonnant par approximation. Cependant, on va plutôt utiliser un résultat puissant sur les désintégrations de mesures :

**Lemme. (Gluing lemma)** Si l'on a deux mesures  $\gamma_+ \in \pi(\mu, \rho)$  et  $\gamma_- \in \pi(\rho, \nu)$ , alors il existe une mesure de probabilité  $\sigma$  sur  $\Omega^3$  telle que

$$\pi_{x,y}\#\sigma = \gamma_- \quad \pi_{y,z}\#\sigma = \gamma_+$$

En admettant momentanément ce lemme, notons  $\gamma_+$  et  $\gamma_-$  deux mesures réalisant les distances  $W_p$ . Il existe une mesure  $\sigma$  sur le triplet cartésien vérifiant  $\pi_{x,y}\#\sigma = \gamma_-$  et  $\pi_{y,z}\#\sigma = \gamma_+$ . Il vient alors :

$$\begin{aligned}
\pi_x\#\gamma &= \pi_x\#\pi_{x,y}\#\sigma = \pi_x\#\gamma_- = \mu \\
\pi_y\#\gamma &= \pi_y\#\pi_{y,z}\#\sigma = \pi_y\#\gamma_+ = \nu
\end{aligned}$$

Ainsi  $W_p(\mu, \nu) \leq (\int |x - z|^p d\gamma)^{1/p} = (\int |x - z|^p d\sigma)^{1/p} = \|x - z\|_{L^p(\sigma)} \leq \|x - y\|_{L^p(\sigma)} + \|y - z\|_{L^p(\sigma)} = W_p(\mu, \rho) + W_p(\rho, \nu)$ , ce qui clos la preuve.  $W_p$  est bien une distance.  $\square$

La preuve du lemme repose sur la désintégration de mesures. La désintégration de  $\gamma_-$  par rapport à  $y$  — en réalité, à  $\pi_y$  — donne :

$$\forall y \in \Omega, \exists \gamma_-^y \in \mathcal{P}(\Omega), \forall \phi \in \mathcal{C}^0(\Omega), \int_{\Omega} \phi(x, y) d\gamma_- = \int_{\Omega} \int_{\Omega} \phi(x, y) d\gamma_-^y(x) d\rho(y)$$

Une désintégration similaire de  $\gamma_+$  donne :

$$\forall y \in \Omega, \exists \gamma_+^y \in \mathcal{P}(\Omega), \forall \phi \in \mathcal{C}^0(\Omega), \int_{\Omega} \phi(y, z) d\gamma_+ = \int_{\Omega} \int_{\Omega} \phi(y, z) d\gamma_+^y(z) d\nu(y)$$

On considère désormais la mesure  $\gamma_-^y \otimes \gamma_+^y$  sur  $\Omega \times \Omega$ . On introduit la mesure  $\sigma$  définie par :

$$\forall \phi \in \mathcal{C}^0(\Omega), \int_{\Omega} \phi(x, y, z) d\sigma = \int_{\Omega} \int_{\Omega} \phi(x, y, z) d(\gamma_-^y \otimes \gamma_+^y)(x, z) d\rho(y)$$

Pour vérifier que cette mesure satisfait bien les propriétés attendues, il suffit d'écrire :

$$\begin{aligned}
\forall \phi \in \mathcal{C}^0(\Omega), \int_{\Omega} \phi(x, y) d\sigma &= \int_{\Omega} \int_{\Omega} \phi(x, y) d(\gamma_-^y \otimes \gamma_+^y)(x, z) d\rho(y) \\
&= \int_{\Omega} \phi(x, y) d\gamma_-^y d\rho(y) = \int \phi \gamma_- \text{ donc } \pi_{x,y}\#\sigma = \gamma_- \\
\forall \phi \in \mathcal{C}^0(\Omega), \int_{\Omega} \phi(y, z) d\sigma &= \int_{\Omega} \int_{\Omega} \phi(y, z) d(\gamma_-^y \otimes \gamma_+^y)(x, z) d\nu(y) \\
&= \int_{\Omega} \phi(y, z) d\gamma_+^y d\nu(y) = \int \phi d\gamma_+ \text{ donc } \pi_{y,z}\#\sigma = \gamma_+
\end{aligned}$$

Ainsi,  $\sigma$  convient bien.  $\square$

## 5.2 Étude de la topologie

Tout d'abord, il faut veiller à distinguer le cas où  $\Omega$  est compact du cas où il ne l'est pas. De plus, si  $\Omega$  n'est pas borné, il faut surveiller les cas où la distance  $W_p$  pourrait devenir infinie. On travaillera sur l'espace  $(\mathcal{P}(\Omega), W_p)$  dans le cas où  $\Omega$  est compact, et sinon on limite l'ensemble des mesures à celui des mesures de moment d'ordre  $p$  finis, noté  $\mathcal{P}_p(\Omega)$ .

Notons que si  $\mu$  et  $\nu$  sont deux mesures de moments d'ordre  $p$  finis, alors  $W_p(\mu, \nu) < +\infty$ . En effet, le moment étant la distance à une masse de DIRAC, on peut majorer la distance entre les deux mesures par la somme de la mesure de l'une à une masse de DIRAC et de cette masse à l'autre, ce qui donne le résultat.

**Théorème.** Si  $\Omega$  est compact, alors

$$W_p(\mu_n, \mu) \xrightarrow{n \rightarrow \infty} 0 \iff \mu_n \xrightarrow{n \rightarrow \infty} \mu$$

**1<sup>er</sup> cas :**  $p = 1$

Dans ce cas, on a prouvé que :

$$W_1(\mu, \nu) = \min_{\gamma \in \pi(\mu, \nu)} \int |x - y| d\gamma = \sup_{\phi \text{ 1-Lip}} \int \phi d(\mu - \nu)$$

(->) Puisque, pour une fonction  $\phi$  lipschitzienne de rapport 1,  $|\int \phi d(\mu_n - \mu)|$  est majorée par  $W_1(\mu_n, \mu)$  qui tend vers 0 si  $\mu_n$  tend vers  $\mu$  au sens de  $W_1$ , on a  $\int \phi d\mu_n$  qui converge vers  $\int \phi d\mu$ . Le critère étant continu, on peut prolonger cette convergence à toute fonction  $\phi$  continue, par densité. On prouve ainsi la convergence faible de  $\mu_n$  vers  $\mu$ .

(<-) Si  $\mu_n$  converge faiblement vers  $\mu$ , introduisons des potentiels de KANTOROVICH  $\phi_n$  pour chaque problème de transport de  $\mu_n$  à  $\mu$  :

$$\forall n \in \mathbb{N}, W_1(\mu_n, \mu) = \sup_{\phi \text{ 1-Lip}} \int \phi d(\mu_n - \mu) = \int \phi_n d(\mu_n - \mu)$$

Puisque modifier le potentiel en un point ne change pas son caractère optimal, puisque la mesure est sans atomes (WHY ?), on peut supposer que les  $\phi_n$  sont toutes nulles en un  $x_0$  fixé. Il vient alors par 1-lipschitzianité :

$$|\phi_n| \leq \text{diam}(\Omega) + \phi_n(x_0) =: C$$

Ainsi, les  $\phi_n$  sont continues et uniformément bornées. Le théorème d'ASCOLI assure alors l'existence d'une extraction faisant converger uniformément la suite. À renommage des indices près, nous pouvons donc supposer que la suite  $(\phi_n)_n$  converge uniformément vers  $\phi$ . Nous pouvons alors écrire :

$$W_1(\mu_n, \mu) = \int (\phi_n - \phi) d(\mu_n - \mu) + \int \phi d(\mu_n - \mu)$$

Le premier terme de la somme converge vers 0 par convergence uniforme de la différence  $\phi_n - \phi$  vers 0, quant au second terme, il converge vers 0 car  $\phi$  est fixée et la différence de mesures  $\mu_n - \mu$  converge vers 0. Ainsi nous avons comme voulu

$$\limsup_{n \rightarrow \infty} W_1(\mu_n, \mu) = 0$$

**2<sup>e</sup> cas :**  $p > 1$

Il suffit de se servir de la croissance des normes  $\mathcal{L}^p$ . En effet, puisque  $W_1 \leq W_p$ , nous avons

$$\|x - y\|_1 \leq \min \|x - y\|_p \leq \int \text{diam}(\Omega)^{p-1} |x - y| d\gamma$$

Ainsi  $W_1 \leq W_p \leq \text{diam}(\Omega)^{\frac{p-1}{p}} N_1(\mu, \nu)^{\frac{1}{p}}$ . La convergence de  $\mu_n$  vers  $\mu$  au sens de  $W_p$  implique donc celle au sens de  $W_1$  (...??).  $\square$ .

**Corollaire.** Si  $\Omega$  est compact, alors  $(\mathcal{P}(\Omega), W_p)$  l'est aussi.

Pour le voir, il suffit de prendre une suite  $(\mu_n)_n$  et en extraire une sous-suite convergente faiblement au sens des fonctions bornées, ce qui est possible par dualité entre  $C(\Omega)$  et  $M(\Omega)$ .  $\square$

**Théorème.** Si  $\mu$  et  $\nu$  sont deux mesures de moments d'ordre  $p$  finis, alors :

$$W_p(\mu_n, \mu) \xrightarrow[n\infty]{} 0 \iff \begin{cases} \mu_n \xrightarrow[n\infty]{} \mu \\ \int |x|^p d\mu_n \xrightarrow[n\infty]{} \int |x|^p d\mu \end{cases}$$

(->) Notons que, par semi-continuité inférieure, on a toujours  $\int |x|^p d\mu \leq \liminf_{n\infty} \int |x|^p d\mu_n$ . On a également les  $W_p$  qui majorent  $W_1$ . Ainsi, s'il y a convergence vers 0 pour  $W_p$ , il en va de même pour  $W_1$ . En particulier

$$\forall \phi \in 1 - Lip, \int \phi d\mu_n \longrightarrow \int \phi d\mu$$

Ce qui donne la convergence faible des  $\mu_n$  vers  $\mu$ . Pour la seconde condition, si  $W_p(\mu_n, \mu)$  converge vers 0, alors on minore par  $W_p(\mu_n, \delta_0) - W_p(\delta_0, \mu)$ , ce qui implique la seconde condition.

(<-) La fonction  $\phi_R : x \mapsto (|x| \wedge R)^p$  est continue et bornée. On a donc  $\int \phi d\mu_n$  qui converge vers  $\int \phi d\mu$ . D'où la convergence des moments d'ordre  $p$  des  $\mu_n$  vers le moment d'ordre  $p$  de  $\mu$ . D'où :

$$\int_{cB(0,R)} (|x|^p - \phi) d\mu_n \longrightarrow \int_{cB(0,R)} (|x|^p - \phi) d\mu$$

Pour  $\varepsilon > 0$  fixé, il existe  $R$  suffisamment grand tel que

$$\int_{cB(0,R)} (|x|^p - \phi) d\mu \leq \int_{cB(0,R)} |x|^p d\mu < \frac{\varepsilon}{6}$$

Pour  $n$  suffisamment grand, on a également

$$\int (|x|^p - \phi) d\mu_n < \frac{\varepsilon}{3}$$

On veut prouver que  $W_p(\mu_n, \mu)$  est inférieur à  $\varepsilon$  pour  $n$  suffisamment grand. Soit  $\pi_k$  la projection sur  $B(0, R)$  :

$$\pi_k : x \mapsto \begin{cases} x & \text{si } x \in B(0, R) \\ \frac{x}{|x|}R & \text{sinon} \end{cases}$$

C'est une application continue, et même lipschitzienne de degré 1. De plus,  $\pi_R \# \mu_n$  converge faiblement vers  $\pi_R \# \mu$ , en vertu d'un lemme déjà énoncé par le passé. On écrit alors, puisque  $|x - R|^p \leq |x|^p - R^p$  pour  $x$  de norme supérieure à  $R$  et donc  $\int (|x| - R)^p d\mu_n \leq \frac{\varepsilon}{3}$  :

$$\begin{aligned} W_p(\mu_n, \mu) &= \leq W_p(\mu, \pi_R \# \mu_n) + W_p(\pi_R \# \mu_n, \pi_k \# \mu) + W_p(\pi_k \# \mu, \mu) \\ &\leq 2 \left(\frac{\varepsilon}{3}\right)^{1/p} + W_p(\pi_R \# \mu_n, \pi_R \# \mu) \end{aligned}$$

Et ce dernier terme tend vers 0 avec  $n$  qui croît compte tenu de la convergence faible de  $\pi_R \# \mu_n$  vers  $\pi_R \# \mu$ .

$\square$

Un contre-exemple est obtenu en prenant une suite  $\alpha_n$  qui converge vers 0, et en considérant

$$\mu_n = (1 - \alpha_n)\delta + \alpha_n\delta_1 \xrightarrow[n \rightarrow \infty]{} \mu = \delta_0$$

Or  $\mu_n$  ne converge pas vers  $\delta_0$  au sens de  $W_p$ , car  $\int \phi d\mu_n$  converge vers  $\phi(0)$ , et  $\int |x|^p d\mu_n = \alpha_n n^p$ , donc avec  $\alpha_n = 1/n^p$ , on obtient la contradiction voulue.

### 5.3 Lien avec l'équation de conservation de la charge

Nous avons l'habitude d'étudier le champ de vitesses de courbes à valeurs réelles, ce qui permet de dégager de nombreuses propriétés importantes et de mieux comprendre les comportements et les limites des courbes étudiées. Or dans un cadre plus général, par exemple dans les espaces de WASSERSTEIN, on ne peut définir comme à l'usuel la dérivée d'une courbe, l'accroissement  $\gamma(t+h) - \gamma(t)$  n'étant *a priori* pas défini. Cependant il est possible de généraliser la définition de dérivée dans le cas d'espaces métriques généraux de manière naturelle, et cette dérivée généralisée ainsi introduite est également très riche en propriétés et donne des informations importantes sur la courbe.

On introduit alors la *dérivée métrique* de  $\gamma$  en  $t$  :

**Définition.** Soit  $(X, d)$  un espace métrique et soit  $\gamma : [0, 1] \rightarrow X$  une courbe dans  $X$ . La dérivée métrique de  $\gamma$  en  $t$  est

$$|\gamma'| (t) = \lim_{h \rightarrow 0} \frac{d(\gamma(t+h), \gamma(t))}{h}$$

**Théorème.** Si  $(X, d)$  est un espace métrique séparable et  $\gamma : [0, T] \rightarrow X$  est lipschitzienne, alors  $|\gamma'|$  existe presque partout.

On rappelle que tout espace métrique séparable borné peut être immergé dans l'espace  $l^\infty$  des suites bornées. En effet, si  $(x_i)_i$  est une suite dense dans l'espace, qui existe par séparabilité, alors l'application qui à  $x$  associe la suite des mesures de  $x$  aux  $x_i$  est une isométrie, donc une immersion.

La notion de *champ des vitesses* s'étend également au cadre métrique général :

**Définition.** Soit  $(\mu_t)_t$  une courbe lipschitzienne dans  $W_p(\Omega)$ . Un champ de vitesses est un champ de vecteurs  $(v_t)_t$  de  $\mathbb{R}^d$ , avec  $v_t$  intégrable pour  $\mu_t$  pour presque tout  $t$ , qui est solution faible de l'équation de continuité :

$$\partial_t \mu_t + \nabla \cdot (v_t \cdot \mu_t) = 0$$

Autrement dit, qui vérifie

$$\forall \phi \in \mathcal{D}(\Omega), \quad \frac{d}{dt} \int_{\Omega} \phi d\mu_t = \int_{\Omega} v_t \cdot \nabla \phi d\mu_t$$

Pour une famille de mesures  $(\mu_t)_{t \in [0,1]}$  de  $P_p(\Omega)$ , on cherche à trouver des liens entre le caractère lipschitzien des  $\mu_t$  et l'existence d'équations différentielles sur  $\mu : t \mapsto \mu_t$ . L'équation principale vérifiée par  $\mu$  est l'équation de continuité :

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0, \quad v_t : \Omega \rightarrow \mathbb{R}^n$$

**Proposition.** Si  $(v_t)_t$  est une famille de champs de vecteurs uniformément lipschitziennes et bornées, si  $y_x(t)$  est la solution du problème de CAUCHY :

$$\begin{cases} \dot{y}_x(t) = v_t(y_x(t)) \\ y_x(0) = x \end{cases}$$

qui existe par le théorème de CAUCHY — LIPSCHITZ, et si  $\mu_0$  est une mesure de probabilité donnée sur  $\mathbb{R}^n$  et  $\mu_t := Y_t \# \mu_0$  où  $Y_t(x) = y_x(t)$ , alors  $\mu_t$  est solution de l'équation de continuité (au sens des distributions) :

$$\forall \phi \in \mathcal{C}^\infty, \frac{d}{dt} \int \phi d\mu_t = \int v_t \cdot \nabla \phi d\mu_t \text{ pour presque tout } t$$

Ce qui est équivalent à dire que

$$\forall \phi(x, t), \int_0^1 dt \int_\Omega \frac{\partial \phi}{\partial t} d\mu_t + \int_0^1 dt \int_\Omega \nabla \phi \cdot v_t d\mu_t$$

*Preuve.* Fixons une fonction test  $\phi \in \mathcal{C}^\infty$  telle que  $\int_\Omega \phi d\mu_t = \int_\Omega \phi(Y_t(x)) d\mu_0$ . Il vient alors :

$$\begin{aligned} \frac{d}{dt} \int \phi d\mu_t &= \int \nabla \phi(Y_t(x)) \cdot \dot{y}_x(t) d\mu_0 \\ &= \int \nabla \phi(y_x(t)) \cdot v_t(y_x(t)) d\mu_0 \\ &= \int \nabla \phi \cdot v_t d(Y_t \# \mu_0) \\ &= \int \nabla \phi \cdot v_t d\mu_t \quad \square \end{aligned}$$

Nous avons ainsi l'existence et l'explicitation d'une solution du problème de continuité dans ce cas. Il reste à prouver l'unicité de celle-ci :

**Théorème.** Si  $(\mu_t)_t$  est une solution de l'équation de continuité au sens des distributions, et si  $v_t$  est une famille de champs de vecteurs uniformément bornés et uniformément lipschitzien, alors si  $\mu_0$  est nulle, les  $\mu_t$  sont également nulles pour presque tout  $t$ . En particulier, à un  $\mu_0$  de départ, il existe une unique solution  $(\mu_t)_t$  correspondant à cette condition initiale.

