

— Rapport de projet —  
Transport Optimal

Département de Génie Mathématique  
Semestre 9 - 2017/2018

*Auteur :*  
Timothée SCHMODERER

*Référente :*  
Carole LE GUYADER  
Vincent DUVAL

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Problème de Transport Optimal</b>	<b>4</b>
2.1	Problème de Transport . . . . .	4
2.2	Plan de Transport Optimal . . . . .	5
2.3	Problème dual de plan de Transport Optimal . . . . .	7
2.4	Existence du transport optimal . . . . .	12
2.5	Le cas quadratique sur $\mathbb{R}^n$ . . . . .	13
2.6	Formulation de J.D. Benamou et Y. Brenier . . . . .	15
2.7	Une autre utilité au transport optimal . . . . .	17
2.8	Conclusion . . . . .	18
<b>3</b>	<b>Opérateurs proximaux</b>	<b>19</b>
<b>4</b>	<b>Pbm de transport optimal</b>	<b>26</b>
<b>5</b>	<b>Formulation de Benamou et Brenier</b>	<b>26</b>
<b>6</b>	<b>Méthodes numériques</b>	<b>26</b>
<b>7</b>	<b>Opérateurs</b>	<b>27</b>
7.1	Interpolation . . . . .	27
7.2	Divergence . . . . .	27
7.3	Frontières . . . . .	27
7.4	Problème discrétisé . . . . .	27
<b>8</b>	<b>Résolution par algorithme de séparation de proximité</b>	<b>28</b>
8.1	Opérateur de proximité de $G_2$ . . . . .	28
8.2	Opérateur de proximité de $J$ . . . . .	29
8.3	Opérateur de proximité de $\iota_C$ . . . . .	29
<b>9</b>	<b>Algorithme de Douglas Rachford</b>	<b>29</b>
	<b>Annexes</b>	<b>30</b>
<b>A</b>	<b>Calcul des variations</b>	<b>30</b>
	<b>Table des figures</b>	<b>33</b>
	<b>Liste des Codes</b>	<b>33</b>

# 1 Introduction

Le problème de transport optimal fut étudié dans sa forme primitive par *Gaspard Monge* dans son *Mémoire sur la théorie des déblais et des remblais* de 1781 [?]. Il s'intéressait au déplacement de tas de matériaux d'un site d'extraction à un site de construction en minimisant le coût de transport (le coût de déplacement d'une unité est le produit de la charge déplacée par la distance).

La formulation moderne a été introduite par *Leonid Kantorovitch*, prix Nobel d'économie en 1975 pour sa contribution aux allocations optimales.

Nous essaierons d'être le plus exhaustif possible dans la présentation des résultats, la plupart des énoncés seront démontrés.

Rendons à Santambrogio ce qui lui appartient, la première section vient en quasi-entièreté de son livre [?], agrémenté de quelques précisions qui faciliteront la compréhension.

Le plan de rapport est le suivant, nous commencerons par présenter les notions liées au transport optimal, puis nous ferons un court aparté sur les opérateurs proximaux. Enfin,

Nous attaquerons la discrétisation et la mise en œuvre numérique. Plan :

- 1 - Théorie sur le transport optimal
- 2 - Théorie sur les opérateurs proximaux
- 3 - Numérique
  - 3.1 - Idées générales de l'implémentation
  - 3.2 - Implémentation Grosse matrice
  - 3.3 - Implémentation centrée
    - 3.3.1 - 1D
    - 3.3.2 - 2D
  - 3.4 - Implémentation staggered FFT
- 4 - Galerie du Fun

## 2 Problème de Transport Optimal

Introduisons des notions essentielles qui nous amèneront à définir le problème de transport optimal et sa formulation en terme de géodésiques.

### 2.1 Problème de Transport

La théorie du transport prend sa source dans la théorie de la mesure, en particulier de des mesures de probabilité.

#### Définition 2.1 : *Transport*

Soient deux mesures de probabilité  $\mu$  sur  $\mathcal{X}$  et  $\nu$  sur  $\mathcal{Y}$  munis de leur tribu borélienne  $\mathcal{B}$ . Un transport est une application  $T : \mathcal{X} \rightarrow \mathcal{Y}$  qui envoie la mesure  $\mu$  sur  $\nu$ . C'est à dire :

$$\forall B \in \mathcal{B}, \quad \mu(T^{-1}(B)) = \nu(B) \quad (1)$$

C'est une relation de conservation de masse. Nous notons  $T_{\#}\mu = \nu$ .

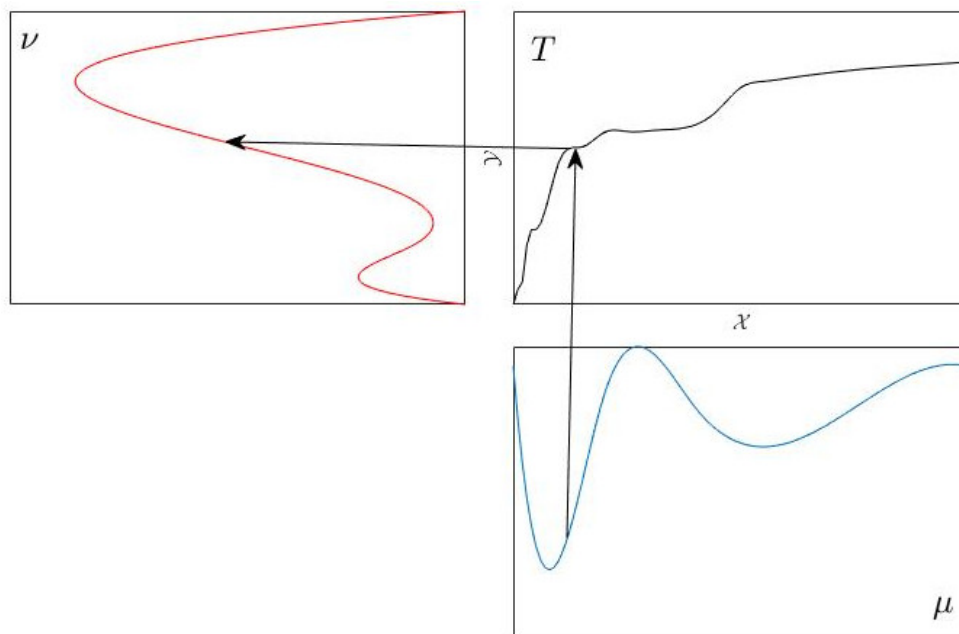


FIGURE 1 – Illustration de l'application de transport

Bien sur, l'application de transport n'est pas unique, il nous faut une manière de sélectionner un transport qui serait "meilleur" que les autres.

#### Définition 2.2 : *Coût*

Le coût est une application

$$C : \begin{cases} \mathcal{X} \times \mathcal{Y} & \longrightarrow [0, +\infty] \\ (x, y) & \longmapsto C(x, y) \end{cases}$$

Cette application va représenter le coût d'affecter  $x$  en  $y$ . Cela nous amène au problème de transport optimal :

Étant donné deux mesures de probabilités  $\mu$  et  $\nu$  et un coût  $C$ , trouver le transport  $T$  réalisant le

$$\inf \left\{ M(T) := \int_{\mathcal{X}} C(x, T(x)) d\mu(x), \quad T_{\#}\mu = \nu \right\} \quad (\text{MP})$$

L'existence du transport optimal n'est absolument pas trivial, suivons la présentation du livre de Santambrogio [?] pour démontrer l'existence d'un transport optimal.

## 2.2 Plan de Transport Optimal

Le problème (MP) est difficile à résoudre du fait de la contrainte. Oublions ce problème pour un temps, et regardons une forme généralisée donnée par Kantorovitch.

Étant donné deux mesures de probabilités  $\mu$  et  $\nu$  et un coût  $C$ , trouver la mesure  $\gamma$  réalisant le

$$\inf \left\{ K(\gamma) := \int_{\mathcal{X} \times \mathcal{Y}} C(x, y) d\gamma(x, y), \quad \gamma \in \Pi(\mu, \nu) \right\} \quad (\text{KP})$$

Où,  $\Pi(\mu, \nu)$  est l'ensemble des plans de transports :

$$\Pi(\mu, \nu) = \left\{ \begin{array}{l} \gamma \text{ une probabilité sur } \mathcal{X} \times \mathcal{Y} \\ \forall A \in \mathcal{B}(\mathcal{X}), \quad \gamma(A \times \mathcal{Y}) = \mu(A) \\ \forall B \in \mathcal{B}(\mathcal{Y}), \quad \gamma(\mathcal{X} \times B) = \nu(B), \end{array} \right. \quad (2)$$

La valeur de  $\gamma(A \times B)$  décrit la quantité de matière transportée de  $A$  en  $B$ . Cette description permet des déplacements plus général qu'un transport classique. A partir d'un point  $x$ , les particules peuvent être transportées en plusieurs destinations. Si c'est le cas, alors cela ne peut pas être décrit au travers de l'application  $T$ . Remarquons que les contraintes (2) sont des contraintes de marginales sur la mesure de probabilité  $\gamma$  et que cela signifie que nous restreignons notre attention sur aux transports qui déplacent des particules distribuées selon  $\mu$  sur des particules distribuées selon  $\nu$ .

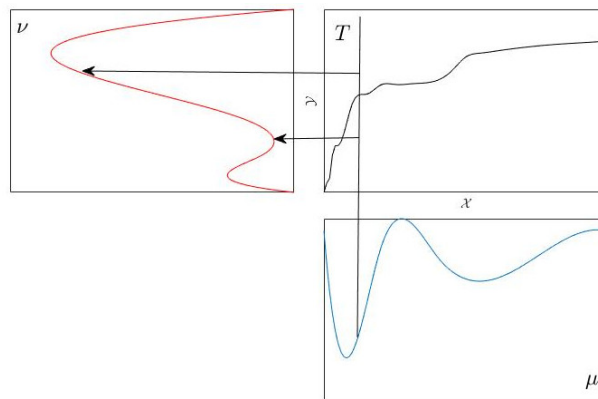


FIGURE 2 – Illustration de la notion de plan de transport

Les minimiseurs de ce problème sont appelés plans de transport optimal entre  $\mu$  et  $\nu$ . Ce problème est bien une généralisation du problème (MP), puisque si  $\gamma = (Id, T)$  pour une certaine fonction mesurable  $T : \mathcal{X} \rightarrow \mathcal{Y}$ . Alors l'application  $T$  serait un transport optimal entre  $\mu$  et  $\nu$ .

Nous allons montrer l'existence de solutions pour ce problème relaxé (KP) à l'aide du calcul des variations. Les outils principaux de la méthode du calcul des variations sont rappelés à l'annexe A.

### **Théorème 2.3 :**

Soit  $\mathcal{X}$  et  $\mathcal{Y}$  des espaces métriques compacts. Supposons que le coût  $C : \mathcal{X} \times \mathcal{Y} \rightarrow [0, +\infty]$  soit continue. Alors le problème (KP) admet une solution.