*Preuve.* Notons  $X = l^\infty$ . Soit  $\gamma = (\gamma_1, \dots, \gamma_n, \dots)$ , avec  $\gamma_i : [0, T] \rightarrow R$  lipschitzienne. En particulier, ce qui précède permet d'affirmer que les  $\gamma'_i$  existent pour presque tout  $t$ . Notons  $N$  le négligeable hors duquel toutes ces dérivées existent.

Soit  $m(t) = \sup_i |\gamma'_i|(t)$  définie hors de  $N$ , et donc à valeurs réelles. En fait on a

$$|\gamma'(t)| = m(t) \frac{d(\gamma(t+h), \gamma(t))}{|h|} \geq \frac{|\gamma_i(t+h) - \gamma_i(t)|}{|h|} \xrightarrow{h \rightarrow 0} |\gamma'_i(t)|$$

On en tire en particulier que, pour tout  $i$ , la limite inférieure de  $\frac{d(\gamma(t+h), \gamma(t))}{|h|}$  est supérieure aux  $|\gamma'_i|(t)$ , donc à  $m(t)$ . Ainsi :

$$d(\gamma(t+h), \gamma(t)) = \sup_i |\gamma_i(t+h) - \gamma_i(t)|/|h| \leq \sup \int_t^{t+h} |\gamma'_i| \leq \int_t^{t+h} m$$

On se sert désormais des résultats usuels sur les points de LEBESGUE. Pour  $m$  dans  $L^1$ , on a pour presque tout  $t$

$$\frac{1}{2h} \int_{t-h}^{t+h} |m(s) - m(t)| ds \xrightarrow{h \rightarrow 0} 0$$

Donc en particulier  $\int_t^{t+h} m \rightarrow m(t)$ .  $\square$

On s'intéresse désormais aux courbes admettant un champ de vitesses. Le résultat qui suit affirme qu'il s'agit essentiellement des courbes lipschitziennes :

**Théorème.** On suppose  $p > 1$ . Soit  $(\mu_t)_t$  une courbe lipschitzienne définie sur  $[0, T]$  dans  $(P_p(\Omega), W_p)$ . Alors pour presque tout  $t$ , il existe un champ de vecteurs  $v_t$  de  $L^p$  satisfaisant

$$\|v_t\|_p \leq |\mu'(t)| \text{ et } \partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0$$

De plus, si  $(\mu_t, v_t)$  satisfait l'équation de continuité et si  $v_t$  est bornée pour la norme  $L^p$ , alors  $\mu_t$  est une courbe lipschitzienne par rapport à  $W_p$  et  $|\mu'(t)| = \|v_t\|_p$ .

*Preuve.* On remarque que le cas  $p = 1$  pose problème, par exemple avec  $\mu_t = (1-t)\delta_0 + t\delta_1$ . En tous cas, on peut affirmer dans ce cas qu'il existe une famille  $(E_t)_t$  de mesures vectorielles telles que

$$\partial_t \mu_t + \nabla \cdot E_t = 0$$

Pour prouver le théorème, fixons une fonction test  $\phi$  de classe  $C^\infty$  et lipschitzienne. Il faut vérifier que  $t \mapsto \int \phi d\mu_t$  est également lipschitzienne, notre but étant de prouver que

$$\frac{d}{dt} \int \phi d\mu_t = \int \langle \nabla \phi, v_t \rangle d\mu_t \text{ pour un certain } v_t$$

On a dans notre cas :

$$\begin{aligned} \int \phi d\mu_{t+h} - \int \phi d\mu_t &= \text{Lip}(\phi) \int \frac{\phi}{\text{Lip}(\phi)} d(\mu_{t+h} - \mu_t) \leq \text{Lip}(\phi) W_1(\mu_{t+h}, \mu_t) \\ &\leq \text{Lip}(\phi) W_p(\mu_{t+h}, \mu_t) \\ &\leq \text{Lip}(\mu_t) \text{Lip}(\phi) |h| \text{ car la courbe est supposée lipschitzienne} \end{aligned}$$

$$\int \phi d\mu_{t+h} - \int \phi d\mu_t = \int (\phi(y) - \phi(x)) d\gamma_h \text{ où } \gamma_h \in \mathcal{O}(\gamma_t, \gamma_{t+h}).$$

On définit alors

$$H : (x, y) \mapsto \begin{cases} \nabla \phi(x) & \text{si } x = y \\ \left| \frac{\phi x - \phi y}{x - y} \right| & \text{sinon} \end{cases}$$

Ce qui fait que, dans tous les cas, on a  $|\phi(x) - \phi(y)| \leq H(x, y)|x - y|$ . Donc :

$$\left| \int \phi d\mu_{t+h} - \int \phi d\mu_t \right| \leq \int H(x, y) |x - y| d\gamma_h \leq \left( \int |x - y|^p d\gamma_h \right)^{1/p} \left( \int H(x, y)^q d\gamma_h \right)^{1/q}$$

$H$  est semi-continue supérieurement car continue hors de la diagonale, et concernant la diagonale, si  $(x_n, y_n)_n$  converge vers  $(x, x)$ , on a :

$$H(x_n, y_n) = \frac{\phi(x_n) - \phi(y_n)}{x_n - y_n} = \langle \nabla \phi(z_n) | x_n - y_n \rangle / |x_n - y_n| \leq |\nabla \phi(z_n)|$$

où  $z_n$  est un point du segment  $[x_n, y_n]$ . Donc par passage à la limite  $H(x_n, y_n) \leq |\nabla \phi(x)| = H(x, x)$ . Ce qui achève la preuve de la semi-continuité supérieure.

Quitte à extraire, ce qui est possible par compacité, on a  $\gamma_n$  qui converge faiblement vers  $\gamma$ , et donc  $\limsup \int H d\gamma_n \leq \int H d\gamma$ . On écrit alors

$$\begin{aligned} \lim_0 \frac{\int \phi d\mu_{t+h} - \int \phi d\mu_t}{h} &\leq \limsup \frac{W_p(\mu_{t+h}, \mu_t)}{|h|} \left( \int H^q(x, y) d\gamma_h \right)^{1/q} \\ &\leq \limsup \frac{W_p(\mu_{t+h}, \mu_t)}{|h|} \left( \int H^q(x, y) d\gamma \right)^{1/q} \text{ où } \gamma = (id \times id) \# \gamma_h \end{aligned}$$

Par le lemme de continuité de l'image d'une mesure par une application fixée, on a :

$$\gamma_h \text{ optimal pour } \pi(\mu_t, \mu_{t+h}) \implies \gamma \text{ optimal pour } \pi(\mu_t, \mu_t)$$

Ce qui prouve en particulier que

$$\lim_{h \rightarrow 0} \frac{1}{h} \left( \int \phi d\mu_{t+h} - \int \phi d\mu_t \right) \leq |\mu'(t)| \left( \int |\nabla \phi|^q(x) d\mu_t \right)^{1/q}$$

Pour tout fonction  $\phi$  de classe  $C^\infty$  et lipschitzienne, on a pour presque pour tout  $t$  :

$$\frac{d}{dt} \int \phi d\mu_t \leq \|\nabla \phi\|_{L^q(\mu_t)} |\mu'(t)|$$

Soit  $D$  un ensemble dense et dénombrable de fraction  $\phi_k$ . On a :

$$\forall \phi, \exists \phi_k \in D, \nabla \phi_k \xrightarrow{L^q(\bar{\mu})} \nabla \phi \text{ où } \bar{\mu} = (\mu_t dt)$$

Il existe alors un ensemble négligeable  $N$  tel que pour tout  $t$  hors de  $N$ , pour toute fonction  $\phi$  dans  $D$ , on ait :

$$\exists \frac{d}{dt} \int \phi d\mu_t \leq \|\nabla \phi\|_{L^q(\mu_t)} |\mu'(t)|$$

Considérons désormais l'application  $T : \phi \mapsto \frac{d}{dt} \int \phi d\mu_t$ , qui est linéaire en  $\nabla \phi$  car on ne change pas la valeur de l'image en ajoutant une constante, donc la fonction est en réalité une fonction de  $\nabla \phi$ . De plus elle est continue car  $T(\nabla \phi) \leq C \|\nabla \phi\|_{L^q(\mu_t)}$ . E???

La théorème de HAHN — BANACH permet de prolonger  $T$  en  $\overset{\circ}{T} : L^q(\mu_t) \rightarrow \mathbb{R}$ , continue et de même norme.  $\overset{\circ}{T}$  est dans le dual de  $L^q(\mu_t)$ , qui est  $L^p(\mu_t)$ . Ainsi, il existe un champ de vecteurs dans  $L^p(\mu_t)$  tel que :

$$\overset{\circ}{T}(x) = \int v_t(x) d\mu_t$$

On a ainsi trouvé un champ de vecteur  $v_t$  tel que  $\|v_t\|_p = \|\overset{\circ}{T}\|_q$ , et :

$$\forall \phi \in SPAN???( \phi ), \exists \frac{d}{dt} \int \phi d\mu_t = \int v_t \nabla \phi d\mu_t, \|v_t\|_p \leq |\mu'(t)|$$

Cette égalité reste vraie pour tout  $\phi$  de classe  $C^\infty$  par densité.

## 2e PARTIE

Prenons  $(\mu_t, v_t)$  et régularisons  $v_t$  pour la rendre suffisamment régulière :  $\mu_t^\varepsilon = \mu_t \star \rho_\varepsilon$  où  $\rho_\varepsilon$  est une gaussienne, qui est strictement positive, ce qui implique que  $\mu_t^\varepsilon$  est strictement positive. On pose également  $E_t = v_t \mu_t$  et  $E_t^\varepsilon = E_t \star \rho_\varepsilon$  et  $v_t^\varepsilon = E_t^\varepsilon / \mu_t^\varepsilon$ . Dans ce cas :

$$\partial_t \mu_t^\varepsilon + \nabla \cdot (v_t^\varepsilon \mu_t^\varepsilon) = 0$$

et  $v^\varepsilon$  est un champ de vecteur de classe  $C^\infty$  et localement borné. Le théorème d'unicité s'applique donc et affirme que  $\mu_t^\varepsilon = Y_t^\varepsilon \# \mu_0^\varepsilon$ , où  $\gamma_t^\varepsilon(x) = y_x^\varepsilon(t)$  est solution du problème de CAUCHY

$$\begin{cases} \dot{y}_x^\varepsilon = v_t^\varepsilon(y_x^\varepsilon(t)) \\ y_x^\varepsilon(0) = x \end{cases}$$

On peut alors écrire  $W_p(\mu_t^\varepsilon, \mu_s^\varepsilon) \leq \int |Y_t^\varepsilon(x) - Y_s^\varepsilon(x)| d\mu_0^\varepsilon$  où  $(Y_t^\varepsilon, Y_s^\varepsilon) \# \mu_0^\varepsilon \in \pi(\mu_t^\varepsilon, \mu_s^\varepsilon)$ . Il vient alors :

$$\begin{aligned}
\left| \frac{Y_t^\varepsilon(x) - Y_s^\varepsilon(x)}{t-s} \right|^p &\leq \left( \frac{1}{|t-s|} \int_s^t |y_x^\varepsilon(\tau)| d\tau \right)^p \\
&= \frac{1}{|t-s|} \int_s^t |v_t^\varepsilon(y_x^\varepsilon(\tau))| d\tau \\
&= \frac{1}{|t-s|} \int_s^t |v_t^\varepsilon(y_x^\varepsilon(\tau))|^p d\tau d\mu_0^\varepsilon \\
&= \frac{1}{|t-s|} \int_t^s \int |v_t^\varepsilon(y_x^\varepsilon(\tau))|^p d\mu_0^\varepsilon d\tau
\end{aligned}$$

et  $\int |v_t^\varepsilon(y_x^\varepsilon(\tau))|^p d\mu_0^\varepsilon = \int |v_\tau^\varepsilon(y)|^p d\mu_\tau^\varepsilon(y) = \|v_{tau}^\varepsilon\|_{L^p(\mu_t^\varepsilon)}^p$ .

Puis pour finir :

$$\frac{W_p(\mu_t^\varepsilon, \mu_s^\varepsilon)}{t-s} \leq \left( \frac{1}{t-s} \int_t^s \|v_\tau^\varepsilon\|_{L^p(\mu_\tau^\varepsilon)}^p d\tau \right)^{1/p}$$

Or si on définit la fonction

$$f(a, b) = \begin{cases} \frac{|b|^p}{pa^{p-1}} & \text{si } a > 1 \\ 0 & \text{si } a = b = 0 \\ +\infty & \text{sinon} \end{cases}$$

$f$  est convexe et semi-continue inférieurement car  $f(a, b) = \sup \{at + b \cdot w \mid (t, w) \in \mathbb{R} \times \mathbb{R}^n, t + |w|^q/q \leq 0\}$ .  
L'inégalité de JENSEN donne pour toute fonction  $f$  convexe :

$$f\left(\int \alpha d\rho\right) \leq \int f(\alpha) d\rho$$

On l'applique en particulier au noyau de convolution :

BLA BLA BLA

On obtient alors, en fin de compte, une majoration par :

$$(\dots) \leq \left( \int f(\mu_t, E_t) dt \right)^{1/p} = \left( \frac{1}{t-s} \int \|v_\tau\|_p^p d\tau \right)^{1/p}$$

On sait que  $\|v_\tau\|$  est majorée par  $C$ , donc la distance est majoré par  $C|t-s|$ , donc on a la même inégalité lorsque l'on fait tendre  $\varepsilon$  vers 0. Ainsi pour tout point de LEBESGUE :

$$\frac{W_p(\mu_t^\varepsilon, \mu_s^\varepsilon)}{t-s} \underset{s \rightarrow t}{\leq} \|v_t\|_p$$

On a ainsi majoré la dérivée métrique par la norme  $\|\cdot\|$ .  $\square$

$$|\mu'(t)| \leq \|v_t\|_{L^p(\mu_t)}$$

## 5.4 Quelques « rappels »

### 5.4.1 Chemins et longueurs en espace métrique

On se place dans un espace métrique général  $(X, d)$ . Une *courbe* est une application continue d'un segment  $[a, b]$  dans cet espace, on se limite donc habituellement au segment  $[0, 1]$  par reparamétrisation. On définit sa *longueur* par

$$l_d : \gamma \mapsto \sup \left\{ \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) \mid 0 = t_0 < t_1 < \dots < t_n = 1 \right\}$$

Dans le cas d'une courbe *lipschitzienne*, la longueur est toujours finie et majorée par la constante de LIPSCHITZ de l'application. Dans ce cas, et dans le cas plus général où la dérivée métrique existe, on a

$$\forall \gamma, l_d(\gamma) = \int_0^1 |\gamma'(t)| dt$$

N.B. : La dérivée métrique de  $\gamma$  est aussi égale à  $|\gamma'| : t \mapsto \sup_k |d(\gamma(t), x_k)'|$  si la suite  $(x_k)_k$  est une suite dense dans l'espace. Il s'agit donc d'une propriété valable dans les espaces séparables.

### 5.4.2 Géodésiques

Une *géodésique* est une courbe  $\gamma$  minimisant la longueur parmi toutes les courbes à deux extrémités fixées. Plus précisément, si on impose  $\gamma(0) = x$  et  $\gamma(1) = y$ , on recherche à minimiser la longueur d'un chemin de  $x$  à  $y$ . On peut donc se limiter aux courbes lipschitziennes. Notons que l'on peut d'ailleurs toujours reparamétriser une courbe pour la rendre lipschitzienne (**WHY?**).

Les géodésiques sont donc les courbes  $\gamma_0$  vérifiant

$$l_d(\gamma_0) = \min \left\{ \int_0^1 |\gamma_0'(t)| dt \mid \gamma_0 \in Lip([0, 1], X), \gamma_0(0) = x, \gamma_0(1) = y \right\}$$

Dans le cas qui nous intéresse, nous n'aurons pas à nous soucier de l'existence de géodésiques : on les connaît explicitement !

### 5.4.3 Espaces de longueurs

$(X, d)$  est un *espace de longueurs* si

$$\forall x, y \in X, d(x, y) = \inf \{ l_d(\gamma) \mid \gamma \in Lip([0, 1], X), \gamma(0) = x, \gamma(1) = y \}$$

À toute distance  $d$ , on peut associer une *distance géodésique*  $d_g$  par cette formule. En considérant la partition triviale de  $[0, 1]$  par lui-même, on obtient que la distance géodésique est toujours supérieure à la distance. Un espace est un espace de longueurs si, et seulement si, on a l'égalité entre la distance et la distance géodésique associée. Par ce qui précède, il suffit donc de prouver l'inégalité  $d_g \geq d$ .

$(X, d)$  est un *espace géodésique* si c'est un espace de longueurs et si les distances géodésiques sont atteintes, autrement dit si les chemins optimaux existent bel et bien *dans* l'ensemble des chemins sur cet espace.

Une courbe  $\gamma$  est dite à *vitesse constante* si sa dérivée métrique est constante presque partout, ce qui traduit bien, dans le cas de dérivées usuelles, la notion de vitesse constante.

### 5.4.4 Une caractérisation des géodésiques à vitesse constante

Si  $p > 1$  et si on cherche à minimiser  $\int_0^1 |\gamma'|^p$ , les minimiseurs sont les géodésiques à vitesse constante. En effet, par l'inégalité de JENSEN, on a :

$$\int_0^1 |\gamma'|^p \geq \left( \int_0^1 |\gamma'| \right)^p$$

et il y a égalité si, et seulement si,  $\gamma$  est un chemin à vitesse constante, par le cas d'égalité de l'inégalité de JENSEN pour les fonctions strictement convexes, puisque la fonction  $x \mapsto |x|^p$  est strictement convexe.

Si  $\bar{\gamma}$  est une géodésique, on a sa longueur qui est égale à la distance géodésique de  $x$  à  $y$ . Donc si sa vitesse est constante, elle vaut nécessairement la distance géodésique entre ces deux points, d'où

$$\int |\gamma'|^p \geq d_g(x, y)^p = \left( \int_0^1 d_g(x, y) \right)^p = \left( \int |\gamma'| \right)^p$$

**Théorème.** Dans un espace géodésique, une courbe  $\gamma$  est une géodésique à vitesse constante si, et seulement si, on a  $d(\gamma(t), \gamma(s)) = d(x, y)|t - s|$ .

( $\Leftarrow$ ) Tout d'abord, on a

$$l_d(\gamma) = \sup \sum_{i=0}^{i=n-1} d(\gamma(t_i), \gamma(t_{i+1})) = \sup \sum_{i=0}^{i=n-1} d(x, y)|t_{i+1} - t_i| = d(x, y)$$

donc  $\gamma$  est une géodésique, et de plus :

$$\frac{d(\gamma(t), \gamma(s))}{|t - s|} = d(x, y)$$

donc

$$|\gamma'(t)| = \lim_{s \rightarrow t} \frac{d(\gamma(t), \gamma(s))}{|t - s|} = d(x, y)$$

$\gamma$  est donc bien à vitesse constante.

( $\Rightarrow$ ) Pour tout  $t, s \in [0, 1]$ , on a

$$d(\gamma(t), \gamma(s)) \leq \int_t^s |\gamma'| = d(x, y)|s - t|$$

donc si  $0 \leq s \leq t \leq 1$  alors

$$\begin{aligned} d(\gamma(0), \gamma(s)) &\leq d(x, y)s \\ d(\gamma(s), \gamma(t)) &\leq d(x, y)|t - s| \\ d(\gamma(t), \gamma(1)) &\leq d(x, y)|1 - t| \end{aligned}$$

En sommant, on obtient que  $d(\gamma(0), \gamma(1)) \leq d(x, y)$  donc les inégalités sont toutes des égalités, donc  $d(\gamma(s), \gamma(t)) = d(x, y)|t - s|$ .

## 5.5 Géodésiques et transport optimal

On se place désormais dans l'espace  $(P_p(\Omega), W_p)$ , et on fixe deux mesures  $\mu$  et  $\nu$  de cet espace. L'existence d'un transport optimal de  $\mu$  à  $\nu$  correspond à l'idée intuitive de minimiser le « trajet » qu'il y a à faire pour passer de  $\mu$  à  $\nu$ , ce qui correspond aussi à l'idée intuitive que l'on peut se faire d'une géodésique de entre  $\mu$  à  $\nu$ . Cette intuition est en réalité justifiée, et c'est ce que nous allons voir par la suite.

### 5.5.1 Géodésiques et plans de transferts optimaux

**Théorème.** Si  $\Omega$  est convexe et s'il existe un plan de transfert optimal  $\gamma$  de  $\mu$  à  $\nu$  dérivant d'un transport  $T$ , alors  $\mu_t = ((1-t)id + tT) \# \mu$  est une géodésique de  $\mu$  à  $\nu$ .

Si l'on n'a pas d'hypothèse sur  $\Omega$  ni sur l'existence de  $T$ , alors il existe un plan de transfert optimal  $\gamma$  tel que  $\pi_t \# \gamma$  (?) est une géodésique, où  $\pi_t(x, y) = (1-t)x + ty$ .

Autrement dit, on peut toujours trouver une géodésique qui est la projection d'un plan de transfert optimal  $\gamma$  sur un  $\pi_t$ .

Par le théorème précédent, il suffit de prouver que  $W_p(\mu_t, \mu_s) = W_p(\mu, \nu)|t-s|$ , donc l'inégalité  $\leq$  suffit. Il nous suffit donc de trouver un plan de plan de transfert entre  $\mu_t$  et  $\mu_s$ . Le plan  $(\pi_t, \pi_s) \# \gamma$  convient car  $\gamma$  est un transport optimal, donc

MARGINALES

Puis on a

$$W_p(\mu_t, \mu_s) \leq \int |x-y|^p d((\pi_t, \pi_s) \# \gamma) = \int |\pi_t - \pi_s|^p d\gamma = \int |t-s|^p |x-y|^p d\gamma = |t-s|^p W_p^p(\mu, \nu)$$

donc  $W_p(\mu_t, \mu_s) \leq W_p(\mu, \nu)|s-t|$ . La première inégalité provient de l'optimalité de la distance de WASSERSTEIN : on a ici effectué le transport entre les instants  $t$  et  $s$  seulement, mais pas en entier. Cela ne change pas son caractère optimal car s'il ne l'était pas, il suffirait de le rendre optimal sur cette portion et de le recoller au reste inchangé pour obtenir un transport strictement plus efficace.  $\square$

On a alors déjà vu que l'on pouvait trouver un champ de vecteurs  $(v_t)_t$  vérifiant l'équation de continuité :

$$\forall t \in [0, 1], \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0$$

Si  $T_t = (1-t)id + tT$ , alors on a  $v_t(x) = T(y) - y$  avec  $y$  un antécédent de  $x$  par  $T$ . Autrement dit,  $v_t = (T - id) \circ T_t^{-1}$ . (???)

Remarquons que  $v_t$  correspond à la vitesse des particules dans une région considérée *au temps*  $t$ . On ne suit donc pas le comportement d'une particule bien déterminée, mais celle de toutes les particules circulant dans une région de l'espace bien déterminée. Il s'agit donc d'une description *lagrangienne* du problème, et non *eulérienne*.

PARTIE CCM A ECRIRE

Pour  $p = 1$ , les rayons de transport ne se croisent pas. Pour  $p > 1$ , la condition est beaucoup plus faible : ils ne peuvent pas de croiser *au même instant*, mais les traces du mouvement, autrement dit le graphe des rayons de transport considérés comme fonctions du temps  $et$  de l'espace, peuvent se croiser.

### 5.5.2 Géodésiques à vitesses constantes et plans de transferts optimaux

**Proposition.** On suppose que  $p > 1$ . Si  $\mu_t$  est une géodésique à vitesse constante et si  $t \in ]0, 1[$ , alors il existe un unique plan de transfert optimal  $\gamma_-$  de  $\mu_0$  à  $\mu_t$ , et il dérive d'un transport. De même, il existe un unique plan de transfert optimal  $\gamma_+$  de  $\mu_t$  à  $\mu_1$ , et il vient aussi d'un transport.

Le grand intérêt de cette proposition est de pouvoir ensuite composer les deux plans de transfert ainsi obtenus, à la manière du *gluing lemma*, pour obtenir un transport optimal de  $\mu$  à  $\nu$ .

On rappelle le *gluing lemma* : (...)

Soient  $\gamma_-$  et  $\gamma_+$  optimaux respectivement pour  $(\mu_0, \mu_t)$  et  $(\mu_t, \mu_1 = \nu)$ . Soit  $\gamma$  (noté  $\sigma$  dans le *gluing lemma*) obtenu par recollement. Il vient alors

$$\begin{aligned}
W_p(\mu, \nu) &\leq \left( \int |x - y|^p d\gamma \right)^{1/p} = \|x - z\|_{L^p(\sigma)} \\
&\leq \|x - z\|_{L^p(\sigma)} + \|y - z\|_{L^p(\sigma)} \\
&\leq \left( \int |x - y|^p d\gamma_- \right)^{1/p} + \left( \int |y - z|^p d\gamma_+ \right)^{1/p}
\end{aligned}$$

Ce qui donne

$$W_p(\mu, \nu) \leq W_p(\mu, \mu_t) + W_p(\mu_t, \nu) = tW_p(\mu, \nu) + (1-t)W_p(\mu, \nu) = W_p(\mu, \nu)$$

Il y a donc égalité partout et donc  $\gamma$  est un plan de transfert optimal de  $\mu$  à  $\nu$ . Comme la norme  $\|\cdot\|_{L^p}$  est strictement convexe, car  $p > 1$ , on a

$$x - y = \lambda(x - z) \text{ } \sigma\text{-presque partout et } y_z = (1 - \lambda)(x - z) \text{ } \sigma\text{-presque partout}$$

On a en fait utilisé le fait suivant : si  $\|f + g\|_p = \|f\|_p + \|g\|_p$ , alors  $f, g$  et  $f + g$  sont multiples l'un de l'autre. Autrement dit, on peut écrire pour un certain  $t$  :  $f + t(f + g)$  et  $g = t'(f + g)$  avec  $t$  et  $t'$  positifs. Cela implique en particulier que  $t + t' = 1$ , donc que  $t' = 1 - t$ .

Si on intègre « à la  $L^p$  » les égalités ainsi obtenues, on obtient

$$\|x - y\|_p = \lambda\|x - z\|_p = \lambda W_p(\mu, \nu) = W_p(\mu, \mu_t) = tW_p(\mu, \nu)$$

ce qui implique que  $t = \lambda$ . Ainsi, on a trouvé un  $\gamma$  optimal et  $\mu_t = \pi_t \# \gamma$  car  $\mu_t = \pi_y \# \sigma$  et  $y = (1 - t)x + tz$  par les relations précédentes. Il reste à prouver que les plans de transfert optimaux considérés sont des transports.

Mais le  $\gamma$  obtenu dépend *a priori* de  $t$  ! L'égalité  $\mu_t = \pi_t \# \gamma$  n'est donc vraie, *a priori*, qu'à l'instant  $t$  fixé au début du théorème. Cette dépendance est gênante et ne permet pas de construire un recollement « canonique » entre  $\gamma_+$  et  $\gamma_-$ .

Montrons que  $\gamma_-$  vient d'un transport. Soient  $(x, y)$  et  $(x, y')$  dans le support de  $\gamma_-$ . On veut prouver que  $y = y'$ , autrement dit que la relation est fonctionnelle. Il y aurait alors  $z$  et  $z'$  tels que  $y = (1 - t)x + tz$  et  $y' = (1 - t)x' + tz'$ , avec  $(x, y, z)$  et  $(x', y, z')$  qui sont dans le support de  $\sigma$ . Il vient alors que  $(x, z)$  et  $(x', z')$  est dans le support de  $\gamma$ . Or, les rayons de transport ne peuvent se croiser *au même instant* !  $\square$

Comme habituellement, on prouve l'unicité de ce plan de transport optimal en raisonnant avec des combinaisons convexes.

**Propriété.** On suppose que  $p > 1$ . Soit  $\mu_t$  une géodésique à vitesse constante. Alors il existe un plan de transfert optimal  $\gamma$ , et  $\mu_t = \pi_t \# \gamma$ .

On fixe  $t = \frac{1}{2}$ . Maintenant, on sait que le plan de transfert optimal de  $\mu_{1/2}$  à  $\mu$  est unique. Notons  $\nu_s = \mu_{s/2}$  une géodésique à vitesse constante de  $\mu$  à  $\mu_{1/2}$ . Si  $s \in ]0, 1[$ , on sait que  $\nu_s = \pi_s \# \hat{\gamma}$  où  $\hat{\gamma}$  est le seul plan de transfert optimal de  $\mu$  à  $\mu_{1/2}$ , par ce qui précède. On note aussi  $\tilde{\gamma}$  l'unique plan de transfert optimal de  $\mu_{1/2}$  à  $\nu$ , et  $\bar{\gamma}$  le recollement de ces deux plan de transfert.

Puisque  $y = \frac{1}{2}x + \frac{1}{2}z$   $\sigma$ -presque partout, car  $t = \frac{1}{2}$ . On a alors

$$\begin{aligned}
\nu_s &= \pi_s \# \hat{\gamma} = \pi_s \# \sigma \\
&= \left( (1 - s)x + \frac{s}{2}x + \frac{s}{2}z \right) \# \sigma \\
&= \left( \left(1 - \frac{s}{2}\right)x + \frac{1}{2}z \right) \# \bar{\gamma} \\
&= \pi_{s/2} \# \bar{\gamma}
\end{aligned}$$

Ainsi, on a prouvé que chercher les géodésiques revenait à rechercher les plans de transferts optimaux, conformément à ce que nous suggérait l'intuition géométrique.

Par contre pour  $p = 1$ , tout cela est faux. En effet, avec l'exemple usuel  $\mu_t = t\delta_0 + (1-t)\delta_1$ , on a une géodésique qui n'est pas obtenue par interpolation de la position des particules, sinon on aurait nécessairement  $\mu_t = \delta_t$ , mais ici la masse est transportée partiellement en 0 et en 1 : c'est un transfert continu de masse d'un point à un autre.

## 5.6 Convexité géodésique et fonctionnelles

### 5.6.1 Généralisation de la notion de convexité

Dans un espace géodésique, on ne peut plus utiliser la définition usuelle de la convexité, les combinaisons linéaires convexes n'étant plus définies. On définit alors une fonction  $F$  comme étant *géodésiquement convexe* comme une fonction vérifiant

$$\forall (x, y) \in \Omega^2, \forall \gamma \text{ géodésique à vitesse constante de } x \text{ à } y, F(\gamma(t)) \leq (1-t)F(x) + tF(y)$$

Un espace  $Y$  est dit *géodésiquement convexe* si pour tout couple de points de  $Y$ , toute géodésique à vitesse constante liant ces deux points prend ses valeurs dans  $Y$ . Autrement dit, on définit la convexité comme la convexité usuelle le long des géodésiques à vitesse constante sur l'espace.

### 5.6.2 Quelques fonctionnelles centrales

Dans  $W_p(\Omega)$  avec  $\Omega$  convexe, on considère les trois classes de fonctionnelles suivantes, représentant des *énergies* intervenant de manière assez générale :

$$\begin{aligned} J_V^1 : P_p(\Omega) \ni \mu &\longmapsto \int_{\Omega} V d\mu \text{ où } V : \Omega \longrightarrow \mathbb{R} \cup \{\infty\} \\ J_W^2 : P_p(\Omega) \ni \mu &\longmapsto \int_{\Omega \times \Omega} W(x-y) d\mu(x) d\mu(y) \\ J_U^3 : P_p^a(\Omega) \ni \mu &\longmapsto \int_{\Omega} U(p(x)) \quad U(0) = 0 \end{aligned}$$

où  $P_p^a(\Omega)$  représente l'ensemble des mesures de  $P_p(\Omega)$  absolument continues par rapport à la mesure de LEBESGUE. On observe que  $P_p^a \subset P_p$  est géodésiquement convexe. En effet, soient  $\mu$  et  $\nu$  deux mesures de  $P_p^a(\Omega)$ , et soit  $(\mu_t)_{t \in [0,1]}$  une géodésique à vitesse constante entre elles. On sait, pour  $p > 1$ , que  $\mu_t = \pi_t \# \gamma$  pour un certain plan de transfert optimal  $\gamma \in \pi(\mu, \nu)$ . Puisque  $\mu$  est absolument continue par rapport à  $\lambda$ , ce  $\gamma$  est unique et dérive d'un transport. Dans ce cadre, on a aussi prouvé que  $\mu_t$  est absolument continue par rapport à  $\lambda$  pour  $t \in ]0, 1[$  lors de l'étude de la densité de transport  $\sigma_{\gamma}$ <sup>1</sup>.

Notons quelques cas particuliers de la fonctionnelle  $J_U^3$  : si  $U$  est linéaire, cela correspond à la masse de la mesure ; si  $U$  est quadratique, c'est une fonctionnelle qui donne plus d'importance aux concentrations de masse.

### 5.6.3 De la convexité des fonctionnelles

#### Théorème.

- \*  $V$  convexe  $\iff J_V^1$  géodésiquement convexe
- \*  $W$  convexe  $\implies J_W^2$  géodésiquement convexe
- \*  $U$  convexe et  $s \mapsto U(s^{-d})s^d$  convexe décroissante avec  $d = \dim(E) \implies J_U^3$  géodésiquement convexe

Un cas très utile de cette dernière assertion est  $U(t) = t^p$  pour  $p > 1$ .

*Preuve.* Détaillons séparément le preuve pour chacune des trois fonctionnelles. On considère à chaque fois une géodésique à vitesse constante entre  $\mu$  et  $\nu$ , qui s'écrit  $\mu_t = \pi_t \# \gamma$  par théorème.