#### **Preuve :**

Il s'agit de montrer que l'ensemble  $\Pi(\mu, \nu)$  est compact et que l'application  $\gamma \rightarrow K(\gamma)$  est continue, puis d'appliquer le théorème de Weierstrass (A.3).

Nous choisissons la notion de convergence donnée par la convergence faible des mesures de probabilités (voir (A.9) par dualité avec  $C_b(\mathcal{X} \times \mathcal{Y})$  ce qui ici est la même chose que  $C(\mathcal{X} \times \mathcal{Y})$  ou  $C_0(\mathcal{X} \times \mathcal{Y})$  car  $\mathcal{X}$  et  $\mathcal{Y}$  sont compacts). Cela nous donne la continuité de  $K$  puisque  $C \in C(\mathcal{X} \times \mathcal{Y})$ .

Soit une suite  $\gamma_n \in \Pi(\mu, \nu)$ . Ce sont des mesures de probabilités, donc de masse 1 et sont donc bornées dans le dual de  $C(\mathcal{X} \times \mathcal{Y})$ . Ainsi, la compacité faible-\* dans les espaces duaux, garanti l'existence d'une sous suite  $\gamma_{n_k} \rightharpoonup \gamma$  convergeant vers une probabilité  $\gamma$ . Vérifions que  $\gamma \in \Pi(\mu, \nu)$ .

Fixons  $\phi \in C(\mathcal{X})$ , comme  $\int \phi d\gamma_{n_k} = \int \phi d\mu$ , par passage à la limite, on a :  $\int \phi d\gamma = \int \phi d\mu$ . De même on montre que la seconde contrainte de marginale (2) est vérifiée.

Ainsi, l'espace  $\Pi(\mu, \nu)$  est compact et le théorème de Weierstrass s'applique.  $\square$

Allons vers plus de généralités.

### **Proposition 2.1 :**

Supposons que le coût soit seulement semi continue inférieurement et bornée en dessous. Alors le problème (KP) admet une solution.

#### **Preuve :**

La seule différence avec le cas précédent est que  $K$  n'est plus continue mais semi continue inférieurement pour la convergence faible des mesures de probabilités. Cela vient du lemme suivant appliqué à  $f = C$  et  $\Omega = \mathcal{X} \times \mathcal{Y}$ .

**Lemme :**

Si  $f : \Omega \rightarrow \mathbb{R} \cup \{\infty\}$  est une fonction semi-continue inférieurement bornée par dessous sur un espace métrique  $\Omega$ . Alors la fonctionnel  $J : \mathcal{M}_+(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$  définie sur les mesures positives par  $J(\lambda) = \int f d\lambda$  est semi continue inférieurement pour la convergence faible des mesures.

**Preuve :**

Soit une suite  $(f_k)$  de fonctions continues, bornées convergeant vers  $f$  par valeurs croissantes. Soit  $J_k(\lambda) = \int f_k d\lambda$ ,  $J_k \rightarrow J$  par convergence monotone. Nous pouvons alors écrire  $J(\lambda) = \sup_k J_k(\lambda)$ . Alors  $J$  est semi-continue inférieurement comme le sup de fonctionnelles continues.  $\square$

Et donc, le théorème de Weierstrass (A.3) s'applique.  $\square$

**Proposition 2.2 :**

Supposons que  $\mathcal{X}$  et  $\mathcal{Y}$  sont des espaces métriques complets et séparables (i.e. polonais). Supposons toujours que le coût  $C$  est semi-continue inférieurement et bornée par dessous. Alors le problème (KP) admet une solution.

**Preuve :**

Cette fois, c'est la compacité de  $\Pi(\mu, \nu)$  qui est moins évidente. Nous allons utiliser le théorème de Prokhorov (A.9). Ce qui revient à montrer que toute suite de  $\Pi(\mu, \nu)$  est tendue.

Soient  $\epsilon > 0$  et deux compact  $K_{\mathcal{X}} \subset \mathcal{X}$  et  $K_{\mathcal{Y}} \subset \mathcal{Y}$  tels que  $\mu(\mathcal{X} \setminus K_{\mathcal{X}}), \nu(\mathcal{Y} \setminus K_{\mathcal{Y}}) < \frac{1}{2}\epsilon$ . Cela est possible par la réciproque du théorème de Prokhorov, car une mesure seule est toujours tendue. Alors l'ensemble  $K_{\mathcal{X}} \times K_{\mathcal{Y}}$  est compact dans  $\mathcal{X} \times \mathcal{Y}$ , et pour toute suite  $\gamma_n \in \Pi(\mu, \nu)$ , nous avons

$$\begin{aligned} \gamma_n((\mathcal{X} \times \mathcal{Y}) \setminus (K_{\mathcal{X}} \times K_{\mathcal{Y}})) &\leq \gamma_n((\mathcal{X} \setminus K_{\mathcal{X}}) \times \mathcal{Y}) + \gamma_n(\mathcal{X} \times (\mathcal{Y} \setminus K_{\mathcal{Y}})) \\ &\leq \mu(\mathcal{X} \setminus K_{\mathcal{X}}) + \nu(\mathcal{Y} \setminus K_{\mathcal{Y}}) < \epsilon \end{aligned}$$

Ce qui montre la tension et donc la compacité de toute suite de  $\Pi(\mu, \nu)$ .  $\square$

## 2.3 Problème dual de plan de Transport Optimal

Maintenant que nous avons démontré l'existence de solution au problème de Kantorovitch, nous souhaiterions savoir sous quelles conditions nous pouvons remonter au problème de Monge. Comme le problème (KP) est un problème d'optimisation linéaire sous contraintes d'égalité, il est donc naturel d'étudier le problème dual.

Toujours suivant la présentation du livre de Santambrogio [?], cherchons en premier lieu le problème dual formellement.

La contrainte  $\gamma \in \Pi(\mu, \nu)$  peut s'exprimer sous la forme :

$$\begin{aligned} \sup_{(\phi, \psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{Y})} \int_{\mathcal{X}} \phi d\mu(x) + \int_{\mathcal{Y}} \psi d\mu(y) - \int_{\mathcal{X} \times \mathcal{Y}} \phi(x) + \psi(y) d\gamma(x, y) \\ = \begin{cases} 0 & \text{si } \gamma \in \Pi(\mu, \nu) \\ +\infty & \text{sinon} \end{cases} \end{aligned} \quad (3)$$

Ainsi nous pouvons supprimer la contrainte dans le problème (KP) en ajoutant l'expression (3), puisque si la contrainte est satisfaite, rien n'est ajoutée, si elle ne l'est pas nous obtenons  $+\infty$  ce qui sera évité par la minimisation.

$$\inf_{\gamma} \int_{\mathcal{X} \times \mathcal{Y}} C(x, y) d\gamma + \sup_{(\phi, \psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{Y})} \int_{\mathcal{X}} \phi d\mu(x) + \int_{\mathcal{Y}} \psi d\mu(y) - \int_{\mathcal{X} \times \mathcal{Y}} \phi(x) + \psi(y) d\gamma(x, y)$$

Il n'est pas toujours possible d'échanger le sup et l'inf, mais supposons pour le moment que cela soit possible :

$$\sup_{(\phi, \psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{Y})} \int_{\mathcal{X}} \phi d\mu(x) + \int_{\mathcal{Y}} \psi d\mu(y) + \inf_{\gamma} \int_{\mathcal{X} \times \mathcal{Y}} (C(x, y) - \phi(x) + \psi(y)) d\gamma(x, y) \quad (4)$$

Avec ce problème de maximisation sur les variables  $(\phi, \psi)$ , l'infimum en  $\gamma$  peut être réécrit comme une contraintes sur  $(\phi, \psi)$  :

$$\inf_{\gamma} \int_{\mathcal{X} \times \mathcal{Y}} (C - \phi \oplus \psi) d\gamma = \begin{cases} 0 & \text{si } \phi \oplus \psi \leq C \text{ sur } \mathcal{X} \times \mathcal{Y} \\ -\infty & \text{sinon} \end{cases} \quad (5)$$

Où la fonction,  $\phi \oplus \psi$  est définie par  $(\phi \oplus \psi)(x, y) = \phi(x) + \psi(y)$ . L'égalité précédente est justifiée car, si  $\phi \oplus \psi > c$  quelque part alors en utilisant des mesures  $\gamma$  concentrée en cet endroit et de masse croissante, l'intégrale tend vers  $-\infty$ . Ce qui nous amène au problème suivant :

Étant donné deux mesures de probabilités  $\mu$  et  $\nu$  et un coût  $C$ , trouver deux fonctions continues, bornée  $\phi$  et  $\psi$  réalisant le

$$\sup \left\{ \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\mu, \quad (\phi, \psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{Y}) \quad \phi \oplus \psi \leq C \right\} \quad (\text{DKP})$$

Remarquons que le  $\sup (\text{DKP}) \leq \inf (\text{KP})$ , en intégrant la condition  $\phi \oplus \psi \leq c$  par rapport à  $\gamma$  :

$$\int_{\mathcal{X}} \phi d\mu(x) + \int_{\mathcal{Y}} \psi d\mu(y) = \int_{\mathcal{X} \times \mathcal{Y}} \phi \oplus \psi d\gamma \leq \int_{\mathcal{X} \times \mathcal{Y}} C d\gamma$$

Malheureusement le problème (DKP) n'admet pas trivialement l'existence de ses maximums, car l'espace des fonctions continues bornées n'est pas compact. Montrons dans un premier temps l'existence de ces maximiseurs, puis montrons que le problème (DKP) est bien le dual de (KP) au sens ou  $\max (\text{DKP}) = \min (\text{KP})$ .



**Définition 2.4** :  $C$  et  $\bar{C}$ -transformée,  $C$  et  $\bar{C}$ -concavité

Soit une fonction  $\chi : \mathcal{X} \rightarrow [-\infty, +\infty]$ . La  $C$ -transformée (ou parfois,  $C$  - conjuguée) de  $\chi$  est définie par  $\chi^C : \mathcal{Y} \rightarrow [-\infty, +\infty]$  :

$$\chi^C(y) = \inf_{x \in \mathcal{X}} C(x, y) - \chi(x)$$

De même, la  $\bar{C}$ -transformée est définie pour une fonction  $\tau : \mathcal{Y} \rightarrow [-\infty, +\infty]$  par

$$\tau^{\bar{C}} = \inf_{y \in \mathcal{Y}} C(x, y) - \tau(y)$$

De plus, nous dirons qu'une fonction  $\psi$  définie sur  $\mathcal{Y}$  est  $\bar{C}$ -concave si il existe  $\chi : \mathcal{X} \rightarrow [-\infty, +\infty]$  telle que  $\psi = \chi^C$ . Respectivement, une fonction  $\phi$  définie sur  $\mathcal{X}$  est dite  $C$ -concave si il existe  $\tau : \mathcal{Y} \rightarrow [-\infty, +\infty]$  telle que  $\phi = \tau^{\bar{C}}$ .

Nous noterons  $C\text{-conc}(\mathcal{X})$  et  $\bar{C}\text{-conc}(\mathcal{Y})$  les ensembles des fonctions  $C$  et  $\bar{C}$ -concaves.

Remarquons que lorsque  $\mathcal{X} = \mathcal{Y}$  et que  $C$  est symétrique, il n'y a aucune distinction entre la transformation  $C$  et  $\bar{C}$ , cette notation lourde sera alors abandonnée.

Dans notre cas,  $C$  est continue sur un ensemble compact, et est donc uniformément continue. C'est à dire qu'il existe une fonction continue croissante  $\omega : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  et  $\omega(0) = 0$  (module de continuité) telle que

$$|C(x, y) - C(x', y')| \leq \omega(d(x, x') + d(y, y'))$$

En reprenant la définition de  $\chi^C(y) = \inf_x g_x(y)$  avec  $g_x(y) = C(x, y) - \chi(x)$  qui vérifie  $|g_x(y) - g_x(y')| \leq \omega(d(y, y'))$ . Ce qui montre que  $\chi^C$  et  $C$  partagent le même module de continuité.

Ensuite remarquons qu'étant une paire  $(\phi, \psi)$  du problème (DKP), elle peut toujours être remplacée par la paire  $(\phi, \phi^C)$  puis par  $(\phi^{C\bar{C}}, \phi^C)$ , car les contraintes sont préservées et les intégrales augmentées. Nous pourrions espérer itérer ce processus, mais pour toute fonction  $\phi$ , nous avons  $\phi^{C\bar{C}C} = \phi$ . Cependant les considérations précédentes nous amènent le théorèmes d'existence suivant.

**Théorème 2.5** :

Supposons que  $\mathcal{X}$  et  $\mathcal{Y}$  soient compact et que  $C$  est continue. Alors il existe une solution  $(\phi, \psi)$  au problème (DKP), de la forme  $\phi \in C\text{-conc}(\mathcal{X})$ ,  $\psi \in \bar{C}\text{-conc}(\mathcal{Y})$  et  $\psi = \phi^C$ . En particulier,

$$\max(\text{DKP}) = \max_{\phi \in C\text{-conc}(\mathcal{X})} \int_{\mathcal{X}} \phi \, d\mu + \int_{\mathcal{Y}} \phi^C \, d\nu \quad (6)$$

**Preuve :**

Prenons une suite maximisante  $(\phi_n, \psi_n)$  et améliorons là par la  $C$  et la  $\bar{C}$ -transformée. Par les considérations précédentes, nous pouvons supposer une borne uniforme sur la

continuité de ces fonctions (le même module de continuité que  $C$ ). Gardons la notation  $(\phi, \psi)$  pour désigner cette suite améliorée.

Vérifions qu'elle est équibornée pour appliquer le théorème d'Ascoli Arzelà (A.10). Remarquons qu'ajouter une constante à  $\phi$  et la retirer à  $\psi$  ne change pas la valeur de la fonctionnelle et n'affecte pas les contraintes. Ainsi, comme  $\phi_n$  est continue sur un ensemble compact, elle est donc bornée, quitte à ajouter le minimum, nous pouvons supposer que  $\min \phi_n = 0$ , nous avons aussi que  $\max \phi_n \leq \omega(d(\mathcal{X}))$ , car les oscillations d'une fonctions sont toujours plus faible que son module de continuité appliqué à la plus grande distance possible dans l'ensemble. Comme nous avons choisis  $\psi_n = \phi_n^C$ , nous avons également,  $\psi_n(y) = \inf_x C(x, y) - \phi_n(x) \in [\min C - \omega(d(\mathcal{X})), \max C]$ . Ce qui nous donne des bornes uniformes sur  $(\phi_n, \psi_n)$  et nous permet d'appliquer le théorème d'Ascoli - Arzelà.

A une sous suite près,  $\phi_n \rightarrow \phi$  et  $\psi_n \rightarrow \psi$  avec convergence uniforme, de ce fait :

$$\int_{\mathcal{X}} \phi_n d\mu + \int_{\mathcal{Y}} \psi_n d\nu \rightarrow \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu \quad (7)$$

De plus,

$$\phi_n(x) + \psi_n(y) \leq C(x, y) \Rightarrow \phi(x) + \psi(y) \leq C(x, y)$$

Ici la convergence simple aurait suffit. Cela montre que  $(\phi, \psi)$  est une paire admissible pour (DKP) et optimale.  $\square$

Montrons à présent que (DKP) est bien le problème dual associé à (KP).

### **Théorème 2.6 :**

Les problèmes (KP) et (DKP) sont bien duaux l'un de l'autre :

$$\min (\text{KP}) = \max (\text{DKP}) \quad (8)$$

### **Preuve :**

Par simplicité, prenons le cas  $\mathcal{X}$  et  $\mathcal{Y}$  compacts, et  $C$  continue.

Pour tout  $p \in C(\mathcal{X} \times \mathcal{Y})$ , posons

$$H(p) = - \max \left\{ \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu, \quad \phi \oplus \psi \leq C - p \right\}$$

C'est l'opposée de la valeur de (DKP) pour le coût  $C - p$ . Par la proposition, nous avons que,  $H(p)$  est bien atteint et le module de continuité de la paire optimale  $(\phi, \psi)$  est le même que celui de  $C - p$ .

### **Lemme :**

La fonction  $H : C(\mathcal{X} \times \mathcal{Y}) \rightarrow \mathbb{R}$  est convexe et semi continue inférieurement pour la convergence uniforme.

**Preuve :**

Pour la convexité, prenons  $p_0$  et  $p_1$  associé avec leur paire optimale  $(\phi_0, \psi_0)$  et  $(\phi_1, \psi_1)$ . Pour  $t \in [0, 1]$  posons,  $p_t = (1-t)p_0 + tp_1$ ,  $\phi_t = (1-t)\phi_0 + t\phi_1$  et  $\psi_t = (1-t)\psi_0 + t\psi_1$ . La paire  $(\phi_t, \psi_t)$  est admissible dans le maximum définissant  $H(p_t)$ . Ainsi,

$$H(p_t) \leq - \left( \int_{\mathcal{X}} \phi_t d\mu + \int_{\mathcal{Y}} \psi_t d\nu \right) = (1-t)H(p_0) + tH(p_1)$$

Ce qui montre la convexité. Pour la semi continuité, prenons  $p_n \rightarrow p$ . A une sous suite près,  $p_n$  réalise la  $\liminf$  de  $H(p_n)$ . De la convergence uniforme, la suite  $p_n$  est équicontinue et bornée (réciproque de théorème d'Ascoli - Arzelà (A.10)). Ainsi, les paires optimales correspondante  $s(p_n, \psi_n)$  sont aussi équicontinues et bornées. A une sous suite près, supposons que  $\phi_n \rightarrow \phi$  et  $\psi_n \rightarrow \psi$  uniformément. Comme  $\phi_n \oplus \psi_n \leq C - p$ , nous avons  $\phi \oplus \psi \leq C - p$ , ainsi :

$$H(p) \leq - \left( \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu \right) = \lim_n H(p_n) = \liminf_n H(p_n)$$

ce qui montre la semi continuité inférieure.  $\square$

Calculons à présent la transformée de Legendre<sup>1</sup> de  $H$ .  $H^* : \mathcal{M}(\mathcal{X} \times \mathcal{Y}) \rightarrow \mathbb{R} \cup \{+\infty\}$ . Pour tout  $\gamma \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ , nous avons

$$H^*(\gamma) = \sup_p \int_{\mathcal{X} \times \mathcal{Y}} p d\gamma + \sup \left\{ \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu, \quad \phi \oplus \psi \leq C - p \right\}$$

Ce qui se réécrit sous la forme d'un sup unique sur  $p, \phi, \psi$ . Remarquons que, si  $\gamma \notin \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$ , alors il existe  $p_0 \leq 0$  tel que  $\int p_0 d\gamma > 0$ , en prenant  $\phi \equiv 0, \psi \equiv 0$  et  $p = C + np_0$  en faisant tendre  $n$  vers  $+\infty$  nous obtenons  $H^*(\gamma) = +\infty$ . D'autre part, si  $\gamma \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$ , en prenant le  $p$  le plus grand possible i.e.  $p = C - \phi - \psi$ , nous obtenons

$$H^*(\gamma) = \sup_{(\phi, \psi)} \int_{\mathcal{X} \times \mathcal{Y}} p d\gamma + \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} \phi d\gamma - \int_{\mathcal{X} \times \mathcal{Y}} \psi d\gamma$$

C'est exactement l'expression obtenue au (3) pour réécrire les contraintes de (KP).

$$H^*(\gamma) = \begin{cases} K(\gamma) & \text{si } \gamma \in \Pi(\mu, \nu) \\ +\infty & \text{sinon} \end{cases}$$

La preuve est presque terminée,  $\max(\text{DKP}) = -H(0) = H^{**}(0)$  car  $H$  est convexe et semi continue inférieurement<sup>2</sup>. De plus,  $H^{**}(0) = \sup_{\gamma} \langle \gamma, \gamma \rangle - H^*(\gamma) = -\inf H^* = -\min(\text{KP})$ .  $\square$

Ainsi, nous voyons que

$$\min(\text{KP}) = \max_{\phi \in C\text{-conc}(\mathcal{X})} \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \phi^C d\nu$$

Ce qui montre que la valeur du minimum de (KP) est une fonction convexe de  $\mu$  et  $\nu$ , comme supremum de fonctionnelle linéaire.

**Définition 2.7 :** *Potentiel de Kantorovitch*

Les fonctions  $\phi$  réalisant le maximum de (6) sont appelés potentiels de Kantorovitch pour le transport de  $\mu$  sur  $\nu$ .

**2.4 Existence du transport optimal**

A partir des résultats précédents sur les problèmes (KP) et son dual (DKP), nous allons montrer l'existence du transport optimal dans le cas où,  $\mathcal{X} = \mathcal{Y} = \Omega \subset \mathbb{R}^n$  compact et  $C(x, y) = h(x - y)$  pour une certaine fonction strictement convexe  $h$ .