- \* Dans le cas de  $J_V^1$ , on écrit simplement, en supposant  $V$  convexe :

<sup>1</sup>. La preuve a été faite dans le cas  $p = 1$  par approximation de la mesure  $\nu$  par des mesures atomes  $\nu_n$  obtenues en concentrant la masse en un point de chaque région d'une subdivision de plus en plus fine. La même preuve tient dans le cas  $p > 1$ .

$$\begin{aligned}
J_V^1(\mu_t) &= \int \int V d\mu_t = \int \int V(\pi_t(x, y)) d\gamma \\
&= \int \int V((1-t)x + ty) d\gamma \\
&\leq \int \int ((1-t)V(x) + tV(y)) d\gamma \text{ par convexité de } V \\
&= (1-t)J_V^1(\mu) + tJ_V^1(\nu)
\end{aligned}$$

ce qui prouve que  $J_V^1$  est géodésiquement convexe. La réciproque est également vraie en introduisant des mesures de DIRAC.

\* Dans le cas de  $J_W^2$ , on écrit simplement, en supposant  $W$  convexe :

$$\begin{aligned}
J_W^2(\mu_t) &= \int \int W(z - z') d\mu_t^2(z, z') = \int \int W(\pi_t(x, y) - \pi_t(x', y')) d\gamma \\
&= \int \int W((1-t)(x - x') + t(y - y')) d\gamma \\
&\leq \int \int ((1-t)W(x - x') + tW(y - y')) d\gamma \text{ par convexité de } W \\
&= (1-t)J_W^2(\mu) + tJ_W^2(\nu)
\end{aligned}$$

ce qui achève la preuve de la convexité de  $W$ . Notons d'ailleurs que la preuve est la même dans le cas un peu plus général, bien que moins courant en pratique, d'une fonction  $W$  de deux variables  $W(x, x')$  vérifiant l'hypothèse de convexité par rapport au couple  $(x, x')$ . De plus, en dimension 1, il suffit que  $W$  soit convexe sur les deux demi-plans délimités par la première bissectrice car le transport dont dérive  $\gamma$  est croissant et stabilise donc ces deux demi-plans.

\* Notons  $\rho$  la densité de  $\mu$ . Pour une application  $s$  injective et lipschitzienne (ce qui implique la différentiabilité presque partout d'après le théorème de RADEMACHER), par changement de variable on a

$$s\#\rho = \frac{\rho \circ s^{-1}}{|\det(\nabla s)|}$$

Les rayons de transport ne peuvent se croiser au même instant par application de la propriété de  $c$ -monotonie cyclique, ce qui fait que les applications  $T_t$  sont injectives. La convexité implique la lipschitzianité dénombrable du gradient (ce qui suffit à avoir le résultat précédent). On a déjà vu que les  $\mu_t$  sont des mesures à densité par rapport à  $\lambda$ , et le résultat qui précède donne alors

$$\rho_t(y) = \frac{\rho \circ T_t^{-1}(y)}{|\nabla T_t(T_t^{-1}(y))|}$$

Il vient alors par changement de variables :

$$J_U^3(\mu_t) = \int U \left( \frac{\rho}{|\nabla T_t|} \circ T_t^{-1}(x) \right) dx = \int U \left( \frac{\rho}{|\nabla T_t|} |\nabla T_t(y)| \right) dy$$

Prouvons d'abord le

**Lemme.**  $\forall A \in S_n(\mathbb{R}), (A \geq -I \implies [0, 1] \ni t \mapsto |I + tA|^{1/d}$  est concave).

*Preuve.* Les matrices symétriques réelles étant diagonalisables, il suffit de prouver le résultat pour les matrices diagonales, celui-ci étant invariant par changement de base car

$$\forall t \in [0, 1], \forall P \in \mathcal{GL}_n(\mathbb{R}), |I + t(PAP^{-1})| = |P(I + tA)P^{-1}| = |I + tA|$$

On se contente donc de raisonner sur les éléments matrices diagonales  $A = (\lambda_i)_i$  :

$$f : t \mapsto \prod_i (1 + t\lambda_i)^{1/d}, f' : t \mapsto \frac{1}{d} \sum_{i=1}^n \lambda_i (1 + t\lambda_i)^{-1} \prod_{j=1}^n (1 + t\lambda_j)^{1/d} = g(t)f(t)$$

où  $g : t \mapsto \frac{1}{d} \sum_{i=1}^n \lambda_i (1 + t\lambda_i)^{-1}$ . On écrit alors simplement

$$f' = fg, f'' = f'g + fg' = fg^2 + fg' = f(g^2 + g')$$

ce qui prouve la concavité puisque  $g' \leq -g^2$ .  $\square$

Dans notre cas, on considère  $f : t \mapsto \rho^{-1/d} |I + tA|^{1/d}$  où  $A = H\psi - I$  et  $T = \nabla\psi$ , et la fonction XXX est bien concave... ???  $\square$

### 5.6.4 Inégalité de Brunn - Minkowski

Si  $X$  et  $Y$  sont des parties de  $E$ , alors

$$\lambda(X + Y)^{1/d} = \lambda(X)^{1/d} + \lambda(Y)^{1/d}$$

*Preuve.* On restreint les mesures aux espaces  $X$  et  $Y$  considérés :

$$\mu \leftarrow \frac{1}{|X|} \lambda|_X, \nu \leftarrow \frac{1}{|Y|} \lambda|_Y$$

On pose  $U(s) = -s^{1-1/d}$  qui est convexe et décroissante comme déjà signalé. Le théorème précédent s'applique, notamment avec  $t = \frac{1}{2}$ , donc le support  $A$  est contenu dans  $\frac{1}{2}(X + Y)$ , d'où

$$-\int \rho_{1/2}^{1-1/d} \leq -\frac{1}{2} |X| |X|^{-1+1/d} - |Y| |Y|^{-1+1/d} = -\frac{1}{2} (|X|^{1/d} + |Y|^{1/d})$$

L'inégalité de JENSEN affirme que

d'où :

$$\frac{1}{2} |X + Y|^d \geq \left| \frac{X + Y}{2} \right|^{1/d} \geq |A|^{1/d} \geq \frac{1}{2} |X|^{1/d} + \frac{1}{2} |Y|^{1/d}$$

ce qui donne  $|X + Y|^d \geq |X|^d + |Y|^d$ .  $\square$

Une conséquence immédiate est l'*inégalité isopérimétrique*, en prenant  $Y = B^d(0, 1)$  et en notant  $\omega_d$  son volume. On a alors

$$|X_r|^{1/d} \geq |X| + \omega_d^{1/d} r \text{ où } \omega_d = \lambda(B^d(0, 1)) \text{ et } X_r = X + rB^d(0, 1)$$

d'où, en écrivant un développement limité à l'ordre 1 du volume :

$$\text{Vol}(X_r \text{ (ou } X \text{ ?)}) \sim (\text{Per}(X)r + |X|)^{1/d} \geq \omega_d^{1/d} r$$

d'où :

$$|X|^{1/d} + d|X|^{1/d-1} \text{Per}(X)r \geq |X|^{1/d} + \omega_d^{1/d} r$$

d'où :

$$\forall X, \text{Per}(X) \geq \frac{1}{d} \omega_d^{1/d} |X|^{1-1/d}$$

car...

### 5.6.5 Autre application

Dans le cas où  $\Omega$  est borné et  $\mu, \nu$  sont bornées, alors les normes  $L^\infty$  des  $\mu_t$  sont bornées par le maximum des normes de  $\mu$  et de  $\nu$ . En effet, le résultat est connu pour les normes  $L^p$ , et on peut passer à la limite puisque  $\Omega$  est de masse totale finie, car il est borné.

Si on revient à  $\sigma_\gamma \leq \int_0^1 \mu_t dt$  et si  $\mu, \nu$  sont dans  $L^q$ , alors ( $\|\mu_t\|_q \leq C \implies \sigma \in L^q$ ). Mais ce résultat n'est pas valable pour  $p \neq 2$ , il faudrait retrouver les géodésiques convexes à la main.

## Chapitre 6

# Application à des problèmes d'équations aux dérivées partielles

Précédemment, on a vu que l'on pouvait travailler sur des espaces métriques adaptés aux problèmes de transport, notamment  $(P(\Omega), W_p)$  si  $\Omega$  est compact, et  $(P_p(\Omega), W_p)$  de manière générale. Nous allons maintenant rechercher les propriétés et les expressions des géodésiques dans ces espaces.

### 6.1 Bénomou – Brenier

#### 6.1.1 Le problème de la minimisation de la longueur

On cherche désormais à minimiser  $\int_0^1 |\mu'(t)| dt$ , à  $\mu_0$  et  $\mu_1$  données. On sait que

$$\int_0^1 |\mu'(t)| dt \leq \min \int \|v_t\|_{L^p(\mu_t)}^p dt \text{ où } \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0$$

car pour toute solution de l'équation de continuité, on a  $|\mu'|^p \leq \|v_t\|_{L^p(\mu_t)}^p$  et l'existence d'un  $v_t$  réalisant l'égalité. Donc il y a égalité, par le théorème du cours précédent.

On travaille donc désormais sur le problème de la minimisation du critère

$$\int \|v_t\|_{L^p(\mu_t)}^p dt \text{ où } \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0$$

Ce problème de minimisation sur les couples  $(\mu_t, v_t)$  est d'une formulation peu exploitable, car les contraintes ne sont pas linéaires et les fonctions mises en jeu ne sont pas convexes en les inconnues, il n'y a donc pas de propriété de convexité pour l'ensemble des solutions, et les méthodes de recherche et de démonstration usuelles en analyse convexe ne s'appliqueront pas.

#### 6.1.2 La formulation de Bénomou – Brenier du problème

On cherche donc à reformuler le problème d'une manière plus judicieuse. Notons  $E_t = v_t \mu_t$ . On ramène ainsi le problème à une optimisation convexe du critère

$$\frac{1}{p} \int \int \left| \frac{E_t}{\mu_t} \right|^p d\mu_t \text{ où } \partial_t \mu_t + \nabla \cdot E_t = 0$$

On introduit alors

$$f : (s, w) \mapsto \begin{cases} \frac{|w|^p}{ps^{p-1}} & \text{si } s > 0 \\ 0 & \text{si } s = 0, w = 0 \\ +\infty & \text{sinon} \end{cases}$$

La fonction  $f$  ainsi définie est convexe et 1-homogène. Plus précisément, posons

$$\forall s, w, g(s, w) = \sup \{as + b \cdot w \mid a \in \mathbb{R}, b \in \mathbb{R}^n, a + \frac{1}{q}|b|^q \leq 0\}$$

$g$  est convexe et 1-homogène, car c'est un supremum de fonctions linéaires. Si  $s$  est positif, on peut réduire les hypothèses :

$$\forall b, \max \left\{ a \mid a + \frac{1}{q}|b|^q \leq 0 \right\} = -\frac{1}{q}|b|^q$$

donc

$$\forall s, w, g(s, w) = \sup_{b \in \mathbb{R}^n} \left( -\frac{s}{q}|b|^q + b \cdot w \right)$$

Le sup est atteint si  $b$  et  $w$  sont colinéaires et que  $|w| = s|b|^{q-1}$ . D'où  $f = g$ . Par suite :

### CHAINE IMMORDE D'EGALITES

On change alors les inconnues du problème pour se ramener à une formulation convexe. Le nouveau problème consiste à minimiser

$$\int \int f(\mu_t, E_t) \text{ où } \partial_t \mu_t + \nabla \cdot E_t = 0$$

On se ramène alors à minimiser

$$\sup \left\{ \int (a\psi_t + b \cdot E_t) \mid a + \frac{1}{q}|b|^q \leq 0, \partial_t \mu_t + \nabla \cdot E_t = 0 \right\}$$

C'est la formulation de BRENIER – BÉNAMOU du problème.

## 6.2 Un peu de dualité...

### 6.2.1 Reformulations du problème et simplification des critères

On absorbe comme usuellement la contrainte en introduisant la fonction  $G : \phi \mapsto \int \phi(1, x)d\mu_1 - \int \phi(0, x)d\mu_0$ . Les solutions de l'équation de continuité vérifient, en traduisant l'équation au sens des distributions :

$$\sup_{\phi} [\phi d\mu]_0^1 - \int \partial_t \phi \mu - \int \nabla \phi \cdot E_t = \begin{cases} 0 & \text{si les contraintes sont satisfaites} \\ +\infty & \text{sinon} \end{cases}$$

Le nouveau critère de minimisation devient alors

$$\min_{\substack{\mu \in P_p(\Omega), E_t = v_t \mu_t \\ \partial_t \mu_t + \nabla \cdot E_t = 0}} \sup_{\substack{a(t, x), b(t, x), \phi \in C^1(\Omega) \\ a + \frac{1}{q}|b|^q \leq 0}} \int \int a d\mu + \int \int b dE - \int \int \partial_t \phi d\mu - \int \nabla \phi \cdot E + G(\phi)$$

Comme toujours, on cherche à savoir si ce problème est le même que le problème dual, obtenu en échangeant les bornes inférieures et supérieures. On a toujours  $\inf \sup \geq \sup \inf$  comme déjà mentionné. Un théorème d'analyse convexe dû à ROCKAFELLAR affirme que l'on a l'interchangeabilité si, et seulement si, on a l'existence d'un *point selle*  $(x^*, y^*)$ , i.e. d'un point vérifiant

$$\begin{aligned} \forall x \quad , \quad F(x, y^*) &\geq F(x^*, y^*) \\ \forall y \quad , \quad F(x^*, y) &\leq F(x^*, y^*) \end{aligned}$$

On pose alors  $m = (\mu, E)$  et  $\xi = (a, b)$  pour alléger les notations. Le produit considéré est le produit scalaire  $L^2$  sur l'espace-temps. On introduit alors

$$K = \{\xi \mid a + \frac{1}{q}|b|^q \leq 0\}$$

qui est un ensemble convexe de  $\mathbb{R}^{n+1}$ . On introduit l'indicatrice  $I_K$  de  $K$  qui vaut 0 si  $\xi$  est presque partout dans  $K$ , et  $+\infty$  sinon. Le critère se réécrit alors

$$\min_m \sup_{\xi, \phi} m \cdot \xi - \nabla_{t,x} \phi \cdot m + G(\phi) - I_k(\xi)$$

Si  $m$  est optimal, alors  $\xi - \nabla_{t,x} \phi = 0$ . On peut donc ajouter le carré de cette quantité, ce qui n'influera pas sur l'optimum. Cette *méthode du lagrangien augmenté* permet de transformer le critère en

$$\min_m \sup_{\xi, \phi} m \cdot \xi - \nabla_{t,x} \phi \cdot m + G(\phi) - I_k(\xi) + \|\xi - \nabla_{t,x} \phi\|^2$$

On est dans une situation où l'on voudrait minimiser une fonction  $F$  de  $m$ , et l'on considère une fonction  $F_-$  de  $m$  qui est inférieure à  $F$ . Si  $\bar{m}$  est optimal pour  $F_-$ , il existe un point selle  $(\bar{x}, \bar{\xi}, \bar{\phi})$  et la quantité ajoutée au critère y est nulle. On a donc  $F_-(\bar{m}) = F(\bar{m})$ , donc le minimum de  $F$  est inférieur à  $F(\bar{m})$  qui est le minimum de  $F_-$ , qui est inférieure au minimum de  $F$  car elle lui est toujours inférieure. Ainsi, on n'a que des égalités, et on est assurés de ne pas avoir changé le problème en ajoutant cette quantité au critère à minimiser.

## 6.2.2 L'algorithme de Brenier

On approche l'optimum grâce à l'algorithme de BRENIER, de type gradient :

$$\begin{cases} F(m) = \sup_x f(m, x) \\ \nabla F(m) = \nabla_m f(m, x^*) \text{ avec } x^* \text{ réalisant le supremum} \end{cases}$$

Le problème est que cet algorithme ne s'applique que pour des fonctions suffisamment régulières pour pouvoir être différenciée presque partout. Dans le cas général, on effectue une discrétisation du problème et on utilise un procédé récursif pour l'approximation :

$$\forall k \in \mathbb{N}, \begin{cases} \phi_{k+1} = \phi \text{ maximisant le critère avec } (m_k, \xi_k) \\ \xi_{k+1} = \xi \text{ minimisant le critère avec } (m_k, \phi_{k+1}) \\ m_{k+1} = m_k - r(\xi_{k+1} - \nabla_{x,t} \phi_{k+1}) \end{cases}$$

On aurait préféré que  $\xi_k$  optimise à  $\phi_{k+1}$  fixé et que, inversement,  $\phi_{k+1}$  optimise à  $\xi_k$  fixé. Pour une fonction convexe, les conditions vérifiées par cette construction sont suffisantes.

Le problème est maintenant de savoir comment faire pour réussir à déterminer les deux optimums intervenant dans la récursion de manière efficace.

Pour le calcul de  $\phi_{k+1}$  ...