Rappelons que pour une mesure  $\gamma$ , son support est le plus petit ensemble fermé sur lequel  $\gamma$  est concentré :

$$\text{supp } \gamma = \bigcap \{A \mid A \text{ fermé}, \gamma(\mathcal{X} \setminus A) = 0\}$$

une autre caractérisation du support de  $\gamma$  est donnée par

$$\text{supp } \gamma = \{x \in \mathcal{X} \mid \gamma(B(x, r)) > 0 \forall r > 0\}$$

De l'égalité entre le minimum de (KP) et le maximum de (DKP) et le fait que les deux soient atteints, considérons un plan de transport optimal  $\gamma$  et un potentiel de Kantorovitch  $\phi$ ,

$$\phi(x) + \phi^C(y) \leq C(x, y) \text{ sur } \Omega \times \Omega \quad (9)$$

$$\phi(x) + \phi^C(y) = C(x, y) \text{ sur } \text{supp } \gamma \quad (10)$$

L'égalité vient de l'inégalité (qui est vraie partout) et de

$$\int_{\Omega \times \Omega} C \, d\gamma = \int_{\Omega} \phi \, d\mu + \int_{\Omega} \phi^C \, d\nu = \int_{\Omega \times \Omega} \phi \oplus \phi^c \, d\gamma \quad (11)$$

ce qui implique l'égalité  $\gamma$  presque partout.

**Proposition 2.3 :**

Si  $C \in C^1(\Omega \times \Omega, [0, +\infty])$ ,  $\phi$  un potentiel de Kantorovitch pour le coût  $C$  pour le transport de  $\mu$  vers  $\nu$ . Soit  $(x_0, y_0) \in \text{supp } \gamma$ . Alors, si  $\phi$  est différentiable en  $x_0$ , on a  $\nabla \phi(x_0) = \frac{\partial}{\partial x} C(x_0, y_0)$ .

En particulier les gradients de deux potentiels de Kantorovitch coïncident en tout point  $x_0 \in \text{supp } \gamma$  où ils sont tous les deux différentiables.

**Preuve :**

Soit  $(x_0, y_0) \in \text{supp } \gamma$  telle que  $\phi$  soit différentiable en  $x_0$ , des calculs précédents, nous déduisons que l'application  $x \mapsto \phi(x) - c(x, y_0)$  atteint son minimum en  $x = x_0$ . Comme  $\phi$  et  $C(\cdot, y_0)$  sont différentiables, nous obtenons directement le résultat souhaité.  $\square$

**Théorème 2.8 :** *Existence du transport optimal*

Soient  $\mu$  et  $\nu$  des mesures de probabilités sur un domaine compact  $\Omega$  de  $\mathbb{R}^n$ , alors il existe un plan de transport optimal  $\gamma$  pour le coût  $C(x, t) = h(x - y)$ , avec  $h$  strictement convexe. De plus,  $\gamma$  est unique, de la forme  $(Id, T)_\# \mu$  si  $\mu$  est absolument continue et  $\partial\Omega$  est négligeable. Enfin, il existe un potentiel de Kantorovitch  $\phi$  qui est lié à  $T$  par :

$$T(x) = x - (\partial h)^{-1}(\nabla \phi(x)) \quad (12)$$

### Preuve :

Le théorème précédent nous donne l'existence d'un plan de transport optimal  $\gamma$  et d'un potentiel de Kantorovitch  $\phi$ .

$\phi$  partage le même module de continuité que  $C$ , c'est une fonction lipschitzienne sur  $\Omega \times \Omega$  puisque  $h$  est localement lipschitzienne et bornée sur  $\Omega$ . Ainsi,  $\phi$  est aussi lipschitzienne.

Pour  $C(x, y) = h(x - y)$ , si  $h$  est différentiable, nous obtenons par la proposition précédente que  $\nabla \phi(x_0) = \nabla h(x_0 - y_0)$ , si ce n'est pas le cas, nous avons  $\nabla \phi(x_0) \in \partial h(x_0 - y_0)$  le sous différentiel de  $h$ <sup>3</sup>. Comme  $h$  est strictement convexe,  $(\partial h)^{-1}$  existe, et nous avons

$$x_0 - y_0 = (\partial h)^{-1}(\nabla \phi(x_0))$$

Comme  $\mu$  est absolument continue,  $\partial\Omega$  et l'ensemble des points où  $\phi$  n'est pas différentiable sont négligeables (par le théorème de Rademacher).

Cela montre que le plan de transport optimal est induit par un transport, que ce transport est donnée par  $x \mapsto x - (\partial h)^{-1}(\nabla \phi(x_0))$ . Ainsi il est uniquement déterminé, car  $\phi$  ne dépend pas de  $\gamma$ . En conséquence,  $\gamma$  est également unique.  $\square$

Ainsi, tout les coûts de la forme  $C(x, y) = \|x - y\|^p$  avec  $p > 1$  sont traités.

## 2.5 Le cas quadratique sur $\mathbb{R}^n$

Introduisons petit à petit des hypothèses simplificatrices mais pas aberrantes. Supposons que les mesures aient une densité :  $\mu = f_0 dx$  et  $\nu = f_1 dx$ . La relation (1) devient alors :

$$\begin{aligned} \forall B \in \mathcal{B}(\mathbb{R}^n), \quad \int_B f_1(y) dy &= \int_{T^{-1}(B)} f_0(x) dx \\ &= \int_B \sum_{x \in T^{-1}(y)} \left( \frac{f_0(x)}{|\det \nabla T(x)|} \right) dy \end{aligned}$$

C'est à dire  $f_1(y) = \sum_{x \in T^{-1}(y)} \left( \frac{f_0(x)}{|\det \nabla T(x)|} \right)$ .

Supposons de plus que le transport est injectif et lisse, i.e.  $x \in T^{-1}(y) \leftrightarrow y = T(x)$ . Nous obtenons alors l'équation dite de Jacobienne :

$$f_1(T(x)) = \frac{f_0(x)}{|\det \nabla T(x)|} \quad (13)$$

Notons  $\mathcal{T}(f_0, f_1)$  l'ensemble des applications qui vérifient (13).

Nous pouvons alors donner la définition qui donne le titre à ce projet.

### Définition 2.9 : Problème de transport optimal

Soient,  $f_0$  et  $f_1$  deux densités de probabilité et un coût  $C$ . Le problème de transport optimal est alors le suivant.

Trouver  $T \in \mathcal{T}(f_0, f_1)$  qui réalise le

$$\min_{T \in \mathcal{T}(f_0, f_1)} \int C(x, T(x)) f_0(x) dx \quad (14)$$

### Remarques :

1. Un ensemble de coût est particulièrement intéressant, ce sont ceux qui s'écrivent sous la forme  $C(x, y) = \|x - y\|^p$ . Et plus précisément dans notre problème  $p = 2$ . Nous verrons que le problème de transport optimal possède alors de jolies propriétés.
2. La valeur de (14) est alors appelée distance  $L^2$  de Wasserstein entre  $f_0$  et  $f_1$ , et est notée  $d(f_0, f_1)$ .

### Théorème 2.10 : Existence et unicité du transport optimal

**Preuve :**

□

### Théorème 2.11 :

Dans le cas du coût quadratique, il existe une unique application de transport optimal qui s'écrit comme le gradient d'une fonction convexe.

$$T(x) = \nabla \Psi(x) \quad (15)$$

**Preuve :**

□

A partir de l'équation de la jacobienne (13),  $\Psi$  est solution de l'équation de *Monge - Ampère*

$$f_1(\nabla \Psi(x)) = \frac{f_0(x)}{\det H\Psi(x)} \quad (16)$$

## 2.6 Formulation de J.D. Benamou et Y. Brenier

Pour le moment, le coût ne dépend que des densités initiales et finale, cependant il semble pertinent de chercher à connaître ce qu'il se passe pendant le transport, cela nous amènera à prendre en compte la notion d'obstacles. Dans leur article de 1999 [?], J.D. Benamou et Y. Brenier ont donné une autre formulation au problème de transport optimal en réintroduisant le temps comme variable. Ils obtiennent alors une interprétation très simple du problème de transport optimal en terme de mécanique des fluides.

### **Théorème 2.12 :** *Benamou, Brenier*

Soient  $f_0$  et  $f_1$  deux densités de probabilité assez régulières. Alors,

$$\min_{T \in \mathcal{T}(f_0, f_1)} \int \|x - T(x)\|^2 dx = \min_{(f, v) \in \mathcal{C}_v} \int_0^1 \int_{\mathbb{R}^n} f(t, x) \|v(t, x)\|^2 dt dx \quad (17)$$

Avec,

$$\begin{aligned} f(t, x) &: \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R} \quad \text{la densité} \\ v(t, x) &: \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n \quad \text{champ de vecteurs vitesses} \end{aligned}$$

et

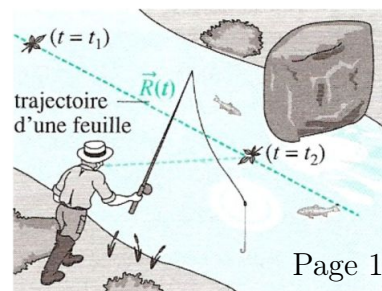
$$C_v = \left\{ (f, v) \mid \frac{\partial f}{\partial t} + \operatorname{div}_x(fv) = 0, f(0, \cdot) = f_0, f(1, \cdot) = f_1 \right\} \quad (18)$$

**Preuve :**

Nous redonnons la preuve donnée par Benamou et Brenier en complétant les points qui nous semble obscurs.

La preuve repose sur l'utilisation des variables lagrangiennes  $X(t, x)$  qui décrivent le comportement d'un fluide en "suivant" les particules. Supposons  $f_0$  et  $f_1$  bornées et à support compact de  $\mathbb{R}^n$ . Considérons  $(f, v)$ , suffisamment régulières, vérifiant (18).

Transport Optimal



En termes formels, cela donne,

$$X(0, x) = x \quad \frac{\partial X}{\partial t}(t, x) = v(t, X(t, x)) \quad (19)$$

Ainsi, pour toute fonction test  $\phi$  :

$$\int \phi(t, x) f(t, x) dx dt = \int \phi(t, X(t, x)) f_0(x) dx dt \quad (20)$$

$$\int \phi(t, x) f(t, x) v(t, x) dx dt = \int \frac{\partial X}{\partial t}(t, x) \phi(t, X(t, x)) f_0(x) dx dt \quad (21)$$

la première égalité vient du changement de variable  $x = X(t, x)$ , or,  $f(t, X(t, x))$  décrit la densité de la particule de fluide déplacée le long du champ de vitesse  $v$  à l'instant  $t$ , c'est donc la densité initiale en  $x$ ,  $f_0(x)$ .

Remarquons que (20) et la condition  $f(0, \cdot) = f_0$ ,  $f(1, \cdot) = f_1$  impliquent que  $T(x) = X(1, x)$  est un transport valide. En effet, pour tout borélien  $B$ ,  $\int_{T(x) \in B} f_0(x) dx = \int_{X(1, x) \in B} f_0(x) dx$  Ce qui signifie, qu'au temps 1, la particule  $x$  est dans le borélien  $B$ , comme la densité au temps 1 est donnée par  $f_1$  on a bien :  $\int_{X(1, x) \in B} f_0(x) dx = \int_B f_1(x) dx$ .

Puis remarquons, en prenant comme fonction test  $\phi(t, x) = \|v(t, x)\|^2$  dans (20), que :

$$\int_{\mathbb{R}^n} \int_0^1 f(t, x) \|v(t, x)\|^2 dx dt = \int_{\mathbb{R}^n} \int_0^1 f_0(x) \|v(t, X(t, x))\|^2 dx dt$$

Nous obtenons,

$$\begin{aligned} &= \int_{\mathbb{R}^n} \int_0^1 f_0(x) \left\| \frac{\partial X}{\partial t}(t, x) \right\|^2 dx dt \quad \text{par (19)} \\ &= \int_{\mathbb{R}^n} f_0(x) \int_0^1 \left\| \frac{\partial X}{\partial t}(t, x) \right\|^2 dt dx \\ &\text{par l'inégalité de Jensen :} \\ &\leq \int_{\mathbb{R}^n} f_0(x) \left\| \int_0^1 \frac{\partial X}{\partial t}(t, x) dt \right\|^2 dx \\ &\leq \int_{\mathbb{R}^n} f_0(x) \|X(1, x) - X(0, x)\|^2 dx \\ &\leq \int_{\mathbb{R}^n} f_0(x) \|X(1, x) - x\|^2 dx \quad \text{par (19)} \end{aligned}$$

Nous reconnaissons ici la forme qui apparaît dans la distance  $L^2$  de Wasserstein, pour le transport  $T(x) = X(1, x)$ . Prenons  $\nabla \Psi(x)$  le transport optimal, nous avons alors :

$$\leq \int_{\mathbb{R}^n} f_0(x) \|\nabla \Psi(x) - x\|^2 dx$$



Finalement, nous avons montré que :

$$\int_{\mathbb{R}^n} \int_0^1 f(t, x) \|v(t, x)\|^2 dx dt \leq \int_{\mathbb{R}^n} f_0(x) \|\nabla \Psi(x) - x\|^2 dx = d(f_0, f_1)^2$$

Il reste à trouver un couple  $(f, v)$  tel que nous ayons l'égalité et la preuve sera complète. Choisissons  $X(t, x) = x + t(\nabla \Psi(x) - x)$ , ce qui correspond à la paire  $(f, v)$  vérifiant pour toute fonction test  $\phi$  :

$$\begin{aligned} \int_{\mathbb{R}^n} \int_0^1 \phi(t, x) f(t, x) dt dx &= \int_{\mathbb{R}^n} \int_0^1 \phi(t, x + t(\nabla \Psi(x) - x)) f_0(x) dx dt \\ \int_{\mathbb{R}^n} \int_0^1 \phi(t, x) f(t, x) v(t, x) dt dx &= \int_{\mathbb{R}^n} \int_0^1 (\nabla \Psi(x) - x) \phi(t, x + t(\nabla \Psi(x) - x)) f_0(x) dt dx \end{aligned}$$

Et en reprenant le raisonnement de la première étape, nous obtenons

$$\begin{aligned} \int_{\mathbb{R}^n} \int_0^1 f(t, x) \|v(t, x)\|^2 dt dx &= \int_{\mathbb{R}^n} \int_0^1 f_0(x) \|v(t, X(t, x))\|^2 \\ &= \int_{\mathbb{R}^n} \int_0^1 f_0(x) \left\| \frac{\partial X(t, x)}{\partial t}(t, x) \right\|^2 dt dx \\ &= \int_{\mathbb{R}^n} f_0(x) \|\nabla \Psi(x) - x\|^2 dx \end{aligned}$$

Ce qui achève la preuve. □

La première équation dans l'ensemble des contraintes correspond à une équation de continuité. Tandis que les conditions sur les bords temporels imposent l'attache au problème initial.

## 2.7 Une autre utilité au transport optimal

Dans les années 1990, la théorie du transport optimal est devenu un outil de démonstration puissant et élégant pour démontrer de nombreuses inégalités géométriques et fonctionnelles. L'exemple le plus simple est fourni par *l'inégalité isopérimétrique*, solution du problème de Didon [?]. Ce problème, peut s'énoncer ainsi : de toutes les courbes fermées de longueur donnée, celle qui entoure l'aire la plus grande est le cercle. Ce problème se généralise, en dimension  $n$ .

### **Théorème 2.13** : *Problème de Didon*

Dans l'espace euclidien  $\mathbb{R}^n$  muni de la mesure de Lebesgue  $\lambda_n$ , pour tout compact  $K$ ,

$$\frac{(\lambda_{n-1} B)^n}{(\lambda_n B)^n} \leq \frac{(\lambda_{n-1} K)^n}{(\lambda_n K)^n}$$

Avec  $B$ , la boule unité.

En particulier en dimension  $n = 3$

**Preuve :**

Considérons les fonctions de densité uniforme et d'intégrale 1, d'une part sur le domaine de volume  $V$  et d'autre part sur une boule unité, donc de volume  $\frac{4\pi}{3}$ . Cela induit l'existence d'un potentiel  $\phi$  permettant le transport de la première vers la seconde, solution de l'équation de Monge - Ampère en tout point  $x$  du volume  $V$  :

$$\frac{1}{V} = \det(I + H\phi(x)) \frac{1}{B}$$

□

**2.8 Conclusion**

Dans cette partie, nous avons introduit le problème de transport optimal. Démontrer l'existence et l'unicité du transport. Le problème se reformule alors.

Et nous avons montré avec Benamou et Brenier que ce problème est équivalent au suivant : qui a une interprétation en terme de mécanique des fluides très pratiques.

Avant d'entamer la résolution numérique du problème de transport optimal, nous présentons les opérateurs proximaux qui joueront un rôle central dans notre modèle.

### 3 Opérateurs proximaux

L'opérateur proximal d'une fonction convexe généralise la notion de projection sur un ensemble convexe. Cette notion fût introduite pour répondre aux problèmes de minimisation de fonctions non nécessairement différentiables.

Suivant la présentation de *P.L. Combettes* et *J.C. Pesquet* [?] nous présentons une série de résultats sur les opérateurs proximaux, en complétant quelques preuves.

**Notation :** Le domaine d'une fonction  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  est  $\mathcal{D}(f) = \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$ . Notons  $\Gamma_0(\mathbb{R}^n)$  l'ensemble des fonctions convexes semi-continues inférieurement à valeurs dans  $]-\infty, +\infty]$  telles que  $\mathcal{D}(f) \neq \emptyset$ .

#### Définition 3.1 : Sous - différentiel

Soit  $f \in \Gamma_0(\mathbb{R}^n)$ , le sous - différentiel de  $f$  est l'application suivante,

$$\partial f : \begin{cases} \mathbb{R}^n & \longrightarrow \mathcal{P}(\mathbb{R}^n) \\ x & \longmapsto \{u \in \mathbb{R}^n \mid \forall y \in \mathbb{R}^n, \langle u, y - x \rangle + f(x) \leq f(y)\} \end{cases}$$

Avec  $\mathcal{P}(\mathbb{R}^n)$ , l'ensemble des parties de  $\mathbb{R}^n$ .

L'interprétation géométrique du sous-différentiel est la suivante. Il est formé par toutes les directions des hyperplan qui passent par le point  $(x, f(x))$  et restent "sous" le graphe de la fonction  $f$ .

**Exemple :** Calculons le sous différentiel en tout point de

$$f : \begin{cases} \mathbb{R} & \longrightarrow \mathbb{R}^+ \\ x & \longmapsto |x| \end{cases}$$

Considérons deux cas,

1.  $x > 0$ , le cas  $x < 0$  est identique. Soit une direction  $u$  appartient au sous différentiel de  $f$  en  $x$ , alors  $\forall y \in \mathbb{R}$  :

$$\begin{aligned} u(y - x) + f(x) &\leq f(y) \\ u(y - x) + x &\leq f(y) \end{aligned}$$

En particulier, pour  $y = 0$  on obtient  $x(1 - u) \leq 0$ , ainsi  $u \geq 1$ . Et donc en prenant  $y > x$  (en particulier,  $y > 0$ ) on obtient  $u(y - x) + x \geq y$  ce qui est contraire à notre définition sauf si  $u = 1$ .

2.  $x = 0$  Soit  $u \in \partial f(0)$ , alors  $u$  doit vérifier pour tout  $y \in \mathbb{R}$ ,

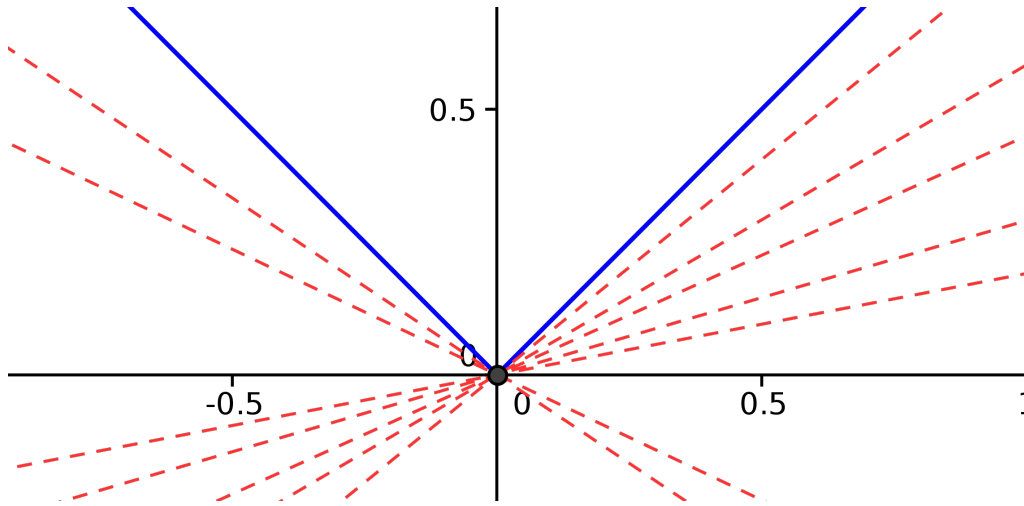
$$uy \leq |y|$$

En prenant  $y > 0$  on obtient que  $u \leq 1$  et en prenant  $y < 0$ , on a  $u \geq -1$ .

En somme,

$$\partial f(x) = \begin{cases} \{-1\} & x < 0 \\ [-1, 1] & x = 0 \\ \{1\} & x > 0 \end{cases}$$

Ce qui correspond bien à l'interprétation géométrique donnée et à la figure 3.