Pour la seconde partie, c'est plus aisé car  $\xi$  doit maximiser le critère pour  $(m_k, \phi_{k+1})$ , donc on doit déterminer

$$\min_K I_K(\xi) + \frac{1}{2} \|\xi - \nabla \phi_{k+1}\| - m_k \xi$$

Ce qui est également équivalent à minimiser  $\frac{r}{2} \|\xi - \nabla_{t,x} \phi_k - \frac{1}{r} m_k\|^2$  pour  $\xi$  dans le convexe  $K$ . Donc pour tout couple  $(tnx)$ , on prend l'élément  $\xi(t, x)$  qui minimise le critère. Notons que l'on peut adopter cette définition ponctuelle sans s'angoisser de la régularité de la fonction  $\xi$  au final : contrairement à  $F$ , celle-ci ne doit pas respecter de contrainte sur son gradient. Ainsi :

$$\xi_K : (t, x) \mapsto P_K(\nabla_{t,x} \cdot \phi_{k+1}(t, x) + \frac{1}{r} m_k(t, x))$$

## 6.3 Cours 7

### 6.4 Introduction aux flots gradients

#### 6.4.1 Formulation du problème

On se place dans un espace de dimension finie que l'on assimile à  $\mathbb{R}^n$ . On considère le problème

$$(E) \begin{cases} x'(t) = -\nabla F(x(t)) \\ x(t_0) = x_0 \end{cases}, \quad t \in \mathbb{R}, \quad F \text{ différentiable}$$

et on cherche à minimiser  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  parmi les solutions du problème (E). Autrement dit, il s'agit d'optimiser le mouvement de  $F$  à partir de la seule condition initiale, donc de suivre en permanence la direction donnée par la plus grande pente, *i.e.* la direction du gradient.

Si  $F \in C^{1,1}(\Omega)$ , le gradient de  $F$  est continu et le théorème de CAUCHY-LIPSCHITZ s'applique et donne l'existence et l'unicité d'une solution maximale.

Si  $F \in C^1(\Omega)$  et **convexe**, on est assurés de l'existence de la solution, mais on ne peut *a priori* rien dire quant à son unicité. Si  $x$  et  $y$  sont deux solutions, il vient :

$$\frac{d}{dt}|x - y|^2 = \langle x - y | x' - y' \rangle = \langle x - y | \nabla F(y) - \nabla F(x) \rangle$$

Cette quantité étant négative par convexité de  $F$ , la distance de  $x$  à  $y$  décroît avec le temps, or elle est nulle à l'origine car la condition initiale est fixée. Donc les deux fonctions sont égales, et on a dans ce cas également l'unicité de la solution.

Si  $F \in C^1(\Omega)$  et  **$\lambda$ -convexe**, les mêmes calculs donnent

$$\frac{d}{dt}|x - y|^2 = \langle x - y | x' - y' \rangle = \langle x - y | \nabla F(y) - \nabla F(x) \rangle \leq -\lambda|x - y|^2$$

Le lemme de GRONWALL donne alors  $g \leq g(0)e^{Ct}$  si  $g(t) = |x - y|^2$ . Puisque  $g(0) = 0$  et que la fonction est positive, elle est constamment nulle, et ces hypothèses suffisent donc également à avoir l'unicité.

#### 6.4.2 Discrétisation temporelle

On fixe un pas temporel  $\tau > 0$  et on cherche une suite  $(x_k)_k$  représentant les positions successives de la courbe de la solution optimale aux temps  $k\tau$ . Le pas  $\tau$  est appelé à tendre vers 0, ce qui fait que la distance entre les  $x_k$  tendra également vers 0 dès que  $F$  est suffisamment régulière. Donc pour la recherche d'une solution minimisante, on peut imposer que  $x_{k+1}$  minimise le critère

$$F(x) + \frac{1}{2\tau}|x - x_k|^2$$

ce qui ne nuit pas à la généralité *ad infinitum*.

Si  $F$  est  $\lambda$ -convexe, le minimum existe en prenant  $\tau \leq -\frac{1}{\lambda}$ . Dans tous les cas, le  $x_{k+1}$  optimisant le critère doit annuler le gradient du critère, autrement dit il doit satisfaire

$$\frac{1}{\tau}(x_{k+1} - x) + \nabla F(x_{k+1}) = 0 \text{ d'où } \frac{x_{k+1} - x_k}{\tau} = \frac{\Delta x}{\Delta t} = -\nabla F(x_{k+1})$$

On définit donc la suite  $(x_k)_k$  récursivement par

$$\begin{aligned} & x_0 \\ \forall k \in \mathbb{N}, & \quad x_{k+1} = x_k - \tau \nabla F(x_{k+1}) \end{aligned}$$

Il s'agit du schéma d'EULER implicite, qui correspond intuitivement bien à ce que l'on cherche : on choisit de suivre à intervalle de temps régulier la plus grande pente possible au point courant, donnée par l'opposé du gradient. On « retrouve » alors  $F$  en interpolant les  $x_k$  lorsque le pas tend vers zéro. En fait tout ceci se généralise dans le cas plus général des espaces métriques, et non plus nécessairement euclidien, en imposant à  $x_{k+1}$  de minimiser  $F(x) + \frac{1}{2\tau}d(x, x_k)^2$ .

## 6.5 Retour sur les flots gradient

On suppose ici que  $\Omega$  est compact et que  $F : P(\Omega) \rightarrow \mathbb{R} \cup \{+\infty\}$ , en utilisant la distance  $W_p$ . On se demande comment minimiser efficacement le critère  $F(x) + \frac{1}{2\tau}|x - x_k|^2$ . À  $\tau > 0$  fixé, notons  $\frac{\delta F}{\delta \mu}$  toute fonction satisfaisant

$$\frac{d}{d\varepsilon} F(\mu + \varepsilon\chi)_{\varepsilon=0} = \int \frac{dF}{d\mu} d\chi, \forall \chi = \mu_1 - \mu, \mu_1 \in \mathbb{P}(\Omega)$$

Par exemple, avec  $F(\mu) = \int V d\mu$ , on trouve  $\frac{\delta F}{\delta \mu} = \int V d\chi$ , ou encore si  $F(\mu) = \int U(\rho)$  si  $\mu = \rho \cdot \lambda$  et  $+\infty$  sinon, alors  $\frac{\delta F}{\delta \mu} = U'(\rho)$

Qui est alors  $\frac{\delta}{\delta \mu} (\frac{1}{2}W_2^2(\cdot, \nu))$  ?

On a  $\frac{\delta}{\delta \mu} (\frac{1}{2}W_2^2(\cdot, \nu)) = \sup (\int \phi d\mu + \int \psi d\nu)$  sur les fonctions vérifiant  $\phi \oplus \psi \leq \frac{1}{2}d^2$ . On note alors  $\phi_\varepsilon(x) = \inf \frac{1}{2}|x - y|^2 - \psi_\varepsilon(y)$  qui est  $\text{diam}(\Omega)$ -lipschitzienne. On écrit alors

$$\int \phi d\chi = \frac{1}{\varepsilon} \left( \int \phi d(\mu + \varepsilon\chi) + \int \psi d\nu - \int \phi d\mu - \int \psi d\nu \right) \leq \frac{1}{2\varepsilon} (W_2^2(\mu + \varepsilon\chi) + W_2^2(\mu, \nu)) \leq \int \phi_\varepsilon d\chi$$

(WHY ?)

Or par le théorème d'ASCOLI, la suite étant uniformément bornée, on peut extraire de cette suite  $(\phi_\varepsilon)_\varepsilon$  une sous-suite convergent uniformément vers une certaine fonction  $\tilde{\phi}$ . On fait de même avec  $\psi$  pour obtenir une fonction limite  $\tilde{\psi}$ . Le couple  $(\tilde{\phi}, \tilde{\psi})$  est nécessairement optimal par construction, donc  $\frac{d}{d\mu}(\dots) = \tilde{\phi}$ .

Ainsi, si  $\mu$  minimise  $F(\mu) + \frac{1}{2\tau}W_2^2(\mu, \mu_k)$ , alors, puisque  $\phi$  est un potentiel de KANTOROVICH :

$$\begin{aligned} \frac{d}{d\varepsilon} \left( F(\mu + \varepsilon\chi) + \frac{1}{2\tau}W_2^2(\mu + \varepsilon\chi, \mu_k)_{\varepsilon=0} \right) &= 0 \text{ donc } \forall \chi, \left( \int \frac{\delta F}{\delta \mu} + \frac{\phi}{\tau} \right) \chi = 0 \\ \text{donc } \forall \chi = \mu_1 - \mu, \int \left( \frac{\delta F}{\delta \mu} + \frac{\phi}{\tau} \right) d\mu_1 &= \int \left( \frac{\delta F}{\delta \mu} + \frac{\phi}{\tau} \right) d\mu \text{ est constante} \\ \text{donc } \frac{\delta F}{\delta \mu} + \frac{\phi}{\tau} &\text{ est constante} \end{aligned}$$

En prenant le gradient, il vient alors

$$\nabla \frac{\delta F}{\delta \mu} + \frac{\nabla \phi}{\tau} = 0 \text{ donc } T(x) = x - \nabla \phi(x) \text{ soit } \nabla \phi = id - T$$

Le théorème de BRENIER donne, pour  $h : x \mapsto \frac{1}{2}|z|^2$ ,  $T(x) = x - (\nabla h)^{-1}(\nabla \phi(x))$ . On introduit alors

$$v_k^\tau : x \mapsto \frac{1}{\tau}(x - T(x)) = -\nabla \frac{\delta F}{\delta \mu}(x) \text{ et } \mu_k = T\#\mu \text{ optimisant le déplacement de la particule}$$

## 6.6 Un cas précis : l'équation de la chaleur

On a vu que si  $F(\rho) = \int f(\rho)$ , on a  $\frac{\delta F}{\delta \mu} = f'(\rho)$ . L'équation de continuité devient alors

$$\partial_t \rho + \nabla \cdot (\rho \nabla f'(\rho)) = 0$$

On cherche à étudier l'équation de la chaleur

$$\partial_t \rho - \Delta \rho = 0$$

Pour s'y ramener par cette voie, on doit nécessairement avoir

I

La fonction

$$G : \mu \mapsto \begin{cases} \int \rho \ln \rho & \text{si } \mu = \rho \cdot \lambda \\ +\infty & \text{sinon} \end{cases}$$

est semi-continue inférieurement pour la topologie faible. En effet, si  $\mu_n \rightarrow \mu$  avec  $\mu_n = \rho_n \lambda$ , on a  $\mu \ll \lambda$  et  $\int \rho \ln \rho \leq \liminf \int \rho_n \ln \rho_n$  dès que  $U(\rho)$  (???) avec  $U$  convexe et superlinéaire, i.e. telle que la limite de  $U(t)/t$  en l'infini soit infinie.

$$F : \mu \mapsto \begin{cases} \int \rho \ln \rho + \int V \rho & \text{si } \mu = \rho \cdot \lambda \\ +\infty & \text{sinon} \end{cases}$$

Pour  $\tau > 0$ , on cherche à approcher une solution par la méthode précédemment exposée. On introduit donc  $\rho_{k+1}$  une mesure optimisant  $\int \rho \ln \rho + \int V \rho + \frac{1}{2\tau} W_2^2(\rho, \rho_k)$  avec  $\rho_0$  donnée.

On se place dans le cas où  $\Omega$  est compact, dans lequel on est assurés de l'existence d'un minimiseur. De plus notons que

- \*  $\rho \mapsto W_2^2(\rho, \nu) = \sum \{ \int \phi d\mu + \int \psi d\nu \mid \phi \oplus \psi \leq W_2 \}$  est convexe à  $\nu$  fixée ;
- \*  $\rho \mapsto \int V d\rho$  est linéaire donc convexe ;
- \*  $\rho \mapsto \int \rho \ln \rho$  est strictement convexe.

La fonction  $F$  étant la somme de ces trois fonctions, elle est strictement convexe, et le minimiseur est donc unique.

Pour une mesure  $\rho$  minimale, on a nécessairement une dérivée par rapport à  $\varepsilon$  qui s'annule en  $\varepsilon = 0$ . On regarde donc le comportement de chaque terme de la dérivée :

### DERIVEES

On regarde les taux d'accroissement :

$$\begin{aligned} \frac{W_2^2(\rho_\varepsilon, \nu) - W_2^2(\rho, \nu)}{\varepsilon} &\geq \frac{\int \phi d\rho_\varepsilon - \int \phi d\rho}{\varepsilon} = \int \phi d(\rho_1 - \rho) \\ &\leq \frac{\int \phi_\varepsilon + \int \psi_\varepsilon d\nu - \int \phi_\varepsilon d\rho - \int \psi_\varepsilon d\nu}{\varepsilon} = \int \rho_\varepsilon d(\rho_1 - \rho) \end{aligned}$$

On associe à  $(\phi, \psi)$  les  $\varepsilon$ -transformées :

$$\begin{aligned} \phi_\varepsilon(x) &= \inf_y \left( \frac{1}{2} |x - y|^2 - \psi_\varepsilon(y) \right) \\ \psi_\varepsilon(x) &= \inf_x \left( \frac{1}{2} |x - y|^2 - \phi_\varepsilon(x) \right) \end{aligned}$$

On fixe alors une suite  $\varepsilon_k$  qui converge vers 0 et telle que

$$\lim_{k \rightarrow \infty} \frac{W_2^2(\rho_{\varepsilon_k}, \nu) - W_2^2(\rho, \nu)}{\varepsilon_k} = \limsup_{k \rightarrow \infty} \frac{W_2^2(\rho_{\varepsilon_k}, \nu) - W_2^2(\rho, \nu)}{\varepsilon_k}$$

On veut prouver que  $\liminf \rho \geq \int \phi d(\rho_1 - \rho) \geq \limsup(\rho)$ . Quitte à extraire, on peut supposer que

$$\varepsilon_{k_i} \rightarrow 0, \phi_{\varepsilon_{k_i}} \xrightarrow{\text{CVU}} \tilde{\phi}, \psi_{\varepsilon_{k_i}} \xrightarrow{\text{CVU}} \tilde{\psi}$$

On va prouver que  $(\tilde{\phi}, \tilde{\psi})$  est également optimal pour  $\rho$  :

$$\begin{aligned} \frac{1}{2} W_2^2(\rho_{\varepsilon}, \nu) &= \int \phi_{\varepsilon} d\rho + \int \psi_{\varepsilon} d\nu + \varepsilon \int \phi_{\varepsilon} d(\rho_1 - \rho) \\ \text{to} \quad \text{to to} & \\ \frac{1}{2} W_2^2(\rho, \nu) &= \int \tilde{\phi} d\rho + \int \tilde{\psi} d\nu \end{aligned}$$

donc on obtient de nouveaux potentiels de KANTOROVICH  $(\tilde{\phi}, \tilde{\psi}, \rho)$ .

2

On aimerait désormais avoir un peu plus d'informations sur ce minimiseur. Posons  $\chi = \rho_1 - \rho$  où  $\rho_1$  est une mesure de probabilité sur  $\Omega$  absolument continue par rapport à la mesure de LEBESGUE, et posons  $\rho_{\varepsilon} = \rho + \varepsilon(\rho_1 - \rho) = (1 - \varepsilon)\rho + \varepsilon\rho_1$ . On voudrait prouver que  $\rho > 0$  presque partout. La mesure  $\chi$  est une mesure absolument continue et de masse nulle *générique*. Il suffit d'utiliser  $\rho_1 = \chi - \rho$ .

$$\frac{d}{d\varepsilon} \int (U(\rho_{\varepsilon}) \rho_{\varepsilon} \ln \rho_{\varepsilon})_{\varepsilon=0} = \int (1 + \ln \rho)(\rho_1 - \rho) = U'(\rho_k)(\rho_1 - \rho) = -\infty \cdot c \text{ là où } \rho = 0$$

De plus, on a

$$W_2^2(\rho_{\varepsilon}, \rho_k) \leq \dots \leq W_2^2(\rho_1, \rho_k) - W_2^2(\rho, \rho_k) \text{ et } \frac{d}{d\varepsilon} \int V d\rho_{\varepsilon} = \int V d(\rho_1 - \rho)$$

Ainsi, toutes les dérivées existent et sont finies, sauf si le logarithme tend vers l'infini *i.e.* lorsque  $\rho$  s'annule, auquel cas le gradient serait négatif et on n'aurait sûrement pas un minimum. Ainsi on a  $\rho > 0$  presque partout, comme voulu.

On a  $T(x) = x - \nabla \phi(x)$  et  $T$  qui est unique presque partout, donc  $\nabla \phi$  est unique  $\rho_1$ -presque partout, donc  $\nabla \phi$  est unique presque partout, donc, puisque  $\Omega$  est convexe,  $\phi$  est unique presque partout à une constante près. On fixe la constante en imposant  $\phi(x_0) = 0$ . Par optimalité, il vient alors

$$\forall \rho_1 \in \mathbb{P}(\Omega), \int (g = 1 + \ln \rho + V + \frac{\phi}{\tau}) d(\rho_1 - \rho) \geq 0$$

et  $g = \frac{d}{d\varepsilon} F(\rho_{\varepsilon}) + \frac{1}{2\tau} W_2^2(\rho_{\varepsilon}, \rho_t)$ . Donc pour tout  $\rho_1$ , on a

$$l + \varepsilon \longleftarrow \int g d\rho_1 \geq \int g d\rho \geq l$$

En choisissant  $\rho_1$  concentrée sur  $\{g < l + \varepsilon\}$ , on en tire que  $\inf g = l$ . Il vient alors que  $g = l$   $\rho$ -presque partout. Ainsi,  $\ln \rho + V + \frac{\phi}{\tau}$  est constante presque partout,  $\rho = \exp(c - V - \frac{\phi}{\tau})$  est donc lipschitzienne et donc admet un gradient presque partout par le théorème de RADEMACHER. Il vient alors  $\nabla \rho + \nabla V + \frac{\nabla \phi}{\tau} = 0$  presque partout.

De même, on a  $\rho_{k+1}^{\tau}$  car  $v_{k+1}^{\tau} = \frac{1}{\tau}(id - T)$  où  $T \# \rho_{k+1} = \rho_k$  qui est optimal, or  $v_{k+1}^{\tau} = \frac{\nabla \phi}{\tau} = -\nabla V - \frac{\nabla \rho_{k+1}}{\tau}$  presque partout.

### 3 – Interpolation

On obtient par la formule qui précède une suite  $(\rho_k)_k = (\rho_k^\tau)_k$  interpolant une mesure optimal  $\rho$ . On veut maintenant interpoler les  $\rho_k^\tau$  pour retrouver la mesure  $\rho$ .

Notons  $\rho^\tau$  la mesure constantes par morceaux définie par  $\rho_0$  entre 0 et  $\tau$ , puis  $\rho_1$  entre  $\tau$  et  $2\tau$ , et ainsi de suite. On réalise alors une interpolation continue par les géodésiques : sur  $](k-1)\tau, k\tau[$ , on prend la géodésique à vitesse constante qui relie  $\rho_{k-1}^\tau$  à  $\rho_k^\tau$ , et l'on définit par ce moyen une interpolée  $\tilde{\rho}_k^\tau$ . L'intérêt est que  $\rho^\tau$  n'est pas continue, et le manque de résultat dans ce cas empêche de continuer, alors que  $\tilde{\rho}^\tau$  est continue, donc on est assurés de l'existence d'un champ de vitesse  $\tilde{v}^\tau$  tel que l'équation de continuité soit vérifiée pour la mesure interpolée :

$$\partial_t \tilde{\rho}^\tau + \nabla \cdot (\tilde{v}^\tau \tilde{\rho}^\tau) = 0$$

et ce champ de vitesses est connu dès que  $\mu_0$  et  $\mu_1$  sont fixées et que  $T$  est un transport optimal :  $v_t = (T - id) \circ T_t^{-1}$ .

Ici, on a  $|\tilde{\rho}^\tau| = |\frac{1}{\tau} W_2(\rho_k, \rho_{k+1})|$ , d'où

$$W_2(\tilde{\rho}_t^\tau, \tilde{\rho}_s^\tau) \leq \text{long} \tilde{\rho}_{|[t,s]}^\tau = \int_t^s |\tilde{\rho}^\tau| \leq \left( \int_t^s |\tilde{\rho}^\tau|^2 \right)^{1/2} |t-s|^{1/2}$$

Donc, puisque  $\tilde{\rho}$  est hölderienne :

$$\int_0^T |\tilde{\rho}'|^2 = \sum \tau \frac{W_2^2(\rho_k, \rho_{k+1})}{\tau^2} \leq C$$

En effet, on a :

$$F(\rho_{k+1}) + \frac{1}{2\tau} W_2^2(\rho_{k+1}, \rho_k) \leq F(\rho_k) + 0$$

et on peut sommer de manière télescopique :

$$\frac{W_2^2(\rho_k, \rho_{k+1})}{\tau} \leq F(\rho_k) - F(\rho_{k+1}) \text{ donc } \sum \frac{W_2^2(\rho_k, \rho_{k+1})}{\tau} \leq C$$

Notons  $v_k^\tau$  la vitesse de  $\rho_k^\tau$  à  $\rho_{k+1}^\tau$ . Soit également  $E^\tau = v^\tau \rho^\tau$  et  $\tilde{E}^\tau = \tilde{v}^\tau \tilde{\rho}^\tau$ . Pour  $E^\tau$  sur  $](k-1)\tau, k\tau[$  on prend  $v_k^\tau$  et  $\rho_k^\tau = \frac{1}{\tau}(x - T(x))$ ; et pour  $\tilde{E}^\tau$ , on prend pour  $\tilde{v}^\tau$  le champ de vecteur valant  $v_k^\tau \circ T_t^{-1}$  sur  $](k-1)\tau, k\tau[$ .

La mesure  $\tilde{\rho}^\tau$  vérifie les hypothèses du théorème d'ASCOLI donc, quitte à extraire, on peut supposer que la suite converge uniformément vers une certaine  $\rho$ .

On a par hölderianité  $W_2(\rho^\tau, \tilde{\rho}^\tau) \leq c\tau^{1/2}$  qui tend vers 0 avec  $\tau$ , donc on a  $\rho^\tau$  qui converge également vers  $\rho$ .

$$\partial_t \rho + \nabla \cdot \tilde{E}^\tau = 0, \quad E = -\nabla \rho + \rho \nabla V$$

$$\partial_t \tilde{\rho}^\tau + \nabla \cdot \tilde{E}^\tau = 0, \quad E^\tau = v^\tau \rho^\tau = \left( -\frac{\nabla \rho^\tau}{\rho^\tau} + \nabla V \right) \rho^\tau = -\nabla \rho^\tau + \rho^\tau \nabla V$$

Supposons que  $\tilde{E}^\tau$  converge faiblement vers  $\tilde{E}$  et que  $E^\tau$  converge faiblement vers  $E$ . Si on peut prouver que  $E = \tilde{E}$ , alors on aura

$$\partial_t \rho - \nabla \cdot (\nabla \rho + \rho \nabla V) = 0 \text{ donc } \partial_t \rho - \nabla \rho - \nabla \cdot (\rho \nabla V) = 0$$

Pour ce faire, il suffit de prouver l'une des convergences. Notons que pour  $v = 0$ , on retrouve l'équation de FORMER – PLANCK, qui a le mérite d'être linéaire et qui permet donc de traiter le problème dans le cadre des distributions.

Si  $\int \int |\tilde{E}^\tau|$  est bornée, alors il en va de même pour  $\tilde{E}^\tau$  dans l'espace des mesures vectorielles  $\mathcal{M}(\Omega \times [0, T], \mathbb{R}^d)$ . Or, on a

$$\int \int |\tilde{E}^\tau| = \int \int \tilde{v} d\tilde{\rho}^\tau = \int \|\tilde{v}^\tau\|_{L^1(\tilde{\rho}^\tau)} \leq \int \|\tilde{v}^\tau\|_{L^2(\tilde{\rho}^\tau)} = \int \|\tilde{\rho}^\tau\|_{W_2} \leq C \left( \int \|\tilde{\rho}^\tau\|_{W_2}^2 \right)^{1/2} = \sum \frac{W_2^2(\rho_k, \rho_{k+1})}{\tau} \leq C$$

(???)

Le caractère borné permet donc, quitte à extraire, de supposer que  $\tilde{E}^\tau$  converge faiblement vers  $\tilde{E}$ . Notons que garder  $\tilde{E}$  et non  $E$  est nécessaire pour que l'équation de continuité soit toujours satisfaite. Cependant, travailler avec  $E$  suffit pour avoir la relation entre  $E$  et  $\frac{E}{\tau}$ .

Nous venons d'obtenir un résultat de convergence sur notre problème « interpolé »  $\tilde{E}$ , et nous aimerions maintenant en tirer des informations sur le champ initial  $E$ . En effet nous savons que l'équation de continuité est vérifiée pour  $(\tilde{\rho}^\tau, \tilde{E}^\tau)$  :

$$\partial_t \tilde{\rho}^\tau + \nabla \cdot \tilde{E}^\tau = 0$$

La convergence de  $\tilde{\rho}^\tau$  vers  $\tilde{\rho}$  et celle de  $\tilde{E}^\tau$  vers  $\tilde{E}$  entrainerait-elle la même relation quant aux limites? Autrement dit, a-t-on comme voulu :

$$\partial_t \rho^\tau + \nabla \cdot E^\tau = 0$$

Pour toute fonction  $f$  lipschitzienne, on cherche à estimer

$$\int \int f d\tilde{E}^\tau - \int \int f dE^\tau$$

Avec  $\tilde{\rho}^\tau = T_t \# \rho_k$  et  $\tilde{v}^\tau = v^\tau \circ T_t^{-1}$ , on a :

$$\begin{aligned} \int \int f d\tilde{E}^\tau &= \int \int f \tilde{v}^\tau d\tilde{\rho}^\tau = \int f(v^\tau \circ T_t^{-1}) d(T_t \# \rho^\tau) \\ &= \int (\rho \circ T_t) v^\tau d\rho^\tau = \int f(T_t) v^\tau d\rho^\tau \\ &= \int f(T_t) dE^\tau \end{aligned}$$

Or  $f$  est lipschitzienne et  $|x - T_t(x)| \leq \tau |v^\tau(x)|$ , ce qui permet d'écrire

$$\begin{aligned} \int \int f d\tilde{E}^\tau - \int \int f dE^\tau &= \int \int f(T_t) dE^\tau - \int \int f dE^\tau \\ &\leq Lip(f) \int \int \tau |v^\tau(x)| |v^\tau| d\rho^\tau \\ &= Lip(f) \tau \int \|v^\tau\|_{L^2(\rho^\tau)}^2 \end{aligned}$$

Or,  $v^\tau = \frac{1}{\tau}(id - T)$  donc  $\int |v^\tau|^2 d\rho^\tau = \int \frac{|x - T(x)|^2}{\tau^2} d\rho^\tau = \frac{W_2^2}{\tau^2}$ , d'où

$$Lip(f) \tau \int \|v^\tau\|_{L^2(\rho^\tau)}^2 = Lip(f) \tau \sum \frac{W_2^2(\rho_k, \rho_{k+1})}{\tau} \leq C \tau$$

Ainsi, à l'image de la convergence  $\tilde{E}^\tau \rightharpoonup \tilde{E}$  obtenue précédemment, on a  $E^\tau \rightharpoonup E$ . L'équation de continuité pour  $(\tilde{\rho}^\tau, \tilde{E}^\tau)$  se traduit au sens des distributions par

$$\forall \phi, \int_0^T \int_\Omega \partial_t \phi d\tilde{\rho}^\tau + \int_0^T \int_\Omega \nabla \phi \cdot \tilde{E}^\tau = 0$$

Pour la première intégrale double, il y a convergence de l'intégrale spatiale par convergence faible de  $\tilde{\rho}^\tau$  vers  $\rho$ , puis convergence de l'intégrale temporelle par convergence dominée. Concernant la seconde intégrale, il y a beaucoup moins de complications car nous avons prouvé la convergence de  $\tilde{E}^\tau$  vers  $E$  sur l'espace-temps  $\Omega \times [0, T]$ .

Ainsi on obtient, puisque  $\rho^\tau \rightharpoonup \tilde{\rho} = \rho$  :

$$\forall \phi, \int \phi \cdot E^\tau = \int (\rho^\tau \nabla \cdot \phi - \rho^\tau \nabla V \cdot \phi)$$

Ainsi, on retrouve la même équation que pour  $(E, V)$  (?).

Finalement, on s'est ramené à considérer un chemin discret approchant les courbes solutions, à chaque étape on a un minimiseur qui est unique comme potentiel de KANTOROVICH, puis (...). On trouve alors une relation entre le déplacement et le transport  $T(x)$ , et on interpole la suite de positions discrètes ainsi construire pour approcher une solution du problème  $(E)$  initial. Résoudre  $(\tilde{E}_c)$  donne une équation entre  $E$  et  $\rho$

$$\frac{\delta J_W^2}{\delta \mu} = \int W(x-y)\rho(y)dy \text{ donc } \tilde{E}^\tau \rightharpoonup \tilde{E}$$

# Chapitre 7

## Annexes

### 7.1 Théorie de la mesure

#### 7.1.1 Convergence faible

On fait un usage exhaustif de la topologie *faible\**, très pratique puisque le théorème de BANACH-ALAOGLU-BOURBAKI garantit que les boules fermées sont compactes pour cette topologie. Ceci nous permettra d'extraire des suites convergentes de mesures en identifiant les espaces de mesures au dual de certains espaces de BANACH.

Lorsque  $\Omega$  est compact, l'ensemble des mesures signées de variation totale finie qu'on note  $\mathcal{M}(\Omega)$  (dont la norme est la variation totale) s'identifie au dual des fonctions continues  $\mathcal{C}(\Omega)'$  via l'isométrie  $\mu \mapsto (f \mapsto \int f d\mu)$ . Clairement, les mesures positives forment un fermé pour cette topologie, et  $Prob(\Omega)$  (mesures positives de masse unitaire) est un compact en tant que partie fermée de la boule unité positive (si  $Prob(\Omega) \ni \mu_n \xrightarrow{*} \mu$  alors  $1 = \int \mathbf{1} d\mu_n \rightarrow \int \mathbf{1} d\mu = \mu(\Omega)$  donc la limite est encore de masse 1).

Si  $\Omega$  est quelconque, l'ensemble des mesures signées de variation totale finie s'identifie cette fois au dual des fonctions continues nulles à l'infini  $\mathcal{C}_0(\Omega)$ . Or  $Prob(\Omega)$  n'est pas nécessairement compact ( $\mathbf{1} \notin \mathcal{C}_0(\Omega)$  donc l'argument précédent ne fonctionne pas, de la masse peut être envoyée à l'infini). Pour pouvoir extraire d'une suite  $\mu_n \in Prob(\Omega)$  une sous-suite convergente, il faut en toute généralité des hypothèses supplémentaires. C'est l'objet d'une autre section sur les suites tendues.

#### 7.1.2 Lemme d'Ulam

Le lemme d'ULAM affirme qu'une mesure de probabilité sur un espace polonais (*i.e.* métrique complet et séparable) est toujours concentrée sur un  $\sigma$ -compact. Autrement dit, il existe une réunion dénombrable de compacts  $K_n$  qui concentre toute la masse. Cela revient également à dire que pour toute mesure de probabilité  $\mu$ , on a

$$\forall \varepsilon > 0, \exists K_\varepsilon \text{ compact, } \mu(K_\varepsilon^c) \leq \varepsilon$$

#### 7.1.3 Suites tendues

Une suite  $(\mu_n)_n$  est dite *tendue* si

$$\forall \varepsilon > 0, \exists K \text{ compact, } \forall n \in \mathbb{N}, \forall A, (A \cap K = \emptyset \implies |\mu_n(A)| \leq \varepsilon)$$

Autrement dit, ce sont des suites pour lesquelles des compacts concentrent la majeure partie de la masse de chacune des mesures. De telles suites jouissent d'une propriété très importante : le théorème de PROKHOROV affirme que si une suite de mesures est tendue, alors on peut en extraire une sous-suite faiblement convergente au sens des fonctions bornées, ce qui revient à dire que :

$$\forall \phi \in \mathcal{C}_B, \int \phi d\mu_{n_i} \xrightarrow{n \rightarrow \infty} \int \phi d\mu$$

## 7.2 Analyse fonctionnelle

### 7.2.1 Semi-continuité inférieure

Une fonction  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  est *semi-continue inférieurement* si pour toute suite  $(x_n)_n$  convergeant vers  $x$ ,  $\liminf f(x_n) \geq f(x)$ . Graphiquement et intuitivement, cela revient à dire que les éventuels points de discontinuités de  $f$  ne correspondent qu'à des sauts *vers le bas*.

Tout l'intérêt de cette notion réside dans le fait que l'existence d'un minimum dans le cas d'une fonctionnelle continue sur un compact tient également sous la seule hypothèse d'une fonctionnelle semi-continue inférieurement, car en effet on a alors  $\liminf f(x_n) \geq f(x) \geq \liminf f(x_n)$ , la première inégalité venant de la définition de la semi-continuité inférieure, et la seconde venant de la définition de la limite inférieure. Les valeurs d'adhérences d'une suite extraite minimisante réalisent, dans ce cas également, le minimum de la fonctionnelle.

Une caractérisation utile des fonctions semi-continues inférieurement et bornées inférieurement affirme que ce sont exactement les bornes supérieures de familles de fonctions continues, autrement dit les limites croissantes de suites de fonctions continues. On peut même imposer l'uniforme continuité des fonctions approchantes. Pour le voir, il suffit de considérer la suite de fonctions :

$$\forall n \in \mathbb{N}, f_n : x \mapsto \inf_{y \in X} (f(y) + nd(x, y))$$

### 7.2.2 Théorème d'Ascoli

Si  $(f_n)_n$  est une suite de fonctions continues sur un compact  $X$ , uniformément bornées sur  $X$ , et équicontinues, *i.e.* vérifiant

$$\forall \varepsilon > 0, \exists \delta > 0, \forall n \in \mathbb{N}, \forall x, y \in \Omega, (|x - y| \leq \delta \implies |f_n(x) - f_n(y)| \leq \varepsilon)$$

alors on peut en extraire une sous-suite convergeant uniformément.

### 7.2.3 Transformation de Legendre

La transformée de Legendre d'une fonction convexe  $h$  est définie par :

$$h^* = \sup_{z \in \Omega} (xz - h(z))$$

Une fonction  $h$  convexe est strictement convexe si et seulement si  $h^*$  est de classe  $\mathcal{C}^1$ . De plus, l'inverse de la sous-différentielle est la différentielle de  $h^*$ , *i.e.*  $p$  est dans la sous-différentielle en  $z$  si, et seulement si,  $z$  est le gradient de  $h$  en  $p$ . En effet,  $h$  est différentiable plus que presque partout : l'ensemble des points pour lesquels le cardinal de la sous-différentielle est supérieur à 1 (*i.e.* où  $h$  n'est pas différentiable) est une réunion dénombrable d'hyperplan. L'ensemble des points où le cardinal de la sous-différentielle est supérieur à  $i$  est de dimension au plus  $n - i$ .