FIGURE 3 – Illustration du sous - différentiel de  $|x|$  en 0**Propriété :**

Soit  $f \in \Gamma_0(\mathbb{R}^n)$ . Si  $f$  est différentiable sur son domaine, alors

$$\forall x \in \mathcal{D}(f)^\circ, \partial f(x) = \{\nabla f(x)\}$$

**Preuve :**

Comme une  $f$  est convexe différentiable, nous avons la caractérisation suivante du gradient :

$$\forall x, y \in \mathbb{R}^n, \quad \langle \nabla f(x), y - x \rangle + f(x) \leq f(y)$$

qui implique que  $\nabla f(x) \in \partial f(x)$ .

Supposons à présent que  $p \in \partial f(x)$ . Soit une suite  $y_n^1$  de  $\mathbb{R}^n$  qui tend vers  $x$ . Écrivons,  $y_n^1 = x + k_n$  avec  $\|k_n\| \rightarrow 0$ . Posons alors  $y_n^2$  une seconde suite de  $\mathbb{R}^n$  telle que  $y_n^2 = x - k_n$ , alors :

$$\begin{cases} \langle p, k_n \rangle \leq f(x + k_n) - f(x) \\ \langle p, -k_n \rangle \leq f(x - k_n) - f(x) \end{cases}$$

Comme  $f$  est différentiable dans son domaine, on peut écrire,

$$f(x \pm k_n) = f(x) \pm \langle \nabla f(x), k_n \rangle$$

Ainsi, nous obtenons

$$\begin{cases} \langle p, k_n \rangle \leq \langle \nabla f(x), k_n \rangle \\ \langle p, k_n \rangle \geq \langle \nabla f(x), k_n \rangle \end{cases}$$

Ainsi  $\langle p, k_n \rangle = \langle \nabla f(x), k_n \rangle$ . En faisant tendre  $n$  vers  $+\infty$ , comme la suite  $k_n$  converge fortement vers 0, elle va aussi converger faiblement. Ce qui implique que  $p = \nabla f(x)$ .  $\square$

Cette propriété est cohérente avec l'exemple précédent,  $\partial f(x) = f'(x)$  partout où,  $f$  est différentiable, c'est à dire  $x \neq 0$ .

**Propriété :**

Si  $f = \alpha f_1 + \beta f_2$ , avec  $f_1, f_2 \in \Gamma_0(\mathbb{R}^n)$  et  $\alpha, \beta \geq 0$ . Alors

$$\begin{aligned}\partial f(x) &= \alpha \partial f_1(x) + \beta \partial f_2(x) \\ &= \{u \in \mathbb{R}^n \mid \exists (u_1, u_2) \in \partial f_1(x) \times \partial f_2(x), u = \alpha u_1 + \beta u_2\}\end{aligned}$$

**Preuve :**

$\alpha \partial f_1(x) + \beta \partial f_2(x) \subset \partial(\alpha f_1 + \beta f_2)(x)$  :  $u \in \mathbb{R}^n$  est dans  $\partial(\alpha f_1 + \beta f_2)(x)$  si et seulement si pour tout  $y \in \mathbb{R}^n$ , nous avons

$$\langle u, y - x \rangle + \alpha f_1(x) + \beta f_2(x) \leq \alpha f_1(y) + \beta f_2(y)$$

En particulier en prenant  $u \in \alpha \partial f_1(x) + \beta \partial f_2(x)$ , au sens donné dans la propriété, cette inégalité est vérifiée.

$\alpha \partial f_1(x) + \beta \partial f_2(x) \supset \partial(\alpha f_1 + \beta f_2)(x)$  : Soit  $u \in \partial(\alpha f_1 + \beta f_2)(x)$ , par contradiction, supposons que  $u \notin \alpha \partial f_1(x) + \beta \partial f_2(x)$ , c'est à dire :

$$\forall (u_1, u_2) \in \partial f_1(x) \times \partial f_2(x), \quad u \neq \alpha u_1 + \beta u_2$$

EMPTY

□

La notion de sous - différentiel est centrale car il permet de caractériser les minimiseurs d'une fonctions (même non différentiable).

**Théorème 3.2 :**

Soit  $f \in \Gamma_0(\mathbb{R}^n)$ . Le point  $x^* \in \mathbb{R}^n$  est un point de minimum global de  $f$  si et seulement si

$$0 \in \partial f(x^*)$$

**Preuve :**

$\Rightarrow$  Soit  $x^*$  le minimum global de  $f$  alors  $\forall y \in \mathbb{R}^n$  :

$$\begin{aligned}f(x^*) &\leq f(y) \\ \langle 0, y - x^* \rangle + f(x^*) &\leq f(y)\end{aligned}$$

Et donc  $0 \in \partial f(x^*)$ .

$\Leftarrow$  Si  $0 \in \partial f(x^*)$  alors  $\forall y \in \mathbb{R}^n$  :

$$\begin{aligned}\langle 0, y - x^* \rangle + f(x^*) &\leq f(y) \\ f(x^*) &\leq f(y)\end{aligned}$$

et  $x^*$  est un minimum global de  $f$ .

□

L'opérateur proximal est un outil important dans l'optimisation de fonctions convexes non différentiables. C'est une application minimisant une fonction fortement convexe dépendant de la fonction originale  $f$ . En trouvant un compromis entre trouver le minimum de la fonction non différentiable  $f$  et en forçant une proximité avec l'argument, nous approchons le minimum de  $f$ .

### Définition 3.3 : Opérateur proximal

Soit  $f \in \Gamma_0(\mathbb{R}^n)$ . L'opérateur proximal de  $f$  est noté  $prox_f$  et il est défini pour tout  $x \in \mathbb{R}^n$  par :

$$prox_f(x) = \arg \min_{y \in \mathbb{R}^n} f(y) + \frac{1}{2} \|x - y\|^2$$

**Exemple :** Prenons, l'indicatrice d'un ensemble convexe non vide  $C$  :

$$i_C = \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases}$$

Alors,

$$\begin{aligned} prox_{i_C}(x) &= \arg \min_{y \in \mathbb{R}^n} i_C(y) + \frac{1}{2} \|x - y\|^2 \\ &= \arg \min_{y \in C} \frac{1}{2} \|x - y\|^2 \\ &= P_C(x) \end{aligned}$$

Où,  $P_C$  désigne la projection orthogonale sur  $C$ .

Sur la figure 4 (tirée de [?]), nous avons en **gras** la frontière de la fonction  $f$ , les lignes fines sont les lignes de niveaux de  $f$ . Évaluer  $prox_f$  sur les points bleus les associent aux points rouges correspondant. Les trois points dans le domaine restent dans le domaine et se déplacent vers le minimum, et les deux à l'extérieur du domaine se placent que la frontière de celui-ci en direction du minimum. Un scalaire  $\gamma > 0$  peut être introduit dans cette définition  $prox_{\gamma f}$ , ce paramètre contrôle le poids apporté à la minimisation de  $f$ . Un  $\gamma$  grand, forcera de se rapprocher fortement du minimum de  $f$ , alors qu'un  $\gamma$  petit favorisera les points proches de  $x$ . Le résultat important est le suivant.

### Propriété :

Soit une fonction  $f$  convexe semi continue inférieurement. Alors,  $prox_f(x)$  existe et est unique pour tout  $x \in \mathbb{R}^n$ . De plus, il est caractérisé par :

$$\forall (x, p) \in \mathbb{R}^n \times \mathbb{R}^n \quad p = prox_f x \quad \Longleftrightarrow \quad x - p \in \partial f(p)$$

En particulier, si  $f$  est différentiable alors :

$$\forall (x, p) \in \mathbb{R}^n \times \mathbb{R}^n \quad p = prox_f x \quad \Longleftrightarrow \quad x - p = \nabla f(p)$$

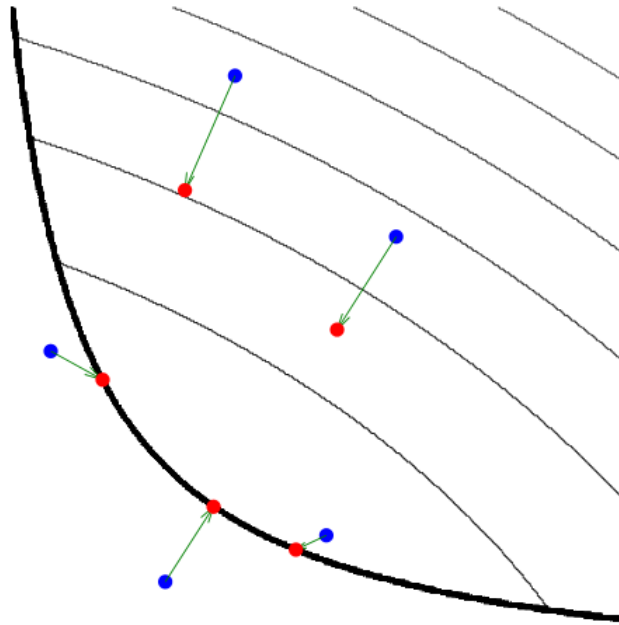


FIGURE 4 – Illustration de l'opérateur proximal

**Preuve :**

$\forall x \in \mathbb{R}^n$ , la fonction  $g : y \rightarrow f(y) + \frac{1}{2}\|x - y\|^2$  est strictement convexe et dans  $\Gamma_0(\mathbb{R}^n)$  (i.e. identiquement égale à  $+\infty$ ). Ce qui montre que le minimum de  $g$  est toujours atteint.

$\Leftarrow$  Supposons que  $x - p \in \partial f(p)$ . Alors, pour tout  $y \in \mathbb{R}^n$ , nous avons

$$\langle x - p, y - p \rangle + f(p) \leq f(y)$$

Or,

$$\begin{aligned} \langle x - p, y - p \rangle &= \frac{1}{2} \langle x - p, y - x + x - p \rangle + \frac{1}{2} \langle x - y + y - p, y - p \rangle \\ &= \frac{1}{2} \|x - p\|^2 + \frac{1}{2} \|y - p\|^2 + \frac{1}{2} (\langle x - p, y - x \rangle + \langle x - y, y - p \rangle) \\ &= \frac{1}{2} \|x - p\|^2 + \frac{1}{2} \|y - p\|^2 - \frac{1}{2} \|x - y\|^2 \end{aligned}$$

Ainsi,

$$\begin{aligned} \frac{1}{2} \|x - p\|^2 + \frac{1}{2} \|y - p\|^2 - \frac{1}{2} \|x - y\|^2 + f(p) &\leq f(y) \\ \frac{1}{2} \|x - p\|^2 + f(p) &\leq f(y) + \frac{1}{2} \|x - y\|^2 \end{aligned}$$

Donc  $p = \text{prox}_f(x)$ .

$\Rightarrow$  Soit  $(x, p) \in \mathbb{R}^n \times \mathbb{R}^n$  telle que  $p = \text{prox}_f(x)$ . Alors  $\forall y \in \mathbb{R}^n$ , je me demande si c'est pas eja prouv   $\square$

De plus,

**Propriété :**

Le point  $x^*$  minimise  $f$  si et seulement si

$$x^* = \text{prox}_f(x^*)$$

**Preuve :**

$\Rightarrow$  Supposons que  $x^*$  minimise  $f$  alors, comme  $\|x^* - y\|^2 > 0$  si  $y \neq x^*$ . Nous avons  $\text{prox}_f(x^*) = x^*$ .