Une fonction  $\chi$  convexe est toujours la transformée de LEGENDRE de sa transformée de LEGENDRE. Tout d'abord il est clair que  $\chi^{**} \leq \chi$  en explicitant  $\chi^{**}$  par définition. Comme toute fonction convexe, on peut écrire  $\chi$  comme supremum de ses minorantes affines, ce qui s'écrit

$$\exists a, \forall x \in \Omega, \chi(x) = \sup\{\langle x|y \rangle - a(y)\} = a^*(x)$$

La première inégalité donne alors  $\chi^* = a^{**} \leq a$ , donc  $\chi^{**} \geq a^* = \chi$  car la transformation de LEGENDRE est décroissante (compte tenu de la présence du signe moins dans la définition). On obtient ainsi le résultat escompté.

### 7.2.4 Fonctions convexes

On sait qu'une fonction convexe définie sur un *ouvert* est supremum en tout point de ses minorantes affines. On peut donc définir une fonction convexe par cette caractérisation.

**Définition** On dit que  $f$  est *convexe* s'il existe une fonction  $a$  à valeur dans  $\overline{\mathbb{R}}$  telle que pour tout  $x$ ,  $f(x) = \sup_y (x \cdot y - a(y)) = a^*(x)$ .

En effet s'il n'existe pas de minorante affine dirigée par  $y$ , on prend  $a(y) = +\infty$ , sinon on prend pour  $a(y)$  le plus grand  $a$  tel que  $x \cdot y - a$  soit minorant. On retrouve ainsi la propriété classique des fonctions convexes : elles sont au dessus de leurs tangentes en chaque point, donc elles sont supérieures au supremum de ces tangentes, et ces supremum sont atteint en chaque point car les tangentes rencontrent le graphe.

Avec cette définition,  $f$  est convexe si et seulement si il existe  $g$  telle que  $f = g^*$ .

On va maintenant énoncer quelques théorèmes et propriétés des fonctions convexes, dont les démonstrations dépendent uniquement de la caractérisation prise en définition. L'intérêt est que l'on va pouvoir définir une notion plus générale que la convexité de manière analogue et nous disposerons alors des mêmes théorèmes. Les fonctions convexes s'appuient sur leurs tangentes minorantes ; qu'obtient-on alors en considérant des courbes minorantes plus générales que des droites ?

**Théorème** Pour une fonction  $f$ ,  $f^{**}$  est la plus grande fonction convexe inférieure ou égale à  $f$ .

**Corollaire**  $f$  est convexe si et seulement si  $f = f^{**}$ .

**Définition**  $\partial h(z)$  est l'ensemble des vecteurs  $p$  de  $\mathbb{R}^n$  tels que  $h(w) \geq h(z) + \langle p, w - z \rangle$  pour tout vecteur  $w$ , il est appelé sous-différentielle de  $h$  en  $z$ . La non différentiabilité de  $h$  en un point  $z$  correspond à l'existence de trop d'éléments dans sa sous-différentielle. Dans le cas où  $h$  est strictement convexe, les sous-différentielles sont soit égales, soit disjointes.

**Propriété**

$$y \in \partial f(x) \Leftrightarrow f(x) + f^*(y) = x \cdot y \Leftrightarrow x \in \partial f^*(y)$$

**Démonstration**

$$\begin{aligned} y \in \partial f(x) &\Leftrightarrow \forall z, f(z) \geq f(x) + \langle y, z - x \rangle \\ &\Leftrightarrow \forall z, -f(z) + \langle y, z \rangle \leq -f(x) + \langle y, x \rangle \\ &\Leftrightarrow \underset{\text{déf}}{f^*(y)} = -f(x) + \langle y, x \rangle \\ &\Leftrightarrow f(x) + f^*(y) = \langle y, x \rangle \\ &\Leftrightarrow x \in \partial f^*(y) \end{aligned}$$

**Remarque** Si  $f$  et  $f^*$  sont différentiables en  $y$ ,  $\nabla f^*(y) = \nabla f^{-1}(y)$ .

**Définition** On dit que  $\Gamma \subset \Omega \times \Omega$  est *cycliquement monotone* si  $\forall K \in \mathbb{N}^*$ ,  $\forall (x_i, y_i)_{i=1..K} \subset \Gamma$ ,  $\forall \sigma \in \mathfrak{S}_K$  on a :

$$\sum_{i=1}^K \langle x_i, y_i \rangle \geq \sum_{i=1}^K \langle x_i, y_{\sigma(i)} \rangle$$

### Remarque

- $k = 1$

$\Gamma$  est le "graphe" de quelque chose de croissant.

- $k = 2$

$$\begin{aligned}\langle x_1, y_1 \rangle + \langle x_2, y_2 \rangle &\geq \langle x_1, y_2 \rangle + \langle x_2, y_1 \rangle \\ \iff \langle x_2 - x_1, y_2 - y_1 \rangle &\geq 0\end{aligned}$$

Intuitivement, cela signifie que  $y$  croît avec  $x$ .

**Définition** Pour  $f$  convexe, on note  $G_f$  le graphe de sa sous-différentielle :

$$\begin{aligned}G_f &= \{(x, y) | y \in \partial f(x)\} \\ &= \{(x, y) | x \in \partial f^*(y)\} \\ &= \{(x, y) | f(x) + f^*(y) = x \cdot y\}\end{aligned}$$

**Propriété** Si  $f$  est convexe,  $G_f$  est cycliquement monotone.

### Démonstration

Soit  $x_1, \dots, x_K \in \mathbb{R}^K$  et  $y_1, \dots, y_K$  tels que  $\forall i, y_i \in \partial f(x_i)$ , c'est-à-dire  $f(x_i) + f^*(y_i) = \langle x_i, y_i \rangle$ .

$$\begin{aligned}\sum_{i=1}^K \langle x_i, y_i \rangle &= \sum_{i=1}^K f(x_i) + f^*(y_i) \\ &= \sum_{i=1}^K f(x_i) + \sum_{i=1}^K f^*(y_{\sigma(i)})\end{aligned}$$

Or par définition de  $f^*$ ,  $f^*(y_{\sigma(i)}) \geq \langle x_i, y_{\sigma(i)} \rangle - f(x_i)$ , de quoi on déduit :

$$\sum_{i=1}^K \langle x_i, y_i \rangle \geq \sum_{i=1}^K f(x_i) + f^*(y_{\sigma(i)})$$

**Théorème** Si  $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$  est cycliquement monotone, alors il existe une fonction convexe  $f$  telle que le graphe de sa sous-différentielle contienne  $\Gamma$ . La démonstration sera vue dans un cas plus général dans la section suivante.

### 7.2.5 Fonctions $c$ -concaves

Dans toute cette section,  $c$  est une fonction **symétrique** de  $\Omega \times \Omega$  dans  $\overline{\mathbb{R}}$ . On pourrait étendre les résultats qui suivent à des fonctions  $c$  quelconques en introduisant des  $c$ -transformées à gauche et à droite. On définit donc des ensembles plus généraux de fonctions concaves qui auront des propriétés analogues à ce qu'on vient de voir.

**Définition** Une fonction  $\phi$  est  $c$ -concave s'il existe une fonction  $\psi$  telle que pour tout  $x$ ,  $\phi(x) = \inf_y (c(x, y) - \psi(y))$ .

**Définition** Pour  $\phi$  quelconque, on définit sa  $c$ -transformée  $\phi^c$ .

$$\phi^c = \inf_y (c(x, y) - \phi(y))$$

**Propriété**  $\phi^{cc}$  est la plus petite fonction  $c$ -concave supérieure ou égale à  $\phi$ .

**Démonstration** On montre que (i)  $\phi^{cc} \geq \phi$  et (ii)  $\phi^{cc}$  est la plus petite.

(i) Par définition  $\phi^{cc} = \inf_y (c(x, y) - \phi^c(y))$ . Or

$$\forall y, c(x, y) - \phi^c(y) = c(y, x) - \phi^c(y) \underset{\text{(d\u00e9f)}}{\geq} \phi(x)$$

Ce qui donne  $\phi^{cc}(x) \geq \phi(x)$ .

(ii) On remarque que si  $a$  et  $b$  sont deux fonctions telles que  $a \leq b$ , alors  $a^c \geq b^c$ .  
Soit alors  $\psi$   $c$ -concave telle que  $\psi \geq \phi$ . Par d\u00e9finition  $\exists \chi, \psi = \chi^c$ .

$$\chi^c \geq \phi \implies \phi^c \geq \chi^{cc} \geq \chi \implies \psi = \chi^c \geq \phi^{cc}$$

**Corollaire**  $\phi^{cc} = \phi \iff \phi$  convexe

### Exemples

**Cas**  $f(x) = \frac{|x|^p}{p}$

$f^*(x) = \sup_y \left( x \cdot y - \frac{|x|^p}{p} \right)$ , le supremum \u00e9tant atteint pour un certain  $y = \lambda x$ ,  $\lambda \geq 0$ .

On a alors  $x \cdot y - \frac{|x|^p}{p} = \lambda |x|^2 - \lambda^p \frac{|x|^p}{p}$  avec  $|x|^2 = \lambda^{p-1} |x|^p$  par maximalit\u00e9. Ce qui donne :

$$f^*(x) = |x|^{\frac{p}{p-1}} - \frac{|x|^{\frac{p}{p-1}}}{p} = \frac{|x|^q}{q}$$

**Cas**  $c(x, y) = \frac{|x-y|^2}{2}$

Que faire ?

**D\u00e9finition** On dit que  $\Gamma \subset \Omega \times \Omega$  est  $c$ -cycliquement monotone si  $\forall K \in \mathbb{N}^*$ ,  $\forall (x_i, y_i)_{i=1..K} \subset \Gamma$ ,  $\forall \sigma \in \mathfrak{S}_K$  on a :

$$\sum_{i=1}^K c(x_i, y_i) \leq \sum_{i=1}^K c(x_i, y_{\sigma(i)})$$

**D\u00e9finition** Pour  $f$   $c$ -concave, on appelle graphe de sa  $c$ -sur-diff\u00e9rentielle l'ensemble  $G_f^c$  d\u00e9fini comme suit :

$$G_f = \{(x, y) | f(x) + f^c(y) = c(x, y)\}$$

**Propri\u00e9t\u00e9** Si  $f$  est  $c$ -concave,  $G_f^c$  est  $c$ -cycliquement monotone.

### D\u00e9monstration

Soit  $x_1, \dots, x_K \in \mathbb{R}^K$  et  $y_1, \dots, y_K$  tels que  $\forall i, y_i \in \partial f(x_i)$ , c'est-\u00e0-dire  $f(x_i) + f^c(y_i) = c(x_i, y_i)$ .

$$\begin{aligned} \sum_{i=1}^K c(x_i, y_i) &= \sum_{i=1}^K f(x_i) + f^c(y_i) \\ &= \sum_{i=1}^K f(x_i) + f^c(y_{\sigma(i)}) \\ &\underset{\text{d\u00e9f}}{=} f^c(y) = -f(x) + c(y, x) \end{aligned}$$

Or par d\u00e9finition de  $f^c$ ,  $f^c(y_{\sigma(i)}) \leq c(x_i, y_{\sigma(i)}) - f(x_i)$ , de quoi on d\u00e9duit :

$$\sum_{i=1}^K c(x_i, y_i) \leq \sum_{i=1}^K f(x_i) + f^c(y_{\sigma(i)})$$

**Théorème** Si  $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$  est  $c$ -cycliquement monotone, alors il existe une fonction  $c$ -concave  $f \neq -\infty$  telle que le graphe de sa  $c$ -sur-différentielle contienne  $\Gamma$ .

**Démonstration** On exhibe une telle fonction  $f$  en considérant le gain sur le coût total obtenu par une permutation circulaire des points d'arrivées dans le plan de transfert, en partant d'un point variable. Soit  $(x_0, y_0) \in \Gamma$  fixé.

$$f(x) = \inf_{(x_i, y_i)_{i=1 \dots K} | K \in \mathbb{N}^*} c(x, y_K) - c(x_K, y_K) + c(x_K, y_{K-1}) + \dots + c(x_1, y_0) - c(x_0, y_0)$$

Montrons que (i)  $f$  est  $c$ -concave, (ii)  $f \neq -\infty$  et (iii)  $\Gamma \subset G_f^c$ .

(i)  $f$  est un infimum de fonctions  $x \mapsto c(x, y_K) - c(x_K, y_K) + c(x_K, y_{K-1}) + \dots + c(x_1, y_0) - c(x_0, y_0)$  "c-affines" en  $x$ , elle est donc  $c$ -concave.

(ii)  $f(x_0) \geq 0$  car il peut s'écrire comme un infimum d'expressions du type  $\Sigma (c(x_i, y_{\sigma(i)}) - c(x_i, y_i))$  où chaque terme de la somme est positif car  $\Gamma$  est  $c$ -cycliquement monotone. En particulier  $f \neq -\infty$ .

(iii) On constate que :

$$\begin{aligned} (x, y) \in \Gamma &\iff f(x) + f^c(y) = c(x, y) \\ &\iff f^c(y) = c(x, y) - f(x) \\ &\stackrel{\text{déf}}{\iff} \forall z, c(z, y) - f(z) \geq c(z, x) - f(x) \end{aligned}$$

C'est cette dernière inégalité que l'on va montrer. Soient donc  $(x_i, y_i)_{i=1 \dots K} \in \Gamma$  tels que pour un certain  $\epsilon \geq 0$  on ait, par définition de l'infimum :

$$f(x) \leq c(x, y_K) - c(x_K, y_K) + \dots + c(x_1, y_0) - c(x_0, y_0) < f(x) + \epsilon$$

En considérant l'ensemble de points  $(x_i, y_i)_{i=1 \dots K+1}$  où  $(x_{K+1}, y_{K+1}) = (x, y)$ , on obtient  $\forall z, f(z) \leq c(z, y) - c(x, y) + c(x, y_K) - \dots - c(x_0, y_0)$ , c'est-à-dire :

$$\begin{aligned} f(z) &\leq c(z, y) - c(x, y) + f(x) + \epsilon \\ &\iff c(z, y) - f(z) > c(x, y) - f(x) + \epsilon \end{aligned}$$

En faisant tendre  $\epsilon$  vers 0, on obtient le résultat.

## 7.2.6 Théorème de Hahn-Banach

Le théorème de HAHN-BANACH, sous ses formes géométriques, donne des résultats de séparation de deux espaces convexes par un hyperplan fermé dans un espace vectoriel normé. Un convexe ouvert non vide, peut être séparé au sens large d'un autre convexe non vide disjoint du premier. Un convexe fermé et un convexe compact disjoints et non vides peuvent être séparés par un hyperplan fermé au sens strict.

## 7.2.7 Relaxation d'une fonction

On est souvent ennuyé lorsque surgit une fonction  $f : \Omega \longrightarrow \mathbb{R} \cup \{+\infty\}$  qui n'a aucune propriété de régularité particulière. La semi-continuité inférieure est une propriété qui est raisonnablement faible tout en donnant des propriétés intéressantes en optimisation, pour cela on aimerait remplacer la fonction  $f$  par une autre fonction  $\bar{f}$ , qui est « proche » de  $f$  mais qui, cette fois, est semi-continue inférieurement et bornée inférieurement.

La meilleure approximation possible vérifiant ces propriétés est naturellement la plus grande fonction semi-continue inférieurement qui est inférieure à  $f$ , que l'on appelle la *relaxée* de  $f$  :

$$\forall x \in \Omega, \bar{f} : x \longmapsto \inf_{x_n \rightarrow x} \liminf_n F(x_n)$$

La fonction  $\bar{f}$  vérifie toutes les propriétés auxquelles on peut s'attendre : elle est semi continue inférieurement et la borne inférieure de  $f$  est la même que celle de  $\bar{f}$ . Pour le premier point il suffit de dire que la borne supérieure de fonctions semi-continue inférieurement est également semi-continue inférieurement (c'est une borne supérieure d'une borne supérieure de fonctions continue). Pour le second point, en notant  $l$  la limite inférieure de  $f$ , on a  $\bar{f}$  qui est supérieure à  $l$  car  $l$  est une fonction semi-continue inférieure qui minore  $f$ , et  $\bar{f}$  qui est inférieure à  $f$  qui est inférieure à  $l$ .

## 7.2.8 Espaces de Sobolev

### 7.2.9 Théorème de FENCHEL – ROCKAFELLAR

Une méthode usuelle et très puissante pour les problèmes de minimisation de critères linéaires sous contraintes convexes est la dualité. L'idée est formellement très simple : puisque la contrainte est convexe, on peut écrire sa fonction indicatrice comme le supremum de fonctions linéaires. Puis, on échange *formellement* la borne inférieure et la borne supérieure apparaissant dans l'expression du problème de minimisation. La « nouvelle » borne inférieure peut-être vue comme l'indicatrice d'une nouvelle contrainte, souvent plus aisée à exploiter que la contrainte initiale. On associe ainsi à chaque problème de minimisation sous contraintes un problème de maximisation sous contraintes.