$\Leftarrow$  Utilisons la caractérisation de minimisation donnée par le sous - différentiel. Le point  $\tilde{x}$  minimise  $f(x) + \frac{1}{2}\|x - v\|^2$ , c'est à dire  $\tilde{x} = \text{prox}_f(v)$ , si et seulement si

$$0 \in \partial f(\tilde{x}) + (\tilde{x} - v)$$

Où la somme est entendu au sens de la somme entre un ensemble et un singleton.

En prenant  $x^* = v = \tilde{x}$ , nous obtenons  $0 \in \partial f(x^*)$ . Et donc  $x^*$  minimise  $f$ .  $\square$

Cela suggère des algorithmes de point fixes. Cela fonctionnerait si  $\text{prox}_f$  était une contraction (i.e.  $c$ -lipschitzienne,  $0 < c < 1$ ). Ce n'est pas le cas, cependant nous avons une propriété de non expansivité qui est suffisante pour une itération de point fixe :

$$\|\text{prox}_f(x) - \text{prox}_f(y)\|^2 \leq \langle \text{prox}_f(x) - \text{prox}_f(y), x - y \rangle$$

C'est un cas particulier d'opérateur non expansif (comme par exemple,  $-Id$  ou les rotations,  $[\cdot, \cdot]$ ).

**Définition 3.4 :** *Application non expansive*

Une application  $N : \mathbb{R}^n \rightarrow \mathbb{R}^n$  est dite non expansive si

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \quad \|N(x) - N(y)\| \leq \|x - y\|$$

C'est à dire que  $N$  est 1-lipschitzienne. Nous avons la caractérisation suivante :

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \quad \|N(x) - N(y)\|^2 \leq \langle N(x) - N(y), x - y \rangle$$

Si  $N$  est un opérateur non expansif alors  $\forall \alpha \in ]0, 1[$  l'opérateur  $T = (1 - \alpha)I + \alpha N$  a les mêmes point fixes que  $N$ , et qu'en itérant  $T$ , il y a convergence vers les point fixes. C'est à dire, la suite

$$x_{n+1} = (1 - \alpha)x_n + N(x_n)$$

converge vers un point fixe de  $N$ .

Soit le problème de minimisation :



Soient  $f_1$  et  $f_2$  deux fonctions de  $\Gamma_0(\mathbb{R}^n)$  telles que  $f_1 + f_2 \in \Gamma_0(\mathbb{R}^n)$ , en particulier le support de  $f_1 + f_2$  est non vide. Supposons que

$$\lim_{\|x\| \rightarrow +\infty} f_1(x) + f_2(x) = +\infty$$

Considérons alors le problème de minimisation, Trouver  $x \in \mathbb{R}^n$  qui réalise le

$$\min_{x \in \mathbb{R}^n} f_1(x) + f_2(x) \quad (22)$$

### Propriété :

Le problème 22 admet une unique solution et pour  $\forall \gamma > 0$ , elle est caractérisée par :

$$\begin{cases} x &= \text{prox}_{\gamma f_2} y \\ \text{prox}_{\gamma f_2} y &= \text{prox}_{\gamma f_1} (2\text{prox}_{\gamma f_2} y - y) \end{cases}$$

### Preuve :

$f_1 + f_2$  est une fonction convexe, infinie à l'infini donc l'existence est l'unicité du minimiseur est immédiate.  $\square$

Suivant les idées d'algorithmes de point fixes sur les opérateurs proximaux, nous introduisons l'algorithme suivant dit de *Douglas - Rachford*.

---

### Algorithme 1 Douglas-Rachford

---

Soit  $\epsilon \in ]0, 1[$ ,  $\gamma > 0$ ,  $y_0 \in \mathbb{R}^n$

**for**  $n = 0, 1, \dots$  **do**

$x_n = \text{prox}_{\gamma f_2} y_n$

$\lambda_n \in [\epsilon, 2 - \epsilon]$

$y_{n+1} = y_n + \lambda_n(\text{prox}_{\gamma f_1}(2x_n - y_n) - x_n)$

**end for**

---

### Propriété :

Toute suite  $(x_n)$  générée par l'algorithme de Douglas - Rachford converge vers une solution du problème 22.

## 4 Pbm de transport optimal

Soient deux densités  $f_0$  et  $f_1$  de même masse, suffisamment régulière, sur un domaine  $[0, 1]^d$ . Un transport de  $f_0$  sur  $f_1$  est une application  $T$  telle que :

$$f_0(x) = f_1(T(x))|det(\partial T(x))|$$

Le transport optimal est celui qui minimise le coût  $\int c(x, T(x))dx$  parmi tous les transports de  $f_0$  sur  $f_1$ . Dans notre problème :  $C(x, y) = \|x - y\|^2$ .

## 5 Formulation de Benamou et Brenier

On montre que la géodésique entre  $f_0$  et  $f_1$  est :

$$f(x, t) = f_0((1 - t)Id + tT(x))|det((1 - t)Id + t\partial T(x))|$$

et que ce chemin minimise le problème suivant :

$$\min_{(f, v) \in C_v} \int_{[0, 1]^d} \int_0^1 f(x, t) \|v(x, t)\|^2 dt dx$$

Où :  $C_v = \{(f, v) | \partial_t f + div_x(v) = 0; v(0, \cdot) = 0, v(1, \cdot) = 0, f(\cdot, 0) = f_0, f(\cdot, 1) = f_1\}$

On pose  $m = fv$  pour obtenir le problème de minimisation

$$\min_{(f, m) \in C} J(m, f) = \int_{[0, 1]^d} \int_0^1 \theta(m, f) dt dx$$

avec  $\theta(m, f) = \frac{\|m\|^2}{f} si f > 0 \quad 0 si(m, f) = (0, 0) \quad \infty sinon$

Rq : on a ramené le pbm de transport optimal à celui de trouver la géodésique entre les deux densités.

## 6 Méthodes numériques

On présente le cas 1D qui nous a accaparé.

Carré espace temps :  $[0, 1]^2$  que l'on discrétise :

$$G_c = \{(x_i = \frac{i}{N}, t_j = \frac{j}{P}) | 0 \leq i \leq N, 0 \leq j \leq P\}$$

On note  $E_c = (\mathbb{R}^2)^{G_c}$  l'espace des variables centrées et  $V = (m_{ij}, f_{ij}) \quad 0 \leq i \leq N, 0 \leq j \leq P$  les variables discrétisées.

Dans le but de capturer l'équation de continuité, on introduit une grille décentrée :

$$G_s^x = \{(x_i = \frac{i + 1/2}{N}, t_j = \frac{j}{P}) | -1 \leq i \leq N, 0 \leq j \leq P\}$$

et

$$G_s^t = \{(x_i = \frac{i}{N}, t_j = \frac{j + 1/2}{P}) | 0 \leq i \leq N, -1 \leq j \leq P\}$$

On note  $E_s = \mathbb{R}^{G_s^x} \times \mathbb{R}^{G_s^t}$  l'espace des variables décentrées, et  $U = (\bar{m}_{ij}, \bar{f}_{ij}) \quad 1 \leq i \leq N, 0 \leq j \leq P, \bar{f}_{ij}, \quad 0 \leq i \leq N, -1 \leq j \leq P$

## 7 Opérateurs

On introduit plusieurs opérateurs pour lier les variables décentrées et les variables centrées.

### 7.1 Interpolation

$$I : E_s \rightarrow E_c$$

tq

$$m_{ij} == (\bar{m}_{i+1/2,j} + \bar{m}_{i-1/2,j})/2.$$

et

$$f_{ij} == (\bar{f}_{i,j+1/2} + \bar{f}_{i,j-1/2})/2.$$

Cet opérateur s'interprète matriciellement :

$$m = \bar{m} I_m$$

et

$$f = I_f \bar{f}$$

### 7.2 Divergence

l'opérateur qui approxime la divergence

$$div : E_s \rightarrow \mathbb{R}^{G_c}$$

et

$$div(\bar{m}, \bar{f})_{ij} = (\bar{m}_{i+1/2,j} - \bar{m}_{i-1/2,j}) + (\bar{f}_{i,j+1/2} - \bar{f}_{i,j-1/2})$$

### 7.3 Frontières

Un opérateur pour extraire les frontières :

$$b(\bar{m}, \bar{f}) = ((\bar{m}_{-1/2,j}, \bar{m}_{N+1/2,j}); (\bar{f}_{i,-1/2}, \bar{f}_{i,P+1/2}))$$

et on impose les conditions aux frontières suivantes :

$$b(\bar{m}, \bar{f}) = b_0 = (0, 0, f_0, f_1);$$

### 7.4 Problème discrétisé

On a le problème suivant

$$\min_{U=(\bar{m}, \bar{f}) \in E_s} \theta(I(\bar{m}, \bar{f})) + \iota_C(U)$$

avec l'ensemble des contraintes :

$$C = \{(\bar{m}, \bar{f}) \in E_s \mid div(\bar{m}, \bar{f}) = 0 \text{ } b(\bar{m}, \bar{f}) = b_0\}$$

## 8 Résolution par algorithme de séparation de proximité

Ces algorithmes sont des généralisation des algos de gradient conjugués.

**Remarque :** On a  $J(m, f) = \int_{[0,1]^d} \int_0^1 \frac{\|m\|^2}{f} dt dx$  donc si  $f \rightarrow \infty$  on a  $J \rightarrow 0$  donc ce n'est pas coercif ce qui rend l'existence de minimiseurs non triviale. Et si  $f \rightarrow 0$ , on a  $j \rightarrow \infty$  donc les méthodes de gradient conjugués ne peuvent pas s'appliquer, le gradient n'est pas lipschitz.

On veut résoudre le pbm suivant :

$$\min_{z=(U,V) \in E_s \times E_c} G_1(z) + G_2(z)$$

où,  $G_1(z) = J(U) + \iota_C(U)$  et  $G_2(z) = \iota_{C_s}(z)$  et  $C_s = \{z = (U, V) \in E_s \times E_c \mid V = I(U)\}$

**Remarque :**  $G_1$  est la fonctionnelle originelle et  $G_2$  vient de notre introduction des variables décalées.

On va alors calculer les opérateurs de proximités de  $G_1$  et  $G_2$ . On dit que  $G_1$  est **simple** dans la mesure où :

$$Prox_{\gamma G_1}(U, V) = (Prox_{\gamma C}(U), Prox_{\gamma J}(V))$$

### 8.1 Opérateur de proximité de $G_2$

$$Prox_{\gamma C_s}(U, V) = \arg \min_{z' \in C_s} \frac{1}{2} \|z - z'\|^2 = Proj_{C_s}(U, V)$$

Suivant l'article de Papadakis, Peyré et Oudet on a :

$$Prox_{\gamma C_s}(U, V) = (\tilde{U}, I(\tilde{U}))$$

et

$$\tilde{U} = (Id + I^* I)^{-1}(U + I^*(V))$$

Où  $I^*$  est l'adjointe de  $I$ . ce qui nous donne en terme de matrice :

$$\tilde{m} = (\bar{m} + m I_m^*)(Id + I_m I_m^*)^{-1}$$

et

$$\tilde{f} = (Id + I_f^* I_f)^{-1}(\bar{f} + I_f^* f)$$

enfin :  $\tilde{m} = \tilde{m} I_m$  et  $\tilde{f} = I_f \tilde{f}$

Comme l'opérateur d'interpolation s'interprète matriciellement, son adjoint est donné par la transposée.