Bien heureusement cet échange formel n'est pas vain et, sous certaines conditions, les deux problèmes ainsi obtenus sont les mêmes. C'est ce que précise le résultat qui suit :

**Théorème.** *Si  $\phi$  et  $\psi$  sont deux fonctions convexes, finies en  $x_0 \in E$  e.v.n., et si  $\phi$  est continue en  $x_0$ , alors*

$$\inf_{x \in E} (\phi(x) + \psi(x)) = \max_{f \in E'} (-\phi^*(-f) - \psi^*(f))$$

*Preuve.* On commence par travailler l'expression du maximum, en explicitant les transformées de LEGENDRE :

$$\forall x \in E', f^*(x) = \sup_{y \in E} (\langle x, y \rangle - f(y))$$

Il vient alors

$$\begin{aligned} \max_{f \in E'} (-\phi^*(-f) - \psi^*(f)) &= \max_{f \in E'} - (\sup_{x \in E} \langle -f, x \rangle - \phi(x)) - (\sup_{x' \in E} \langle f, x' \rangle - \psi(x')) \\ &= \max_{f \in E'} (\inf_{x \in E} \langle f, x \rangle + \phi(x)) - (\inf_{x' \in E} \langle -f, x' \rangle + \psi(x')) \\ &= \max_{f \in E'} (\inf_{x, x' \in E} \langle f, x - x' \rangle + \phi(x) + \psi(x')) \end{aligned}$$

Avec  $x = x'$ , cette expression devient  $\inf_{x, x' \in E} \phi(x) + \psi(x')$  qui est bien évidemment inférieure à  $m := \inf_{x \in E} \phi(x) + \psi(x)$ . Puisqu'on a supposé  $\phi$  et  $\psi$  finies en  $x_0$ , le minimum  $m$  est fini.

On cherche  $f \in E'$  telle que  $\phi(x) + \psi(x') + \langle f, x - x' \rangle \geq m$ . On introduit

$$\begin{aligned} C &:= \{(x, \lambda) \mid \phi(x) < \lambda\} \\ C' &:= \{(y, \mu) \mid \mu \leq m - \psi(y)\} \end{aligned}$$

Les fonctions étant supposées convexes, il en va de même pour  $C$  et  $C'$ . De plus,  $C$  est d'intérieur non vide car la continuité de  $\phi$  affirme que  $(x_0, \phi(x_0) + 1)$  est dans l'intérieur de  $C$  : il existe un voisinage de  $x_0$  sur lequel  $\phi$  reste à une distance inférieure à 1 de  $\phi(x_0)$ . On sait alors que l'intérieur de  $C$  est égale à son adhérence. De plus,  $C$  et  $C'$  sont disjoints car  $m = \inf \phi + \psi$  donc on ne peut avoir  $\phi(x) + \psi(x) < m - 2\lambda$ . Le théorème de HAHN – BANACH permet de séparer au sens large  $C$  et  $C'$  :

$$\exists l \in (E \times \mathbb{R})' \setminus \{0\}, \inf_{c \in C} \langle l, c \rangle = \inf_{c \in \text{Int}(C)} \langle l, c \rangle \geq \sup_{c' \in C'} \langle l, c' \rangle$$

Ce qui revient, par définition de  $C$  et de  $C'$ , à

$$\exists (f, \alpha) \in (E' \times \mathbb{R}) \setminus \{0\}, \langle f, x \rangle + \alpha \lambda \geq \langle f, x' \rangle + \alpha \mu$$

Ceci implique nécessairement  $\alpha > 0$ .  $f' = f/\alpha$  convient, car alors

$$\exists (f, \alpha) \in (E' \times \mathbb{R}) \setminus \{0\}, \langle f, x \rangle + \lambda \geq \langle f, x' \rangle + \mu$$

Ce qui donne, puisque pour tous  $x$  et  $x'$  on a  $\lambda - \mu > \phi(x) + \psi(x') - m$ ;

$$\exists (f, \alpha) \in (E' \times \mathbb{R}) \setminus \{0\}, \forall (x, x') \in E^2, \langle f, x - x' \rangle + \phi(x) + \psi(x') \geq m$$

Ainsi, on obtient l'égalité voulue.

### 7.2.10 Théorème de RADEMACHER

Une fonction  $h$  de  $\mathbb{R}^n$  dans  $\mathbb{R}$  est localement lipschitzienne. Un théorème de RADEMACHER affirme que le gradient d'une fonction lipschitzienne existe presque partout.

## 7.3 Le transport dans le cas discret

### 7.3.1 Théorème de CHOQUET

**Théorème. (Choquet)** Soient  $K$  un convexe compact d'un espace de BANACH  $E$ ,  $L : E \rightarrow \mathbb{R}$  une fonctionnelle linéaire continue, et  $l : K \rightarrow \mathbb{R}$  sa restriction à  $K$ . Alors  $l$  admet un minimum sur  $K$  qui est un point extrémal de  $K$ .

Rappelons que les points extrémaux d'un ensemble  $K$ , que l'on note  $\mathcal{E}(K)$ , sont les points ne pouvant s'écrire comme combinaison linéaire stricte de plusieurs points de  $K$ . Plus précisément :

$$x \in \mathcal{E}(K) \iff \forall (a, b) \in K^2, \forall \lambda \in [0, 1], (x = \lambda a + (1 - \lambda)b \implies a = b \text{ ou } \lambda \in \{0, 1\})$$

*Preuve.* La continuité de  $L$  et la compacité de  $K$  impliquent que  $l$  atteint son minimum sur  $K$ . On veut maintenant prouver qu'au moins un de ces minima est extrémal.

Nous admettons le théorème de KREIN - MILLMAN qui affirme que tout point de  $K$  est combinaison convexe de points extrémaux de  $K$ . Plus précisément :

**Théorème. (Krein - Millman)**  $\forall x \in K, \exists \rho_x \in \mathbb{P}(K), x = \int_{\mathcal{E}(K)} y d\rho_x(y)$

Le résultat attendu en découle simplement : si aucun des minima n'est extrémal, alors tout point extrémal est strictement supérieur à un minimum  $x_0$ . Par le théorème de KREIN - MILLMAN, on peut trouver une mesure de probabilité  $\rho_{x_0}$  telle que  $x_0$  soit combinaison convexe de points extrémaux. On aurait alors

$$x_0 = \int_{\mathcal{E}(K)} y \rho_{x_0}(y) < \int_{\mathcal{E}(K)} x_0 \rho_{x_0}(y) < x_0$$

ce qui est absurde.  $\square$

Nous pouvons cependant donner une preuve ne faisant pas appel au théorème de KREIN - MILLMAN dans le cas de la dimension finie.

*Preuve alternative.* On se place dans un espace  $E$  de dimension finie. Dans ce cas la fonctionnelle  $L$  est nécessairement continue et l'hypothèse du théorème devient superflue. On commence par écrire  $l$  comme limite croissante de fonctions strictement concaves :

$$l = \lim_{n \rightarrow \infty} l - \frac{1}{n} \|\cdot\|^2$$

Cette convergence est de plus uniforme car  $K$  est borné et donc  $l_n - l = \frac{1}{n} \|\cdot\|^2$  converge uniformément vers 0.

Notons ensuite que l'ensemble des points extrémaux est un compact. En effet, il est fermé et contenu dans le compact  $K$ . (???)

Chaque  $l_n = l - \frac{1}{n} \|\cdot\|^2$  a tous ses minimiseurs  $x_n$  dans  $K$  par compacité, et ceux-là sont sûrement dans  $\mathcal{E}(K)$  par stricte concavité, car

$$x_n = \lambda a + (1 - \lambda)b \implies l_n(x_n) = l_n(\lambda a + (1 - \lambda)b) \geq \lambda l_n(a) + (1 - \lambda)l_n(b) \geq \lambda x_n + (1 - \lambda)x_n = x_n$$

avec égalité si, et seulement si,  $a = b$  ou  $\lambda \in \{0, 1\}$ . On est donc nécessairement dans l'un de ces cas, et  $x_n$  n'est pas combinaison convexe stricte dans  $K$ . On peut alors extraire de la suite  $(x_n)_n$  une sous-suite convergente, que l'on note toujours de la même manière quitte à renommer les indices. Notons  $x$  la limite de la suite extraite, et prouvons que ce  $x$  est un minimiseur de  $l$ . Un simple procédé de convergence diagonale donne le résultat, grâce à la convergence uniforme et à la continuité de  $l$  :

$$\begin{aligned} \forall \varepsilon > 0, \exists n_1 \in \mathbb{N}, \forall n \geq n_1, |l(x_n) - l(x)| &\leq \varepsilon \\ \exists n_2 \in \mathbb{N}, \forall n \geq n_2, \|l_n - l\|_\infty &\leq \varepsilon \end{aligned}$$

Il vient alors par inégalité triangulaire, en choisissant un  $n_0$  supérieur à  $n_1$  et à  $n_2$  :

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, |l_n(x_n) - l(x)| \leq |l_n(x_n) - l(x_n)| + |l(x_n) - l(x)| \leq \|l_n - l\|_\infty + |l(x_n) - l(x)| \leq 2\varepsilon$$

Ainsi  $l_n(x_n)$  converge vers  $l(x)$ . Or les  $l_n$  sont toujours inférieurs à  $l$ , donc en passant à la limite dans  $l_n(x_n) \leq l_n(x')$  qui est valable pour tout  $x'$  de  $K$ , on obtient que  $x$  est un minimiseur de  $l$  sur  $K$ . Comme limite de points extrémaux, c'est un point extrémal car  $\mathbb{E}(K)$  est fermé.  $\square$

### 7.3.2 Théorème de BIRKHOFF

**Théorème. (Birkhoff)** Les points extrémaux de l'ensemble  $\mathcal{B}_n$  des matrices bistochastiques sont les matrices de permutation.

Rappelons que l'ensemble des matrices bistochastiques est l'ensemble des matrices représentant des plans de transfert dans le cas discret :

$$\mathcal{B}_n = \left\{ \gamma = (\gamma_{i,j})_{i,j} \in \mathcal{M}_n([0, 1]) \mid \forall (i_0, j_0) \in \llbracket 1, n \rrbracket^2, \sum_{j=1}^n \gamma_{i_0, j} = 1, \sum_{i=1}^n \gamma_{i, j_0} = 1 \right\}$$

*Preuve.* Tout d'abord, les matrices de permutation sont extrémales car les points extrémaux du segment  $[0, 1]$  sont 0 et 1, ce qui fait que pour une matrice de permutation  $P$  s'écrivant  $\lambda A + (1 - \lambda)B$ , on aura la même relation sur les coefficients à une même place, soit l'un des deux cas suivant :

$$\begin{aligned} 0 &= \lambda a_{i,j} + (1 - \lambda)b_{i,j} \text{ donc } a_{i,j} = b_{i,j} = 0 \text{ car } a_{i,j} \leq 1, b_{i,j} \leq 1 \\ 1 &= \lambda a_{i,j} + (1 - \lambda)b_{i,j} \text{ donc } a_{i,j} = b_{i,j} = 1 \text{ car } a_{i,j} \geq 0, b_{i,j} \geq 0 \end{aligned}$$

Donc  $A = B = P$ , ce qui prouve bien que les matrices de permutations sont extrémales. Cela étant dit, il suffit de prouver que les matrices extrémales de  $\mathcal{B}_n$  sont à coefficients dans  $\{0, 1\}$  pour avoir le résultat. En effet, ce sont exactement les matrices de permutation car il doit il y avoir un élément non nul par ligne (la somme des éléments d'une ligne valant 1) et celui-ci vaut nécessairement 1, il n'y en a donc pas d'autre. La condition de stochasticité sur les colonnes prouve alors que les éléments non nuls sont sur des colonnes distinctes, et que la matrice est donc une matrice de permutation.

Prouvons donc qu'une matrice ayant un coefficient dans  $]0, 1[$  n'est pas extrémale. Notons  $A = (a_{i,j})$  la matrice considérée et supposons que  $a_{i_0, j_0} \in ]0, 1[$ . On va construire une suite de coefficients de  $A$  qui vont pouvoir être perturbés en se compensant les uns les autres :

On part du coefficient  $a_{i_0, j_0}$ . Puisque  $A$  est bistochastique, la somme des coefficients de la ligne  $i_0$  vaut 1 et donc il existe un élément autre coefficient  $a_{i_0, j_1}$  dans  $]0, 1[$ . De même la somme des éléments de la ligne  $j_1$  vaut 1, donc il existe un élément autre coefficient  $a_{i_1, j_1}$  dans  $]0, 1[$ . On construit en suivant le même schéma une suite d'indices dans  $]0, 1[$  ayant chacun leur ligne ou leur colonne en commun avec le précédent, et seulement l'un des deux. On modifie un indice à chaque étape, et les places sont en nombre fini dans  $A$ , donc le *principe des tiroirs* prouve que l'on retombe sur une place déjà visitée. Quitte à ignorer  $a_{i_0, j_0}$  et à commencer par  $a_{i_0, j_1}$ , on peut supposer que la première fois que l'on retrouve un indice connu est  $j_k = j_0$ .

On peut alors perturber la matrice comme suit : on choisit  $\varepsilon > 0$  qui soit inférieur au minimum des distances des coefficients de  $A$  avec 0 ou 1, garantissant ainsi que les  $a_{i,j} \pm \varepsilon$  sont tous dans  $]0, 1[$ . On augmente alors tous les  $a_{i_r, j_r}$  de  $\varepsilon$  et on diminue tous les  $a_{i_r, j_{r+1}}$  de  $\varepsilon$ . Cela ne change pas le caractère bistochastique de la matrice car le couple de manipulations précédent ne modifie pas le poids total d'une ligne, et les répercussions sur les colonnes se compensent : on diminue le poids de la colonne  $i_r$  de  $\varepsilon$ , puis l'étape suivant l'augmente de  $\varepsilon$ . Les deux places limites  $a_{i_0, j_0}$  et  $a_{i_k, j_0}$  se compensent de la même manière.

On a ainsi perturbé la matrice  $A$  sans sortir de  $\mathcal{B}_n$ . Les suites ainsi construites  $(a_{i_r, j_r})_r$  et  $(a_{i_r, j_{r+1}})_r$  sont injectives, car l'algorithme s'arrête au premier retour sur un indice déjà connu, qui est donc le seul à apparaître deux fois, mais qui apparaît une fois dans chaque suite. Elles représentent sont des cycles : notons  $P_1$  et  $P_2$  les matrices représentant ces deux permutations. La perturbation faite précédemment à  $A$  correspondant à  $A + \varepsilon(P_1 - P_2)$ , et est toujours bistochastique. Notons que  $A + \varepsilon(P_2 - P_1)$  est également bistochastique car  $\varepsilon$  a été choisi suffisamment petit pour que les coefficient de cette dernière matrice restent positifs. On trouve alors que

$$A = \frac{1}{2} ((A + \varepsilon(P_1 - P_2)) + (A + \varepsilon(P_2 - P_1)))$$

La matrice  $A$  est donc combinaison convexe stricte de deux matrices bistochastiques : elle n'est pas extrémale.  $\square$

## 7.4 Quelques mots sur la théorie des distributions

Le principe général de la théorie des distributions est de généraliser des propriétés et des opérations que l'on ne peut réaliser que sur un certain type de fonctions à une famille plus large. Pour cela, l'idée est de trouver des caractérisations des propriétés ou des reformulations des opérations que l'on cherche à étendre, qui aient un sens pour des fonctions plus générales. Ainsi pour la dérivation, on remarque que

$$\forall \phi \in \mathcal{D}(\Omega), \int f' \phi = - \int f \phi'$$

où  $f$  est une fonction dérivable, et  $\phi$  parcourt l'ensemble des *fonctions test*  $\mathcal{D}(\Omega)$ , qui est l'ensemble des fonctions infiniment dérivables et à support compact. En effet, il s'agit d'une simple intégration par partie, dont la variation des primitives est nulle car  $\phi$  est supposée à support compact. Les *distributions* sont le dual des fonctions test, autrement dit les formes linéaires continues sur  $\mathcal{D}(\Omega)$ . On peut écrire ce fait pour les fonctions  $f$  qui sont continûment dérivables, mais la quantité  $-\int f \phi'$  est généralisable à toutes les fonctions intégrables : on définit alors la dérivée *au sens des distributions* d'une distribution par

$$\left(\phi \mapsto \int f\phi\right)' = \left(\phi \mapsto -\int f\phi'\right)$$

On note que cela généralise plutôt bien la dérivée des fonctions : en effet, bien que l'on travaille sur des formes linéaires continues particulières  $\phi \mapsto \int f\phi$ , celles-ci représentent relativement bien les fonctions  $f$  de départ, elles les caractérisent modulo l'égalité presque partout. On remplace ainsi la connaissance des valeurs ponctuelles de  $f$  par la connaissance de ses moyennes pondérées sur des ensembles compacts, ce qui ne nuit que peu à la généralité (puisqu'il s'agit essentiellement de la même connaissance, à quelques différences qui sont négligeables au sens de la mesure), et cela permet de généraliser toutes les opérations de dérivation et de différentiation.

Notamment, au lieu de rechercher des solutions d'équations aux dérivées partielles du type

$$\frac{\partial}{\partial t}u + \Delta u = 0$$

On se permet d'étendre notre recherche aux fonctions – qui ne sont plus nécessairement supposées dérivables – vérifiant

$$\forall \phi \in \mathcal{D}(\Omega), -\int u \frac{\partial}{\partial t}\phi + \int u \Delta \phi = 0$$

On obtient ainsi des solutions appelées *solutions faible* de l'équation. On se limite souvent à la recherche de telles solutions, la forme de l'équation devenant bien plus simple à manier et les problèmes de régularité des fonctions devenant secondaire, sans perdre en généralité puisque les solutions faibles suffisamment régulières sont exactement les solutions de l'équation de départ.

# Chapitre 8

## Références

### **Théorie générale du transport et applications**

- [1] C. Villani, *Topics in optimal transportation*, AMS, 2003.
- [2] C. Villani, *Optimal transport : old and new*, Springer, 2009.

### **Fondements théoriques et outils**

- [3] H. Brézis, *Analyse Fonctionnelle*, McGraw-Hill, Inc., 1987.
- [4] W. Rudin, *Real and Complex Analysis*, Dunod, 1999.