## 8.2 Opérateur de proximité de $J$

$Prox_{\gamma J}(V) = (Prox_{\gamma J}(V_k))_{k \in G_c}$  et

$$Prox_{\gamma J}(m_k, f_k) = \left( \frac{f_k^* m_k}{f_k^* + 2\gamma}, f_k^* \right) \text{ si } f^* > 0 \quad (0, 0) \text{ sinon}$$

et  $f_k^*$  est la plus grande racine réelle du polynôme de degré 3 donné par :

$$P(x) = (X - f_k)(X + 2\gamma)^2 - \gamma \|m_k\|^2$$

La racine est calculée rapidement avec l'algorithme de Newton.

## 8.3 Opérateur de proximité de $\iota_C$

$$Prox_{\gamma C} = Proj_C$$

on réécrit l'ensemble des contraintes sous la forme :

$$C = \{U \in E_s \mid AU = y\} \quad A = (div, b)^T \quad y = (0, b_0)^T$$

et on obtient  $Proj_C(U) = (Id + A^* \Delta^{-1} A)U + A^* \Delta^{-1} y$  avec  $\Delta^{-1} = (A^* A)^{-1}$  et son problème requiert la résolution d'un problème de poisson.

## 9 Algorithme de Douglas Rachford

On procède à l'itération suivante :  $(z^l, w^l) \in (E_s \times E_c) \times (E_s \times E_c)$  avec un  $w^0$  donné :

$$z^{l+1} = Prox_{\gamma G_2}(w^l)$$

et

$$w^{l+1} = \left(1 - \frac{\alpha}{2}\right)w^l + \frac{\alpha}{2} RProx_{\gamma G_2} \circ RProx_{\gamma G_1}(w^l)$$

Où on a posé :  $RProx_{\gamma G} = 2Prox_{\gamma G} - Id$ .

On montre alors que pour  $\alpha \in ]0, 2[$  et  $\gamma > 0$  la méthode converge  $z^l \rightarrow z^*$  ce qui permet de retrouver la géodésique, car :

$$z^l = (U^l, V^l) \rightarrow (U^*, V^*) \quad V^* = (m^*, f^*)$$

# Appendices

## A Calcul des variations

Rapellons ici la méthode directe du calcul des variations qui nous servira pour notre démonstration.

### Définition A.1 : Fonction continue inférieurement

Dans un espace métrique  $\mathcal{X}$ , une fonction  $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$  est dite semi-continue inférieurement si pour toute suite  $x_n \rightarrow x$

$$f(x) \leq \liminf_n f(x_n)$$

### Définition A.2 : Espace compact

Un espace métrique  $\mathcal{X}$  est dit compact si de toute suite  $(x_n)$  il existe une sous suite convergente  $x_{n_k} \rightarrow x \in \mathcal{X}$ .

### Théorème A.3 : Weierstrass

Si  $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$  est semi-continue inférieurement, et  $\mathcal{X}$  est un espace compact. Alors, il existe  $\bar{x} \in \mathcal{X}$  tel que  $f(\bar{x}) = \inf_{x \in \mathcal{X}} f(x)$

#### Preuve :

Soit  $l := \inf \{f(x) \mid x \in \mathcal{X}\} \in \mathbb{R} \cup \{-\infty\}$ . ( $l = +\infty$  si et seulement si  $f$  est identiquement égale à  $+\infty$  auquel cas, tout point de  $\mathcal{X}$  minimise  $f$ . Par définition de l'infimum, il existe une suite  $(x_n) \in \mathcal{X}$  telle que  $f(x_n) \rightarrow l$ . Quitte à extraire une sous suite, on peut supposer que  $x_n \rightarrow \bar{x}$ . Par semi-continuité inférieure, nous avons  $f(\bar{x}) \leq \liminf_n f(x_n) = l$ . D'autre part,  $f(\bar{x}) \geq l$  puisque  $l$  est l'infimum.

Ainsi,  $l = f(\bar{x}) \in \mathbb{R}$ . □

### Définition A.4 : Convergence faible et faible-\*

Une suite  $x_n$  d'un espace de Banach  $\mathcal{X}$  converge faiblement vers  $x$ , que nous notons  $x_n \rightharpoonup x$ , si pour tout  $\xi \in \mathcal{X}'$  (où,  $\mathcal{X}'$  est le dual topologique de  $\mathcal{X}$  et  $\langle \cdot, \cdot \rangle$  désigne le produit de dualité entre ces deux espaces), nous avons  $\langle x_n, \xi \rangle \rightarrow \langle x, \xi \rangle$ .

Une suite  $\xi_n \in \mathcal{X}'$  converge faiblement-\* vers  $\xi$ , que nous notons  $\xi_n \xrightarrow{*} \xi$ , si pour tout  $x \in \mathcal{X}$  nous avons  $\langle x, \xi_n \rangle \rightarrow \langle x, \xi \rangle$ .

### Théorème A.5 : Banach - Alaoglu

Si  $\mathcal{X}$  est séparable et  $\xi_n$  est une suite bornée de  $\mathcal{X}'$ . Alors il existe une sous suite  $\xi_{n_k}$  qui converge faiblement-\* vers un  $\xi \in \mathcal{X}'$ .

**Définition A.6 :** *Mesures de Radon*

Une mesure  $\lambda$  sur la tribu borélienne d'un espace  $\Omega$  séparable est dite mesure de Radon si

1.  $\lambda$  est intérieurement régulière :  $\forall B \in \mathcal{B}(\Omega), \lambda(B) = \sup_{K \subset B, K \text{ compact}} \lambda(K),$
2.  $\lambda$  est localement finie : pour tout point  $x$  il existe un voisinage  $B$  tel que  $\lambda(B) < +\infty.$

Notons  $\mathcal{M}(\Omega)$  l'ensemble des mesures de Radon sur  $\Omega$ . Pour ces mesures, il est possible d'associer une mesure positive  $|\lambda| \in \mathcal{M}_+(\Omega)$  :

$$|\lambda|(B) = \sup \left\{ \sum_i |\lambda(B_i)| \mid B = \bigcup_i B_i, B_i \cap B_j = \emptyset, i \neq j \right\}$$

**Théorème A.7 :** *Riesz pour les mesures de Radon*

Soient  $\Omega$  une espace séparable et localement compact,  $\mathcal{X} = C_0(\Omega)$  l'espace des fonctions continues qui s'annulent à l'infini, muni de la norme sup.

Alors, tout élément de  $\mathcal{X}'$  est représenté de façon unique par un élément de  $\mathcal{M}(\Omega)$ . C'est à dire,  $\forall \xi \in \mathcal{X}'$  il existe un unique  $\lambda \in \mathcal{M}(\Omega)$  tel que  $\langle \phi, \xi \rangle = \int \phi d\lambda \quad \forall \phi \in \mathcal{X}.$

De plus  $\mathcal{X}'$  est isomorphe  $\mathcal{M}(\Omega)$  muni de la norme  $\|\lambda\| = |\lambda|(\Omega).$

Pour les mesures de Radon de  $\mathcal{M}(\Omega)$ , la converge faible-\* devrait être celle de la dualité avec  $C_0(\Omega)$ . Cependant, une autre notion de convergence est intéressante, celle avec la dualité de  $C_b(\Omega)$ , les fonctions continues bornées sur  $\Omega$ . Par abus de notations, nous la noterons aussi  $\rightharpoonup$  :  $\lambda_n \rightharpoonup \lambda$  si et seulement si pour toute fonction  $\phi \in C_b(\Omega)$  nous avons  $\int \phi d\lambda_n = \int \phi d\lambda$ . Remarquons que si  $\Omega$  est compact,  $C(\Omega) = C_0(\Omega) = C_b(\Omega)$  et ces notions de convergences sont alors égales.

Les mesures de probabilités sur  $\Omega$ ,  $\mathcal{P}(\Omega)$ , sont un sous-ensemble particulier de  $\mathcal{M}(\Omega)$ .  $\mu \in \mathcal{P}(\Omega) \Leftrightarrow \mu \in \mathcal{M}_+(\Omega)$  et  $\mu(\Omega) = 1$ , dans ce cas,  $\mu$  et  $|\mu|$  coïncident.

**Définition A.8 :** *Tension d'une suite de mesure de probabilité*

Une suite  $\mu_n$  de probabilité sur  $\Omega$  est dite tendue si pour tout  $\epsilon > 0$  il existe un sous-ensemble compact  $K \subset \Omega$  tel que  $\mu_n(\Omega \setminus K) < \epsilon$  pour tout  $n$ .

**Théorème A.9 :** *Prokhorov*

Soit  $\mu_n$  une suite tendue de mesure de probabilité sur un espace métrique complet et séparable  $\Omega$ . Alors il existe une mesure de probabilité sur  $\Omega$ ,  $\mu$ , et une sous-suite  $\mu_{n_k}$  telle que  $\mu_{n_k} \rightharpoonup \mu$  (dans la dualité avec  $C_b(\Omega)$ ).

Réciproquement, toute suite  $\mu_n \rightharpoonup \mu$  est nécessairement tendue.

Rappelons le résultat principal concernant la compacité dans l'espace des fonctions continues.

**Théorème A.10** : *Ascoli - Arzelà*

Si  $\Omega$  est un espace métrique compact et  $f_n : \Omega \rightarrow \mathbb{R}$  est équicontinue et équiborné. Alors il existe une sous suite  $f_{n_k}$  qui converge uniformément vers une fonction continue  $f : \Omega \rightarrow \mathbb{R}$ . Réciproquement, un sous-ensemble de  $C(\Omega)$  est relativement compact pour la convergence uniforme si et seulement si tout ses éléments sont équicontinus et équibornés.



## Table des figures

1	Illustration de l'application de transport . . . . .	4
2	Illustration de la notion de plan de transport . . . . .	5
3	Illustration du sous - différentiel de $ x $ en 0 . . . . .	20
4	Illustration de l'opérateur proximal . . . . .	23

## Liste des Codes